Politecnico di Bari

Department of Mechanics, Mathematics and Management
MECHANICAL AND MANAGEMENT ENGINEERING
Ph.D. Program
SSD: ING-IND/15 – DESIGN METHOD FOR INDUSTRIAL ENGINEERING

Final Dissertation

OPTIMIZATION OF AUGMENTED REALITY INTERFACES FOR TECHNICAL DOCUMENTATION

by
Enricoandrea Laviola
Enricoandrea Laviola

Referees:

Prof. Joseph L. Gabbard

Prof. Markku Turunen

Supervisors:

Prof. Antonio E. Uva

Prof. Michele Gattullo

Michele Gattullo

Coordinator of Ph.D Program:

Prof. Giuseppe P. Demelio

Course n°36, 01/11/2020-31/10/2023

Index

Abstract

Augmented Reality (AR) is one of the enabling technologies of Industry 4.0. It aims to provide information that seamlessly integrates with the physical environment. In the industrial context, AR plays an important role in combining real and virtual assets, particularly in complex maintenance and assembly procedures. Industrial Augmented Reality (IAR) facilitates technical communication by offering spatially aligned information and instructions linked to real objects. Consequently, IAR enriches the physical environment with supplementary information that surpasses what is naturally present, resulting in reduced cognitive load for workers compared to traditional paper-based procedures.

While numerous methods have been proposed to generate visual assets for IAR interfaces, the literature lacks agreement on the most effective approach for delivering AR instructions. Therefore, the main goal of this dissertation is to establish comprehensive guidelines for creating the next-generation IAR Technical Documentation (TD). The research project aims to provide clear recommendations for optimizing the presentation of instructions within an industrial context, specifically focusing on assessing the tangible economic benefits AR can bring to companies.

The findings of this dissertation show that this optimization not only enhances operators' performance by providing them with more easily understandable information but also reduces developers' computational costs for companies in terms of programming and 3D modeling. Finally, by following the provided guidelines, companies can streamline the authoring process of Augmented Reality Technical Documentation (ARTD), even for technical writers who are not particularly skilled in AR.

List of publications

[1] E. Laviola and A. E. Uva, "From Lab to Reality: Optimization of Industrial Augmented Reality Interfaces," 2022 IEEE Int. Symp. Mix. Augment. Real. Adjun. Dr. Consort., pp. 931–934, 2022, doi: 10.1109/ISMAR-Adjunct57072.2022.00208.

[2] M. Gattullo, E. Laviola, and A. E. A. E. Uva, "From Therbligs to Visual Assets: a Technique to Convey Work Instructions in Augmented Reality Technical Documentation," in Advances on Mechanics, Design Engineering and Manufacturing IV, Springer, Cham, 2022, pp. 1327–1339. doi: 10.1007/978-3-031-15928-2_116.

[3] E. Laviola, M. Gattullo, V. M. Manghisi, M. Fiorentino, and A. E. Uva, "Minimal AR: visual asset optimization for the authoring of augmented reality work instructions in manufacturing," Int. J. Adv. Manuf. Technol., 2021,doi: 10.1007/s00170-021-08449-6.

[4] M. Gattullo, E. Laviola, A. Evangelista, M. Fiorentino, and A. E. Uva, "Towards the Evaluation of Augmented Reality in the Metaverse: Information Presentation Modes," Int. J. Interact. Des. Manuf., 2022, doi: https://doi.org/10.3390/app122412600.

[5] E. Laviola, A. E. Uva, and M. Gattullo, "The Minimal AR authoring approach: validation in a real assembly scenario," Comput. Ind., vol. 154, no. October 2023, 2023, doi: 10.1016/j.compind.2023.104026.

[6] E. Laviola, S. Romano, M. Gattullo, and A. E. Uva, "Evaluating the Worker Technology Acceptance of a Mixed Reality Technical Documentation," in Virtual Reality and Mixed Reality, 2023.

[7] S. Romano, E. Laviola, M. Gattullo, M. Fiorentino, and A. E. Uva, "More Arrows in the Quiver: Investigating the Use of Auxiliary Models to Localize In-view Components with Augmented Reality," 2023 IEEE Int. Symp. Mix. Augment. Real., vol. PP, pp. 1–11, 2023, doi: 10.1109/TVCG.2023.3320229.

[8] E. Laviola, M. Gattullo, A. Evangelista, M. Fiorentino, and A. E. Uva, "In-situ or side-by-side? A user study on augmented reality maintenance instructions in blind areas," Comput. Ind., vol. 144, no. June 2022, p. 103795, 2023, doi: 10.1016/j.compind.2022.103795.

[9] E. Laviola, S. Romano, and M. Gattullo, "Virtual Mirror for Maintenance: an Augmented Reality Support Tool," in Design Tools and Methods in Industrial Engineering, 2023.

[10] S. Romano, E. Laviola, and M. Gattullo, “ADAM: Automatic Development of Auxiliary Models. An Authoring Tool for Augmented Reality Technical Documentation,” in Design Tools and Methods in Industrial Engineering, 2023.
[11] E. Laviola, M. Gattullo, and A. Evangelista, “Displaying Augmented Reality Manuals in the Design Phase of the Product Lifecycle,” in Advances on Mechanics, Design Engineering and Manufacturing IV, Springer, Cham, 2022, pp. 1316–1326. doi: 10.1007/978-3-031-15928-2_115.
[12] M. Gattullo, E. Laviola, M. Fiorentino, and A. E. Uva, “Positive Computing in Virtual Reality Industrial Training,” in 2021 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct), Oct. 2021, pp. 364–366. doi: 10.1109/ISMAR-Adjunct54149.2021.00083.
[13] M. Gattullo et al., “Biophilic Enriched Virtual Environments for Industrial Training: a User Study,” in 2022 IEEE International Symposium on Mixed and Augmented Reality (ISMAR). doi: 10.1109/ISMAR55827.2022.00035.
[14] S. Romano, E. Laviola, A. E. Uva, and M. Gattullo, “Biophilic Design of Virtual Workplaces: Effect of Animations on User Attention,” 2023.
[15] M. Gattullo, E. Laviola, and A. E. Uva, “Exploiting Augmented Reality in LEGO Therapy for Children with Autism Spectrum Disorder,” in Virtual Reality and Mixed Reality, 2022, pp. 73–85. doi: 10.1007/978-3-031-16234-3_5.
[16] E. Laviola et al., “Mixed Reality in STEM Didactics: Case Study of Assembly Drawings of Complex Machines,” in Design Tools and Methods in Industrial Engineering II, 2022, pp. 157–164. doi: 10.1007/978-3-030-91234-5_16.
[17] M. Gattullo et al., “Design of a Mixed Reality Application for STEM Distance Education Laboratories,” Computers, 2022, doi: https://doi.org/10.3390/computers11040050.

List of abbreviations

AM	Auxiliary Model
AR	Augmented Reality
ARTD	Augmented Reality Technical Documentation
CAD	Computer-Aided Design
GUI	Graphical User Interface
HMD	Head-Mounted Display
HHD	Hand-Held Display
IAR	Industrial Augmented Reality
MR	Mixed Reality
MRTD	Mixed Reality Technical Documentation
MRTK	Mixed Reality Toolkit
OOBB	Object-Oriented Bounding Box
OST	Optical See-Through
SAR	Spatial Augmented Reality
SUS	System Usability Scale
TAM	Technology Acceptance Model
TD	Technical Documentation
TTD	Traditional Technical Documentation
UI	User Interface
VR	Virtual Reality
VST	Video See-Through

Introduction[1]

To ensure the safety of individuals, machinery, and systems, it is fundamental to present maintenance, assembly, and training instructions with reliable Technical Documentation (TD). Traditional Technical Documentation (TTD), such as printed manuals or Electronic Work Instructions, often consists of extensive texts or complex images that may be challenging to comprehend with real working products. Augmented Reality (AR) is a powerful tool for supporting operators in manual procedural tasks, including assembly and maintenance [1], [2]. This innovative technology has prompted industrial companies to transition from TTD to a new AR version. In Augmented Reality Technical Documentation (ARTD), virtual work instructions can be registered in 3D directly on real products. Moreover, the information can be presented more concisely and tailored using specific visual assets, such as Computer-Aided Design (CAD) models, drawings, and videos [3].

Introducing this new form of TD can potentially enhance operators' performance. However, it also brings new challenges during the authoring phase, primarily due to the complex requirements inherent in industrial settings [4]. These challenges stem from the intricate interplay between the real environment and the larger variety of visual assets available that can be used. Misinterpreting these aspects could result in the inclusion of visual assets that demand more authoring effort and are not suitable for operators, ultimately leading to a decline in performance. Consequently, the design of instructions for ARTD necessitates highly skilled developers. However, many companies lack the necessary expertise in AR, which either inhibits the adoption of AR in the industry [5] or results in ARTD interfaces where visual assets are not effectively optimized. In the absence of established guidelines, ARTD interfaces may rely solely on the personal preferences of the ARTD author. Therefore, it is necessary to conduct research studies that provide clear guidelines and case studies, enabling the authoring of optimized ARTD interfaces even by individuals without prior expertise in AR.

An optimized ARTD interface is necessary to enhance operators' performance, regardless of their experience level, by increasing their speed and efficiency during task execution. This, in turn, will lead to cost reduction for the company. As proposed in [6], the initial phase in designing an

[1] The purpose of this dissertation described in this Section is summarized in the following paper:
E. Laviola and A. E. Uva, “From Lab to Reality: Optimization of Industrial Augmented Reality Interfaces,” 2022 IEEE Int. Symp. Mix. Augment. Real. Adjun. Dr. Consort., pp. 931–934, 2022, doi: 10.1109/ISMAR-Adjunct57072.2022.00208.

ARTD involves analyzing work instructions. The latter should be decomposed into smaller pieces of information, and appropriate visual assets should be suggested for each elemental information. Nevertheless, the existing literature lacks agreements regarding the parameters for analyzing work instructions and choosing corresponding visual assets for elemental information. Some studies propose visual assets with varying levels of informational detail, considering the operator's experience or preference [7]. This often implies for every information type using animated product models, i.e., 3D virtual models of products and parts [8], [9]. Other studies decompose the assembly task into basic operations without a standardized approach [10]. In certain cases, task complexity is also considered [11]. However, determining the complexity of a task is challenging due to the involvement of multiple factors, making it difficult to establish a precise definition [10].

In the industrial context, many studies on AR have primarily focused on evaluating the concept of this technology itself rather than emphasizing the importance of information content [12]. Despite the numerous methods proposed for generating visual assets for industrial AR interfaces, there is no agreement in the literature regarding the optimal approach for providing AR instructions [13]. Based on the analysis of the current state of the art, the following observations can be made:

- Companies desire to use AR for their TD with a standardized framework that enables the development of an ARTD, even for technical writers who are not AR experts.
- Designing an ARTD requires decomposing work instructions into elemental information, for which appropriate visual assets must be suggested.
- Clear guidelines for decomposing work instructions into elemental information are absent.
- Clear guidelines on conveying information through visual assets in ARTD are lacking.
- Clear guidelines on selecting the most suitable visual assets based on the working context are missing.
- Product models are the most used visual assets for every information type; however, the selection process often lacks an objective criterion.

The absence of these guidelines can lead to the ineffectiveness of AR prototypes tested in controlled laboratory environments when used in real industrial settings. However, despite this limitation, the global AR market demonstrated significant growth, with a value of USD 37.0 billion in 2022. Projections indicate that it is expected to reach USD 114.5 billion by 2027, with an impressive annual growth rate of 25.3% over the five-year forecast period [14]. Furthermore, the potential of AR to optimize work processes is estimated to reduce manufacturing plant downtime by up to 50%, resulting in substantial cost reductions and time savings for companies, thereby enhancing their competitive advantage. Nonetheless, these statements remain predictions, and concrete validation of the economic impact of AR for companies can only be achieved after establishing appropriate standards for optimizing ARTD authoring through multiple case studies in various industrial contexts.

Therefore, the contribution of my thesis work is to address the existing gaps in both the literature and the industrial domain, focusing on assessing the real economic impact that AR can offer to companies. The research project aims to identify the most effective approach for conveying instructions in AR for manual documentation. The resulting outcomes will provide guidelines to enhance operators' performance and simplify the ARTD authoring process, particularly for technical writers lacking expertise in AR.

In Chapter 1, Industrial Augmented Reality (IAR) is introduced as one of the enabling technologies of Industry 4.0. A comprehensive overview of AR technology and its specific application for TD within the industrial context is provided. Then, an overview of the techniques for TD authoring is described, highlighting the lacks in the literature. Finally, the objectives and scope of this dissertation are presented.

In Chapter 2, an information model is described to decompose instructions into elemental information. A preliminary study is described to determine if it is effectively possible to decompose instructions into the proposed elemental information.

In Chapter 3, the “minimal AR” authoring approach is presented to associate visual assets based on the information conveyed. Its main goal is to provide only the information needed to accomplish a task through visual assets.

In Chapter 4, two preliminary user studies focused on applying our “minimal AR” authoring approach are described to extract important guidelines for possible association of visual assets based on information types.

In Chapter 5, it is demonstrated that the identified guidelines are also valid in a real industrial context through a user study including real maintenance workers.

In Chapter 6, tangible solutions for visual asset optimization based on the information type and the working context are presented with user studies that examine factors such as surroundings, lighting conditions, and blind areas.

Finally, Chapter 7 describes our developed systems to simplify the authoring based on the previous guidelines identified by the "minimal AR" approach and the optimizations investigated for the information type and the working context.

Chapter 1. State of the art on Augmented Reality technical documentation

1.1. Augmented Reality technology

AR seeks to present information that is seamlessly integrated with the physical surroundings. An augmented scene typically consists of three main elements:

(1) a real-world object (feature)
(2) its designated position within the augmented scene (anchor)
(3) a virtual model associated with the real-world object (visual asset).

In 1997, Azuma [15] provided an AR definition, which is widely accepted in the scientific community nowadays. According to this definition, AR must have the following three fundamental attributes:

- it combines real and virtual elements
- it is interactive in real time
- it has a registration in 3D space.

Specific output devices are necessary to integrate virtual and real stimuli following Azuma's definition. Additionally, AR technology requires spatial registration and interactivity, which entails that virtual elements must be accurately aligned with the real environment in real-time. This alignment is facilitated by a tracking system, which dynamically determines the spatial characteristics of the virtual elements within the physical environment. Lastly, the interactivity aspect also requires an AR system to have a user interaction module that can seamlessly interact with the tracking system. Analyzing Azuma's definition, then, three key technologies can be identified: display, tracking, and User Interface (UI).

1.1.1. Display

The main factor distinguishing AR displays from conventional computer displays is the ability to merge a virtual word into the real one. They can be distinguished into two main categories: visual and nonvisual. The difference lies in the presence or absence of the sense of sight in the AR experience. Out of the visual domain, nonvisual displays include audio [16], haptic [17], olfactory and gustatory [18] devices. However, for obvious reasons, visual displays are found to have a much wider application. Among these, three main types of devices can be distinguished based on the user's eye distance: Head-Mounted Displays (HMD), Hand-Held Displays (HHD), and projection-based displays [19] (Fig. 1).

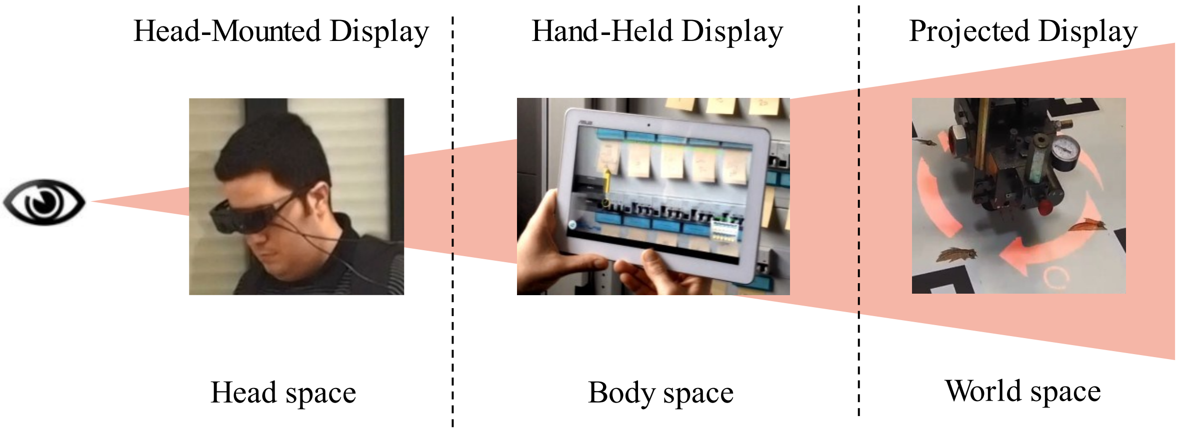


Fig. 1 The most widely used visual displays classified by user's eye distance.

HMDs have the main advantage of being able to be used by granting free hands to the user to perform a task. They can be Optical See-Through (OST) or Video See-Through (VST). The OST HMDs enable users to see the real word using their natural eyes while overlaying graphics onto their view thanks to a holographic optical element. The VST HMDs, instead, are a type of display technology in which the user is provided with a video view of the real world with graphics overlaid on it. HHDs (e.g., mobile phones, tablets) are considered a favorable alternative to HMDs because they offer several advantages such as minimal intrusion, social acceptability, easy availability and mobility. However, the main disadvantage is that they require using the user's hands. Projection-based displays, also called Spatial Augmented Reality (SAR) systems, are a good option for applications where multiple users are involved, and the goal is to minimize intrusiveness. This type of display technology projects augmented content onto a surface or object, eliminating the need for users to wear any specific device or equipment.

1.1.2. Tracking

The tracking system plays a crucial role in AR by facilitating the registration of virtual objects in a real environment. Its primary function is to ensure that virtual elements are precisely aligned with the physical surroundings in real-time. As a result, the AR tracking system is considered one of the critical components for delivering a compelling AR experience. It dynamically determines the spatial properties, specifically the six degrees of freedom of virtual elements within the real environment. This includes information about the position and orientation of the virtual objects in relation to the user's viewpoint. By accurately tracking the user's movements and real-world objects, the tracking system enables seamless integration of virtual elements with the physical space.

Three types of tracking techniques can be distinguished: sensor-based, vision-based, and hybrid. Sensor-based tracking techniques rely on various types of sensors to track and determine the spatial properties of virtual objects within the real environment. These sensors can include magnetic, acoustic, inertial, optical, and mechanical ones [20]. Vision-based tracking techniques use image processing methods to calculate the camera pose relative to real-world objects [21]. We can distinguish feature-based and model-based techniques [22]. The rationale behind feature-based methods is to establish a correspondence between 2D image features and their corresponding 3D coordinates in the real-world frame. Model-based methods, instead, rely on constructing models using lines or edges present in the model itself. Among these features, edges are commonly preferred due to their computational efficiency and robustness against changes in lighting conditions. In certain AR applications, relying solely on computer vision for tracking may not offer a sufficiently robust solution. As a result, hybrid methods have been developed to combine multiple sensing technologies to enhance tracking accuracy and robustness. One example is the work proposed by Azuma et al. [23], who suggested that AR systems designed for outdoor environments benefit from a tracking system that integrates GPS, inertial sensors, and computer vision sensing.

1.1.3. User Interface

One of the most important aspects of AR technology is to establish intuitive interaction techniques between the user and the virtual content. Four types of user interfaces are commonly employed in AR applications: tangible, collaborative, hybrid, and multimodal. Tangible interfaces directly interact with the real world by leveraging physical objects and tools. An example of the effectiveness of tangible user interfaces is the application developed by Kato et al. [24]. This application enables users to select and rearrange furniture within an AR living room design application by employing a real, physical paddle as an interaction tool. While this type of user interface may appear relatively easy to use, it gives rise to a significant issue: demonstrating to the user the proper way to employ tangible objects when interacting with the system. Collaborative AR interfaces involve the use of multiple displays to facilitate both remote and co-located activities. In co-located scenarios, 3D interfaces enhance physical collaboration within a collaborative workspace. On the other hand, remote sharing leverages AR to seamlessly integrate multiple devices across different locations, enhancing teleconferencing experiences. Hybrid interfaces integrate a set of distinct yet complementary interfaces, offering the flexibility to interact through various interaction devices [19]. This approach creates a versatile platform for spontaneous, everyday interactions, where the specific type of interaction displays to be used is not predetermined in advance. Multimodal interfaces merge input from physical objects with naturally occurring modes of communication, including speech, touch, natural hand gestures, and gaze. This type of interaction is progressing rapidly and is expected to become increasingly favoured in future AR applications [25]. Multimodal interfaces offer a compelling combination of robustness, efficiency, expressiveness, and mobility, making them highly suitable for human-computer interaction. They align with users' preferred interaction style and can support users in combining different modalities or transitioning between input modes based on the specific task or context.

1.2. Industrial Augmented Reality interfaces for technical documentation

The main goal of AR is to convey information that is directly registered into the real world. In the industrial context, AR allows merging real objects with virtual assets to support complex maintenance, assembly, and disassembly procedures. The term "Industrial Augmented Reality" (IAR) was introduced by Georgel [26] to describe the application of AR technology in industrial settings. IAR helps better comprehend and execute industrial tasks, enhancing workers' performance across the production chain [27], [28]. As a result, the well-established advantages of this technology and its rapid technological advancements have led to an increasing number of companies investing in AR.

Although AR has experienced significant growth, its widespread adoption in real industrial settings remains limited. This can be attributed to the complex demands that IAR applications need to address, as stated by Lorenz et al. [4]. These challenges stem from various factors, including user-related issues, technical constraints, environmental considerations, regulatory requirements, and economic aspects. To overcome these barriers, IAR implementations must meet requirements of reliability, safety, helpfulness, operator acceptance, and adherence to strict rules. However, according to Rolim et al. [13], there is a lack of agreement in the literature regarding the optimal approach for presenting information and delivering instructions to users via AR.

In this regard, Gattullo et al. [8] conducted a comprehensive literature review focusing on visualization methods employed in IAR prototypes for technical instructions. The goal was to identify the most used visual assets and examine their use in IAR interfaces for maintenance, assembly, and disassembly tasks. The following subsections report the main results obtained from this study regarding the characteristics of the visual assets used (Fig. 2).

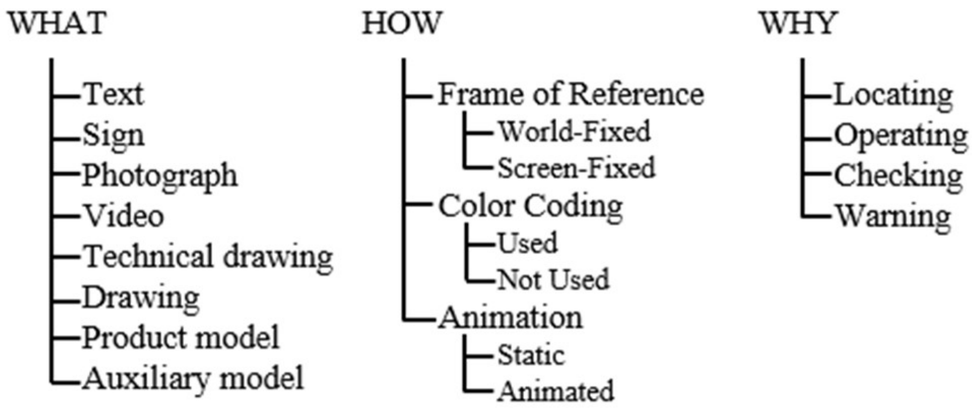


Fig. 2 Proposed classification of visual assets for IAR interfaces (image taken from [8]*).*

1.2.1. Visual asset classification

An initial distinction was made based on the different approaches in authoring needed for each visual asset. Fig. 3 shows examples of using different visual assets in the same industrial context.

TEXT. Traditionally, text has been used to communicate verbal information. Authoring text is a straightforward process, as it only requires defining the content of the text. In this category, the authors include 2D and 3D text, presented within bounding boxes or without them.

PHOTOGRAPH. This category includes assets comprising content generated through photographs captured by a camera depicting real-world objects or scenes. Photographs are prevalent in various forms of manuals, particularly digital manuals, and instructional websites.

VIDEO. This class focuses on assets that involve content generated through video recordings captured by a video camera or webcam, depicting real-world situations or events.

SIGN. Signs can be classified into 3 types: icons, indices, and symbols. Following Peirce's definition [29]: "a sign is a thing which serves to convey knowledge of some other thing, which it is said to stand for or represent."

AUXILIARY MODEL. Following the definition of Wang et al. [30]: "auxiliary models are virtual models for auxiliary instructions." Auxiliary Models (AMs), such as 2D and 3D annotations, are employed by technical authors to provide additional guidance to operators. These annotations direct the operator's visual attention toward specific details or elements. Examples of AMs include arrows, circles, and abstract sketches.

DRAWINGS. This category includes all digitized 2D drawings that do not adhere to formal standards. This includes various examples such as freehand sketches, maps, and charts. Additionally, annotated photographs are also included in this group. They refer to combining a photograph and additional 2D AMs and/or text.

TECHNICAL DRAWING. Creating technical drawings can be more challenging as they need to adhere to international standards to effectively convey constructive and functional information about products (ISO 128-1:2003 [31]). This category encompasses 2D representations in the form of technical drawings presented as static images on a canvas and 3D graphical annotations following ASME Y14.41 – 2003 [32].

PRODUCT MODEL. Following the definition of Wang et al. [30]: "product models are 3D virtual models of products and parts". Therefore, product models refer to the digital representations of real-world objects such as machinery parts, components, and tools. The authoring of product models is commonly accomplished using 3D CAD and 3D modeling tools like SolidWorks, CATIA, and Blender.

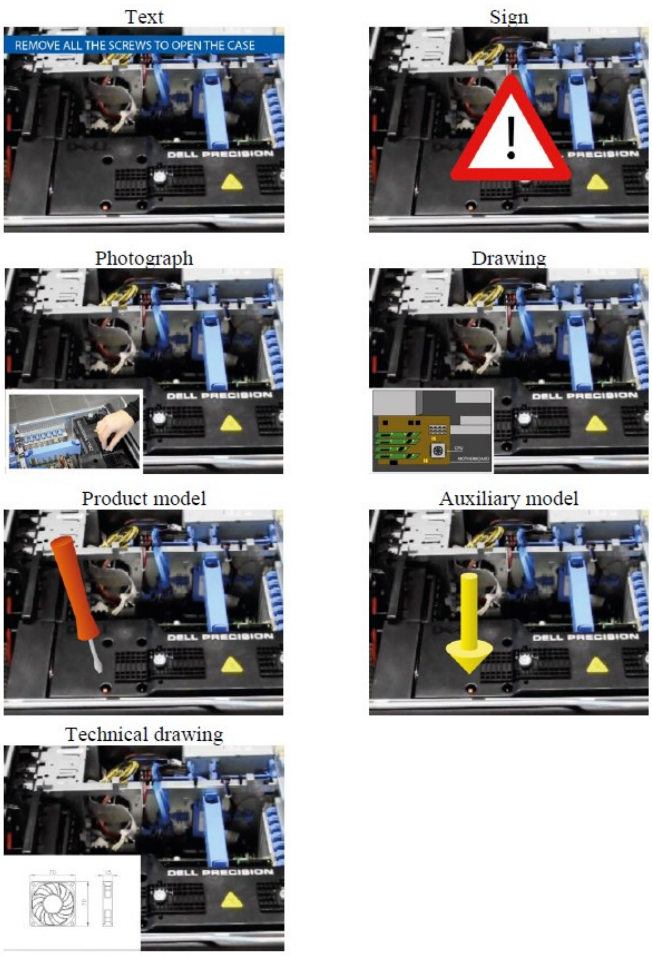


Fig. 3 Classification of visual assets according to different approaches in authoring for the same industrial scenario (image taken by [9]*]).*

1.2.2. Visual asset properties

A second distinction of visual assets can be made according to their properties that can be assigned during the creation of the AR scene to provide additional information.

FRAME OF REFERENCE. Following the definition of Gabbard et al. [33], two types of frames of reference can be distinguished from the user's viewpoint:

- Screen-fixed AR interfaces refer to visual elements that are rendered at a fixed position on the display screen without being spatially anchored to real objects in the scene.
- World-fixed AR interfaces are rendered in a way that they appear to exist at precise locations in the real world.

COLOR CODING. In the industrial context, color usage is subject to international standards and internal guidelines. Consequently, it is possible to differentiate visual assets based on color associations. Some visual assets employ purposeful color semantics, where specific colors convey intended meanings or information according to established standards. On the other hand, there are visual assets whose color choices may appear arbitrary or lack a specific semantic meaning, as determined by the figure itself, its associated caption, and accompanying text.

ANIMATION. Animated visual assets can be differentiated from static ones based on their dynamic nature, as animation provides additional directional or temporal information to users. Animated visual assets can change position, rotation, or scale within the interface over time while maintaining a fixed point of view aligned with the real world.

1.2.3. Information conveyed by visual assets

The last distinction of visual assets refers to the relationship between the type of visual asset employed in an IAR interface and the information it conveys.

LOCATING. Localizing a real object is an important application of visual assets in IAR interfaces. Visual elements can help users in identifying and finding specific parts or components. Locating is a supporting task, as it does not involve direct actions or changes in the system status. Rather, it is a prerequisite for subsequent task actions, ensuring that users can accurately identify and locate the relevant objects before proceeding with their intended actions.

OPERATING. Operating tasks encompass all actions the user performs, whether assisted by tools or not, that result in a change in the system state. These tasks involve actively interacting with the physical or virtual elements manipulating objects, adjusting settings, performing assembly or disassembly actions, or any other task that leads to a tangible change in the system state.

CHECKING. Checking tasks involve thoroughly examining an object or a specific aspect of the scene to make a decision or assessment. These tasks typically involve evaluating the condition of an object, detecting any potential issues or anomalies, or verifying the presence or absence of certain characteristics or elements. Unlike operating tasks, checking tasks do not involve subsequent operations or actions.

WARNING. Safety is very important in an industrial environment, and this consideration also extends to IAR interfaces. It is essential to incorporate special warnings highlighting potential hazards or conditions that demand heightened attention. These warnings are crucial for preventing injuries and mitigating risks that could compromise the operator's health and safety.

1.3. Augmented Reality technical documentation authoring

In the industrial context, many studies on AR have primarily focused on evaluating the concept of AR itself rather than specifically examining the information content it provides [12], [34]. In fact, the literature extensively discusses the challenge of appropriately selecting visual assets for guiding operators in AR [7], [11], [35], [36]. However, research studies often fail to justify their choices of visualization methods. When they do, the focus is mainly on technical aspects such as occlusion, geometric consistency, and computational cost [37]. Major companies like Microsoft [38], Google [39], and Apple [40] have offered guidelines for creating user-friendly AR interfaces using their software libraries. However, these guidelines are provided for generic AR systems and lack specific insights into designing visual assets that effectively convey information from work instructions, as highlighted by Tainaka et al. [7].

Most of the AR authoring systems discussed in the literature typically offer a selection of visual assets based on technical or empirical considerations. The final decision on which visual assets to use is left to the end-users, who choose based on their preferences. For instance, Tainaka et al. [7] suggested a decomposition of assembly tasks into subtasks with a list of candidate visual assets for end-users to pick from. Similarly, Chu et al. [41] conducted a study that associated visual asset proposals with the assembly steps of Dougong Chinese architecture. However, both studies employed an empirical approach that lacks scalability to other industrial applications, as it involves challenges in defining assembly steps and selecting appropriate visual assets.

Other research studies have shown that end-users do not always have the final say based on their preferences. Instead, context-aware AR systems, such as those proposed in [42], [43], recommend the most suitable visual assets based on user capabilities. Syberfeldt et al. [12] and Holm et al. [44] have developed a framework called ARES (Augmented Reality Expert System), which adjusts the level of AR information displayed to operators based on their learning progress when performing new tasks. Similarly, Wolfartsberger et al. [45] used assets with varying levels of detail in working instructions, depending on the operator's qualification level. However, these authors did not explain why a specific asset was chosen over others.

Geng et al. [35] suggest a hybrid authoring approach considering user preferences and capabilities. This approach involves creating a semantic structure for an industrial procedure, breaking it down into a sequence of tasks, each comprising specific actions. To convey virtual information, indicators are used for each action. Similarly, Scheffer et al. [46] introduce an adaptive architectural framework for everyday maintenance. This framework combines expertise with user preferences to enhance overall performance efficiency.

Stork and Schubo [10] introduced an alternative approach to creating AR interfaces by designing predetermined visual assets. This approach was based on the analysis of task-related information, which is also influenced by the complexity of the task itself. This model for assembling tasks served as the theoretical basis in the study of Yang et al. [47], which also examined the joining subtask. In their research, Yang et al. proposed using AR assistance by highlighting the target part with a black rectangle during the commissioning subtask. Conversely, a 3D model of the part to be assembled for the joining subtask was displayed in the assembly zone. Radkowski et al. [11] further extended this investigation, considering the impact of task difficulty. They hypothesized that the complexity of the visual feature should align with the difficulty level of the assembly step.

Thus, the various techniques authors use for the visual assets choice in the TD authoring can be summarized as follows:

- Both technical considerations and end-user preferences [7], [41].
- End-user performance [12], [44], [45].
- Both end-user preferences and performance [35], [46].
- The information type involved [10], [47].
- Task complexity [11], [47].

In contrast to previous studies, this dissertation introduces a novel authoring approach which is designed to propose through AR visual assets only the minimal amount of information required to complete a task, considering various factors such as the real objects involved, the end-user, and the complexity of the task. A similar approach was previously explored by Wang et al. [30], where they suggested a selection of visual assets to reduce the data quantity transmitted between a local technician and a remote expert during AR-based collaborative maintenance. However, the authors did not clearly explain their choices in visual assets, and their proposed framework lacked validation through a user study.

1.4. Objectives and scope

While several methods have been proposed to generate visual assets [37], [48], [49] for industrial AR interfaces [50]–[56], it is evident that there is no agreement in the literature on the best approach for delivering AR instructions [13]. The review of industrial AR interfaces [8] highlighted an interesting finding. For simple tasks like "locating," the most used visual assets are AMs, which offer limited information and use simple virtual shapes (e.g., circles and arrows) to indicate the 3D position of real objects. On the other hand, for more complex tasks like "operating," product models that provide more detailed information, including the shape of the virtual object and its orientation relative to real objects, are predominantly used. One possible explanation for this difference is that additional information does not necessarily enhance operator performance in simple tasks and may lead to information overload, as observed in previous research [11]. This is just an example that focuses on two categories of visual assets. Still, we can generalize by stating that when a substantial amount of information is presented, it tends to create a cluttered presentation. Consequently, extracting meaningful insights from the data becomes challenging for a viewer [57]. However, existing literature does not provide a clear understanding of how various designs of visual assets impact the operator's performance in accomplishing industrial tasks [11], [47].

To fill this gap, in this dissertation we initially propose a replicable method to decompose instructions into elemental information associated with specific visual assets. Thus, we formulated the following research question: "*What is the elemental information contained in work instructions?*". To address it, we developed an information model inspired by the "Therbligs" method [59] described in the following Chapter.

Afterward, we formulated the subsequent research question: " *Which virtual content is recommended for each information type?*". In response, we developed an information model based on Norman's definitions of affordance and signifier [58]. This model establishes the complexity of a task by considering the amount and type of information required to complete the tasks included in a work instruction. This information can be conveyed through object affordance and/or AR signifiers, which involve one or more visual assets with their respective properties. Subsequently, we introduce a new authoring approach named "minimal AR," which focuses on providing the minimum information through AR visual assets to accomplish a task.

Finally, we applied our authoring approach in a real industrial context in which selecting appropriate visual assets for ARTD is a complex process. This complexity arises because AR depends on many factors that must be considered simultaneously, including the surroundings, component availability, worker technology acceptance, lighting conditions, and blind areas. Therefore, a modeling issue occurs. In fact, once the most suitable visual asset for a specific information type has been identified by applying the "minimal AR" authoring approach (e.g., an AM for a locating task), it is necessary to understand its optimized properties (shape, frame of reference, color coding, and animation) according to the working context. Then, for each information type, we formulated the last research question of this dissertation: "*What are the optimized properties of the recommended virtual content?*"

Chapter 2. Decomposition of work instructions[2]

In the previous Chapter, a classification of visual assets with their properties and a summary description of the information they can transfer was provided through the literature review conducted by Gattullo et al. [8]. As anticipated, a significant gap in the existing literature concerns the absence of clear guidelines on three key aspects: (i) the decomposition of work instructions into elemental information, (ii) the effective transmission of information through visual assets in ARTD, and (iii) the optimization of their properties according to the working context. This dissertation intends to fill these gaps.

This Chapter addresses the first issue: how to decompose instructions into elemental information. After presenting the proposed information model, a preliminary study is described to determine if it is effectively possible to decompose instructions into the proposed elemental information.

2.1. Therbligs method

Whether in the form of printed manuals or Electronic Work Instructions [35], TTD consists of a simple collection of work instructions with large texts or complex images that are difficult to relate directly to the real products being worked on. In contrast, ARTD allows work instructions to be precisely positioned on real products, using tailored visual assets [3] such as CAD models, drawings, and videos to convey information more concisely and targeted. While this new form of TD can potentially improve operator performance, it also presents new challenges in the authoring process.

The literature lacks consensus regarding the parameters for analyzing work instructions and selecting appropriate visual assets for elemental information. Some studies propose visual assets with varying levels of information detail, considering factors such as the operator's experience or preference [7], [41]. Other studies decompose assembly tasks into subtasks or basic operations without a standardized method [10], [36]. Task complexity is also a controversial point some studies address [60], [61]. However, task complexity is influenced by numerous factors, such as

[2] The results of the study described in this Section were published in the following paper:
M. Gattullo, E. Laviola, and A. E. A. E. Uva, "From Therbligs to Visual Assets: a Technique to Convey Work Instructions in Augmented Reality Technical Documentation," in Advances on Mechanics, Design Engineering and Manufacturing IV, Springer, Cham, 2022, pp. 1327–1339. doi: 10.1007/978-3-031-15928-2_116.

the number and shapes of the components to be assembled, making it challenging to establish a definitive definition [10]. In this study, we have introduced a technique to convert TTD into ARTD. The goal is to address the following research questions: "*What is the elemental information contained in work instructions?*"

2.2. Analysis of work instructions

To address our research question, we developed an information model that analyzes work instructions by separating the information they contain. Our approach was inspired by the "Therbligs" method [59], which Oyekan et al. [62] suggested as a suitable framework for describing any tasks.

The Therblig method invented and refined by Frank and Lillian Gilbreth in the early 1900s, provides a framework for describing the various motions involved in performing a task [59]. The main goal of this method is to identify unnecessary or fatigue-inducing motions associated with different activities. In [59], a comprehensive list of the eighteen Therbligs serves as descriptors for any given task. These Therbligs include actions such as searching, finding, selecting, grasping, holding, transporting (loaded or empty), positioning, assembling, using, disassembling, inspecting, pre-positioning, releasing load, experiencing unavoidable delays, experiencing avoidable delays, planning, and resting.

Oyekan et al. [62] applied the Therblig method in their study, specifically focusing on fitting instructions for assembling a 'flat-pack' table. They extended the method by introducing the concept of a Therblig model, which suggests that humans examine the environment and perform actions based on the "information to start" and "information to end" associated with each Therblig. For every Therblig, they identified the specific "information to start" and "information to end" required. For instance, in the case of the Therblig "search," operators receive information regarding the "identity of the searched part" and start the action of "looking around." The action concludes once they receive information about the "location of the searched part."

The model proposed in [62] appears to be a viable approach for analyzing work instructions within an assembly manual. Consequently, in an ARTD, each resulting "information to start" and "information to end" could be effectively conveyed through appropriate visual assets. However, our observations indicate that many of the information elements listed in [62] exhibit similarities, even if they belong to different Therbligs. As a result, it is possible to use the same visual asset to convey these similar information elements. For example, the Therblig "search" has as information to start the "identity of the search part," while the Therblig "use” has as information to start the "identity of the object to use." However, both could be generalized as the "identity of an object" and conveyed using the same type of visual asset.

In our study, we expanded upon the model introduced by Oyekan et al. [62] by organizing the "information to start" and "information to end" associated with different Therbligs into distinct clusters. This led us to identify six comprehensive and distinct classes of information called "Information Types:" identity, location, way to, notification, order, and orientation. We illustrated the outcome of this clustering process in Table 1. Afterward, we proposed to analyze work instructions, performing the following steps:

1. decomposing work instructions into single small tasks
2. identifying the specific information required for each task from Table 1
3. associating each information element to one of the six information types.

Table 1 Result of clustering the Therblig and relative information (information to end has a gray background to distinguish from information to start) into information types.

IDENTITY			
SEARCH	Identity of the searched part	ASSEMBL.	Identities of the Parts to assemble
FIND	Recognizable characteristic of the searched part	FIND	Matched characteristic of a part in the search area
TR. LOADED	Identity of the part to move	DISASSEM.	Identities of the Parts to disassemble
SELECT	Criteria of selection	INSPECT	Identity of the object
GRASP	Identity of the object to grasp	PREPOS.	Identity of the part to position
HOLD	Identity of the object to hold	INSPECT	predetermined standard
USE	Identity of the object to use	REL. LOAD	object to release
POSITION	Identity of the part to position		
LOCATION			
TR. LOADED	Location of the part	TR. LOADED	Destination of the part
SELECT	Location of the part that matches the criteria	TR. EMPTY	Destination of the hand
		SEARCH	Location of the searched part
WAY TO			
GRASP	The way to grasp it	USE	The way to use it
ASSEMBL.	The way to assemble them	DISASSEM.	The way to disassemble them
NOTIFICATION			
GRASP	Confirmation that the object is grasped	REL. LOAD	Confirmation that the object has been released
TR. LOADED	Confirmation that the part has reached its destination	INSPECT	Result of the comparison between the object and the standard
TR. EMPTY	Confirmation that the hand has reached its destination	POSITION	Confirmation that the part is oriented as desired
UN. DELAY	Problem occurring: shortage, delay	AV. DELAY	Problem occurring
ASSEMBL.	Confirmation that the parts are assembled	UN. DELAY	Information that the problem has been solved
USE	Confirmation that the use motion is finished	AV. DELAY	Information that the problem has been solved
DISASSEM.	Confirmation that the parts are disassembled	PREPOSIT.	Confirmation that the part is oriented as desired
REST	Need for the worker to rest	REST	End of the resting period

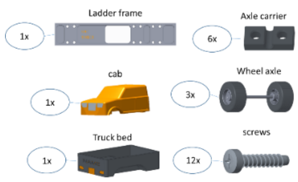
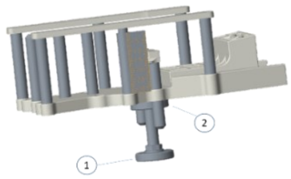
ORDER			
HOLD	<u>Order</u> to hold	PLAN	Structured <u>sequence</u> defining how things are going to happen
HOLD	<u>Order</u> to do the next motion		
ORIENTATION			
POSITION	Initial <u>orientation</u> of the part	PREPOSIT.	Initial <u>orientation</u> of the part
POSITION	Desired <u>orientation</u>	PREPOSIT.	Desired <u>orientation</u>

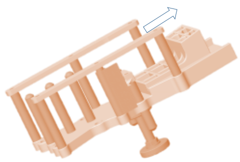
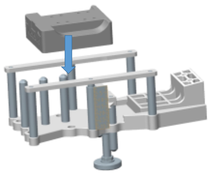
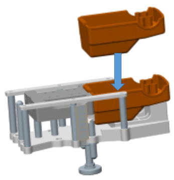
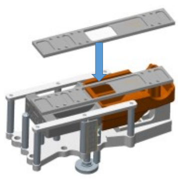
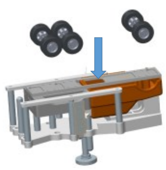
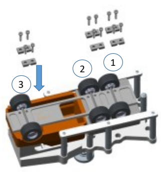
2.3. Material and Methods

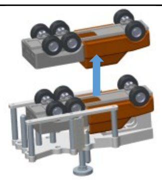
We demonstrated the application of our proposed information model using a case study involving assembly instructions for a "model pick-up truck." The entire production cycle of the pick-up truck was simulated in the C-Factory, a laboratory at the University of Applied Sciences Würzburg-Schweinfurt, which replicates a smart factory environment on a smaller scale. The assembly instructions for the pick-up truck contained all six information types identified in our model.

The original manual consisted of nine instructions, presented to the operator in the C-Factory through a Microsoft PowerPoint document containing text and drawings. Each instruction focused on a single elemental task. We analyzed these instructions, extracting the required information set using the six information types proposed in our work. For the analysis in Table 2, we excluded information that did not necessitate a visual asset, such as notifications indicating that an object was successfully assembled. Such information is directly acquired by the operator while performing the task.

Table 2 Result of the analysis of the work instructions contained in the case study of the pick-up truck manual.

Step	Work instruction	Information types
1	*To start the assembly you should have the following components:*  TASK: <u>"check to have the components"</u>	IDENTITY of the ladder frame IDENTITY of the truck bed IDENTITY of the 3 wheel axles IDENTITY of the cab IDENTITY of the 6 axle carriers IDENTITY of the 12 screws
2	*Please fit the assembly device to your pick-up size.* *Adjust the platform support to 17.5mm by turning the lower knurled nut 1.* *Clamp the position by turning the upper knurled nut 2.*  TASK: <u>"fit the assembly device to the pick-up size"</u>	IDENTITY of the lower knurled nut 1 LOCATION of the lower knurled nut 1 WAY TO adjust the platform support (by turning the lower knurled nut 1) IDENTITY of the height of 17.5 mm (predetermined standard) NOTIFICATION that the height is reached IDENTITY of the upper knurled nut 2 LOCATION of the upper knurled nut 2 WAY TO clamp the position (by turning the upper knurled nut 2)

		NOTIFICATION that the assembly device is clamped
3	*Move the slide into the displayed position.*  TASK: <u>“move the slide”</u>	IDENTITY of the slide LOCATION of the slide (initial position) WAY TO move the slide (translation) LOCATION of the slide (final position) NOTIFICATION that the final position is reached
4	*Insert your truck bed into the assembly device as shown.*  TASK: <u>“insert the truck bed”</u>	IDENTITY: the truck bed LOCATION of the truck bed on the assembly device ORIENTATION of the truck bed WAY TO insert the truck bed
5	*Place the cab into the displayed position.*  TASK: <u>“place the cab”</u>	IDENTITY: the cab LOCATION of the cab on the assembly device ORIENTATION of the cab WAY TO place the cab
6	*Place the ladder frame onto the cab and truck bed with the inscription facing upwards.* *Make sure to align bores.*  TASK: <u>“place the ladder frame”</u>	IDENTITY of the ladder frame LOCATION of the ladder frame on the cab and truck bed ORIENTATION of the ladder frame (inscription upwards) WAY TO place the ladder frame NOTIFICATION: inspect that bores are aligned
7	*Add the axles by inserting the tires into the wheel-arches provided.*  TASK: <u>“add the axles”</u>	IDENTITY of the axles LOCATION of the (three) wheel-arches WAY TO add the axles
8	*Position the axle carriers on the axles.* *Tighten all screws on the truck, starting from rear to front.*  TASK: <u>“position the axle carriers”</u>	IDENTITY of the screws LOCATION of the (twelve) screws on the axle carrier WAY TO tighten the screws ORDER of tightening (from rear to front)
9	*You completed the assembly of your pick-up truck.* *Please remove your truck from the assembly device.*	NOTIFICATION: the assembly is completed IDENTITY of the truck WAY TO remove the truck



TASK: “remove the truck from the assembly device”

2.4. Lessons learned

By proposing an information model based on the decomposition of work instructions, we have effectively addressed our first research question (“*What is the elemental information contained in work instructions?*”). Interestingly, these six information types (identity, location, order, way to, notification, and orientation) proved sufficient for analyzing all the instructions within the TTD that needed to be converted. Thus, our model revealed to be a replicable method to decompose instructions into elemental information to be associated with specific visual assets.

In this way, it was possible to provide the theoretical basis for developing the "minimal AR" authoring approach that focuses on providing the minimum information through AR visual assets to accomplish a task.

Chapter 3. The "minimal AR" authoring approach[3]

The authoring approach presented in this dissertation could orient designers in selecting appropriate visual assets for ARTD. The main goal is to provide only the information needed to accomplish a task through visual assets. To achieve this, it is important to establish an information model that clarifies what information is necessary and how it can be conveyed using visual assets.

3.1. The information model

Creating manufacturing work instructions involves their decomposition into individual, simple tasks, as explained in [6]. For each elemental task, it is possible to determine the set of information needed. Work instructions can encompass various types of information, and there is no universally standardized classification for them. However, a previous study [9] examined ten manufacturing manuals and identified six distinct information types used in these instructions:

- IDENTITY: e.g., the identity of a part to position, of an object to grasp.
- LOCATION: e.g., location and destination of an object.
- ORDER: e.g., order to do a motion or a structured sequence in a plan.
- WAY TO: e.g., the way to dis/assemble parts, to use an object.
- NOTIFICATION: e.g., confirmation that parts are dis/assembled or information that a problem is occurring/ has been solved.
- ORIENTATION: e.g., the initial or desired orientation of an object.

[3] The results of the study described in this Section were published in the following papers:
- E. Laviola, M. Gattullo, V. M. Manghisi, M. Fiorentino, and A. E. Uva, "Minimal AR: visual asset optimization for the authoring of augmented reality work instructions in manufacturing," Int. J. Adv. Manuf. Technol., 2021, doi: 10.1007/s00170-021-08449-6.
- M. Gattullo, E. Laviola, A. Evangelista, M. Fiorentino, and A. E. Uva, "Towards the Evaluation of Augmented Reality in the Metaverse: Information Presentation Modes," Int. J. Interact. Des. Manuf., 2022, doi: https://doi.org/10.3390/app122412600.

Each elemental task could have a variable complexity depending on the specific information required and its amount. As a result, designers have to individually decide on the most appropriate way to present this information to users in the AR interface. It can be provided either by exploiting the affordance of the objects involved in the task or through an AR signifier, i.e., one or more visual assets with their properties (Fig. 4).

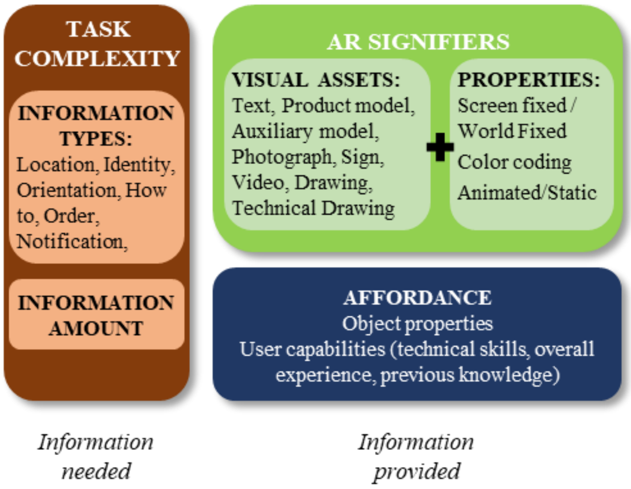


Fig. 4 The information model used in minimal AR: the task complexity is defined by the amount and type of information needed [9]*; they can be provided by exploiting affordance* [58] *and AR signifiers, i.e., visual assets with their properties* [8].

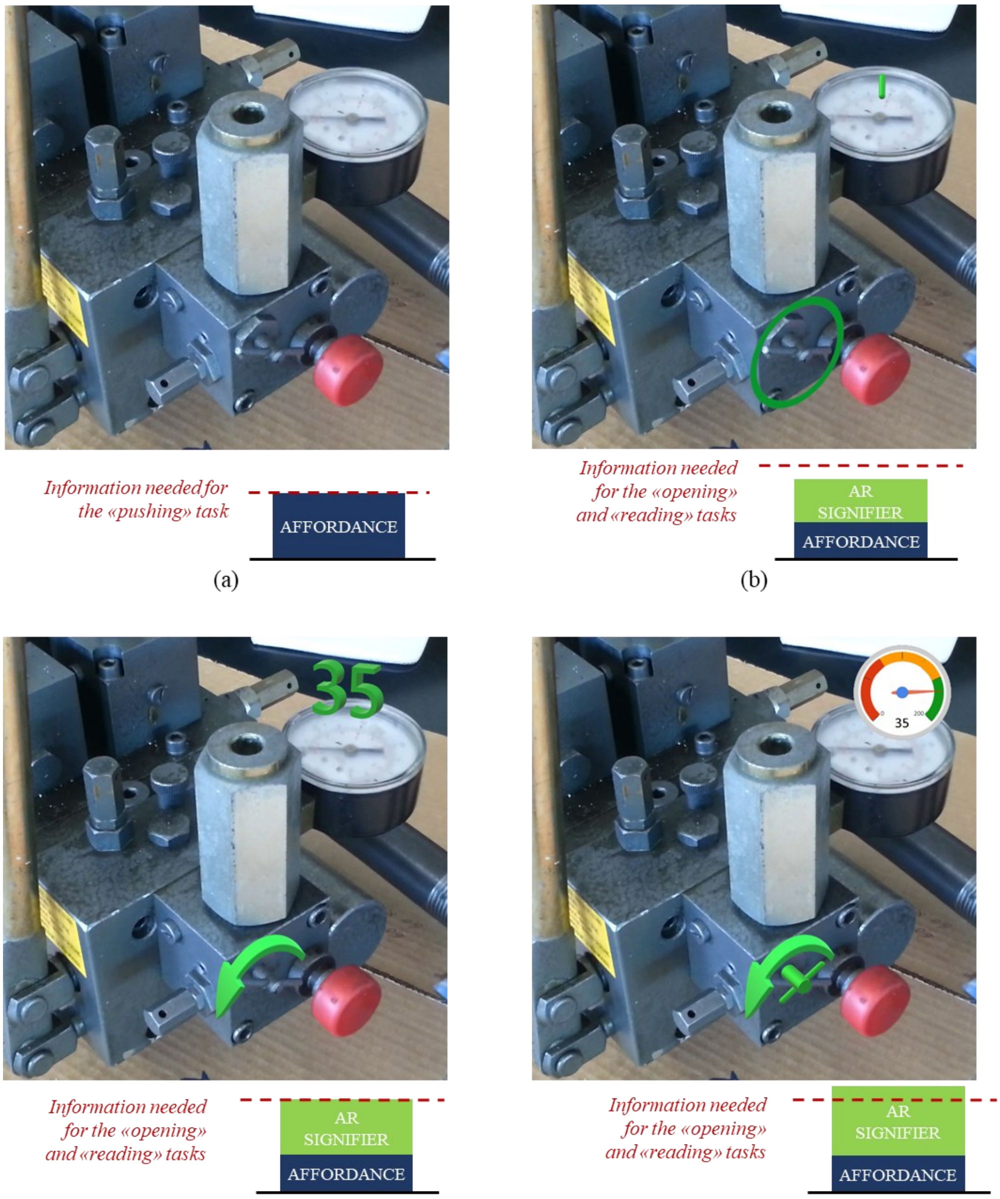
According to Norman's concept of affordance, it refers to the relationship between the properties of an object and the abilities of a person, which determines the potential ways in which the object can be used [58]. Measuring affordance lacks a standardized method, necessitating an analysis of both the object properties and the capabilities of the individuals involved in the task. When an object exhibits high affordance, it implies that its design naturally communicates how to use it effectively. An example is the panic bar in emergency doors, which intuitively indicates how to open the door during an emergency. In cases where an object affordance allows for multiple ways of use, users can rely on their skills and capabilities to complete the task. These skills may vary among individuals based on their attitudes, overall experience in similar tasks, and previous knowledge relevant to that specific task. On the other hand, if an object affordance is low, external sources of information become necessary. In this context, Norman introduced the concept of a "signifier," which refers to any perceivable indicator, a mark or sound, that effectively communicates the appropriate behaviour to the user.

An AR signifier refers to a combination of visual assets that are used to convey a set of information. These visual assets are categorized as in the literature survey on industrial AR interfaces [8], which includes text, signs, photographs, drawings, videos, product models, and AMs. Each type of visual asset has specific design properties, such as its frame of reference (screen/world fixed), color coding, and any associated animations. These design properties allow each visual asset to communicate one or more information. For instance, a world-fixed visual asset provides information about its location with respect to the same screen-fixed visual asset. Color coding can be used to deliver notifications (e.g., warnings in yellow) or to indicate the identity of parts that need to be assembled (e.g., a pin and its corresponding hole). Animations are useful for presenting dynamic information, like demonstrating how to orient objects during an assembly, explaining specific tasks, or showing the order in which multiple parts should be assembled.

3.2. The minimal AR signifier

When creating ARTD, designers face a crucial decision of selecting the appropriate AR signifiers that convey the necessary information to operators to complete a task successfully. This decision holds great significance for operators, as they may encounter three distinct scenarios:

1) Incomplete information: the AR signifier is not well designed because the information provided is not clear or incomplete (Fig. 5b).
2) Minimal information: the AR signifier is tailored to the information required to accomplish the task (Fig. 5a, c).
3) Redundant information: the AR signifier is not optimized because it allows for either redundant information or information not needed to accomplish the task (Fig. 5d).



(a) (b)

(c) Minimal AR (d)

Fig. 5 The minimal AR authoring approach applied to the case study of a hydraulic valve. For the instruction "push the manual descent button," the affordance is enough to accomplish the task, then no AR signifier is needed (a). For the instruction "open regulator #7 until the nominal pressure is read on the pressure gauge," affordance is not enough, and AR signifiers are needed: information provided through signifiers can be incomplete or not clear (b), minimal (c), redundant (d).

The advancement in Norman's theory is significantly evident in the presentation of previous scenarios, primarily enabled by the possibilities offered by AR for interface design. Unlike TTD, which was limited to visual assets suitable for printing, such as text and drawings, AR interfaces open a new realm of opportunities. This allows designers to create interfaces more tailored to the operator's knowledge and experience. In fact, for experts, the traditional documentation could be overwhelming with redundant information. At the same time, it might be unclear or lacking essential details for novices, leaving them to rely solely on their own experiences. On the contrary, AR interfaces present a unique advantage as they permit the careful design of visual assets. This enables the creation of optimal conditions, where information is presented in a more intuitive and contextually relevant manner, benefiting both experienced operators and novices.

In the “minimal AR” authoring approach, the focus is on creating AR signifiers aimed to minimize the difference between the information they provide and that needed to accomplish a task. We called them "minimal AR signifiers." The underlying idea behind this approach, which was drawn from existing literature [8], [11] and observations from industrial projects involving local companies, is as follows: compared to the minimal AR signifier, user performance (completion time, mental workload, errors) does not improve with those conveying additional information.

3.3. Application to industrial case studies

To better understand the information model and how to define minimal AR signifiers, we used our authoring approach on two typical industrial tasks commonly found in TD: pressing a button and adjusting a regulator. These tasks were extracted from an authentic technical manual for a hydraulic valve used in elevators, focusing on pressure regulation for the stem.

The first instruction given is to "push the manual descent button." This instruction consists of a single elemental task, which is to press a button. To successfully perform this task, three essential information types are required: (i) the *identity* of the button, (ii) its *location*, and (iii) the *way to* push it. In this case, the affordance of the valve is sufficient to convey all the necessary information, and no AR signifier is needed (Fig. 5a). In fact, the manual descent button is the only button present on the valve, making it easy for even a novice operator to recognize and locate it. Moreover, the unique mushroom-shaped design of the button provides a high level of affordance, meaning that users naturally understand how to interact with it. Consequently, adding any further AR signifiers, such as an animated 3D arrow demonstrating how to push it, would be unnecessary and redundant.

The second instruction is “open regulator #7 until the nominal pressure is read on the pressure gauge.” It can be divided into two parts: opening the regulator and reading the pressure value. To carry out each elemental task, we can use the proposed information model to identify the specific set of information required. Additionally, we can explore whether this information can be provided through affordance or signifiers.

To begin the task, the operator requires specific information: (i) the *identity* of the regulator, (ii) its *location*, and (iii) the *way to* open it. This information is accessible through affordance but only if the operator has prior experience completing this task. The valve has multiple regulators that can be turned either clockwise or counterclockwise. Novice operators may struggle to identify the correct regulator and direction (counterclockwise) for the opening action. To address this issue, AR signifiers are necessary. One example of a minimal AR signifier for this set of information could be a static counterclockwise arrow placed near the physical regulator (Fig. 5c). A simple circle would not be enough because it lacks information about the opening action (Fig. 5b), leading some operators to guess the direction and potentially make mistakes. On the other hand, using animated arrows or additional AR signifiers would result in redundant information (Fig. 5d).

In the reading task, certain information is required, which includes: (i) the *identity* of the pressure gauge, (ii) its *location*, (iii) the *identity* of the nominal pressure value, and (iv) the *notification* when the pressure reaches the nominal value. In this scenario, when using an AR signifier, the only information that needs to be provided is the identity of the nominal pressure value. This is because there is only one pressure gauge, which is easily locatable even for inexperienced operators. The nominal pressure value (35 bar in this case) can be presented using various visual assets. One option for a minimal AR signifier is to display a text showing the nominal pressure value (Fig. 5c). While it might be appealing to use a drawing of the pressure gauge to indicate the nominal pressure value, it becomes redundant and obstructs the real-world view (Fig. 5d). Another approach is to use a simple line that highlights the position to be reached by the needle on the actual pressure gauge, which is less obstructive (Fig. 5b). However, this solution may not provide clear information, as tracking errors could cause the line to indicate an incorrect number on the pressure gauge.

Chapter 4. Visual asset association based on the information to convey

This Chapter addresses the second issue of this dissertation: how to associate visual assets based on the information to convey. After presenting our "minimal AR" authoring approach, an exploratory study conducted involving a focus group of ARTD experts is described to extract initial considerations of possible associations of visual assets based on information types. Then, a second user study is presented to show an application of our authoring approach in establishing these associations.

4.1. Exploratory study

4.1.1. Methods

We set up a focus group of ten people (2 females, 24 to 48 years old, mean = 35.7, SD = 9.42). The participants were selected based on their expertise in the design of ARTD. First, we presented the six information types using examples of instructions extracted from real manuals, including the case study we conducted. This allowed us to illustrate the practical application of the information types. Subsequently, we introduced a list of the eight different types of visual assets with their properties that can be used to convey information in IAR interfaces following that proposed in [8] described in Chapter 1.

Based on their experience in authoring AR interfaces, the group members discussed the most suitable visual assets for each information type. During the discussions, participants were asked to propose their preferred visual asset and its associated properties. They needed to justify their choice without revealing it to other group members until everyone made their own selection.

Then, the two most preferred visual assets (or more, in the case of a tie) were revealed. Each participant was then asked to decide between the presented options. By following this approach, the visual asset with the highest number of votes for each information type was determined as the recommendation made by the focus group.

4.1.2. Results

This section provides an overview of the key points covered in the discussion and summarizes the proposed visual asset for each information type. The examples provided in the figures highlight the proposed visual asset using a yellow sketch, making it distinguishable from the CAD model of the assembly device used to simulate the real object unavailable due to COVID-19.

IDENTITY. Based on the feedback of the focus group, the most suited visual asset for conveying this information type is a screen-fixed drawing that is static and uses the colors of the real object to be identified (Fig. 6). The participants said that a world-fixed visual asset was unnecessary for this information type, leading them to exclude the option of using a product model, which also requires access to the CAD model. An alternative to the drawing could have been a photograph; however, the group recognized that obtaining photographs of real objects might not always be feasible due to safety or privacy policies. Furthermore, the participants noted that a drawing allows the inclusion of additional information, such as text, to convey the identity of specific measurements (e.g., tightening torque). In addition, the dynamic content provided by a video was not seen as advantageous for this information type. Although text is commonly used in TTD to convey identity, the focus group believed that a drawing offers a clearer understanding of the component, reducing the cognitive load on operators. Lastly, the focus group members reached a consensus that certain visual assets were not suitable for conveying the information type IDENTITY. Signs and technical drawings were considered to provide only limited information, while AMs were deemed relevant only when associated with a real object.

Fig. 6 Example of displaying the information type IDENTITY through a drawing (highlighted in yellow) screen-fixed, static, with the colors of the real truck bed.

LOCATION. The focus group identified the AM as the most suitable visual asset for conveying this information type. They recommended a static world-fixed AM without color coding (Fig. 7). The participants agreed that color coding and animation did not significantly enhance the communication of this information. However, using a world-fixed visual asset was deemed essential for effectively highlighting the location in the real world. As a result, the focus group excluded visual assets that may occlude a significant portion of the real world, such as photographs, drawings, technical drawings, and videos. Regarding the product model, users acknowledged that it could potentially convey more information than just the location, which could lead to information overload. Similar considerations to those expressed for the identity information type were made for text and signs.

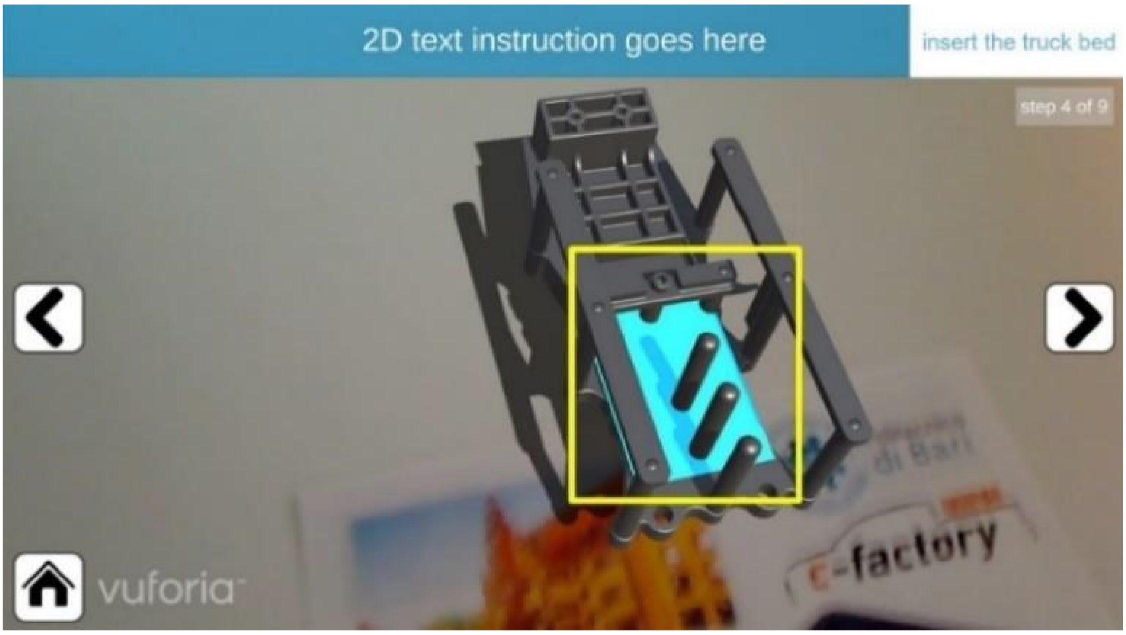


Fig. 7 Example of displaying the information type LOCATION through an AM (highlighted in yellow) world-fixed, without color coding and static.

WAY TO. To convey this information type, the focus group determined that a world-fixed product model without color coding was the most suitable visual asset (Fig. 8). Using animations was considered particularly effective in providing a preview of the operation. By watching the animation, operators could easily replicate the steps they observed. As a result, visual assets such as drawings, technical drawings, photographs, signs, and texts, which could not produce this interactive preview, were rejected for conveying this information. While a video could provide an operation preview, the focus group preferred CAD models due to their ability to precisely align with the real components that need to be handled. Finally, the focus group concluded that an animated AM would not be suitable for conveying this information type, particularly for complex tasks that require reproducing the real components.

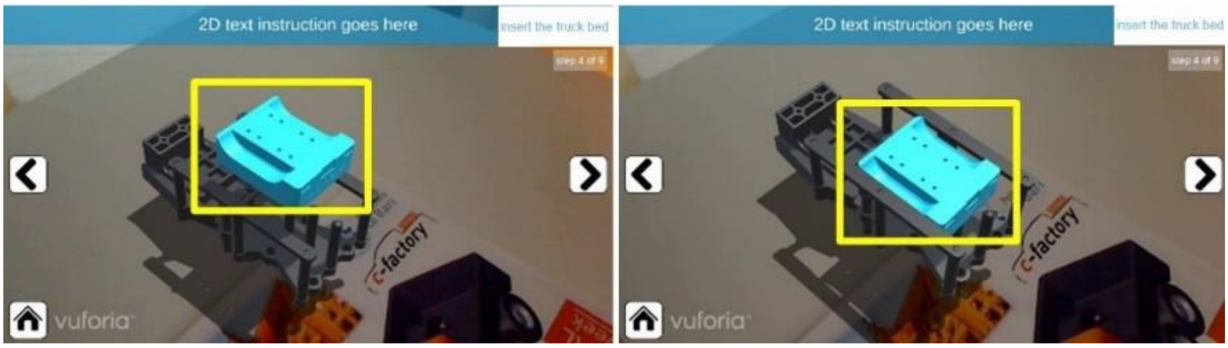


Fig. 8 Example of displaying the information type WAY TO through the animation of a world-fixed product model (highlighted in yellow) without color coding; on the left, it is displayed the initial position of the animation, while on the right, it is displayed the final position.

NOTIFICATION. This information type posed a challenge for the focus group to agree. Participants argued that a visual asset is unnecessary to convey this information type in many cases, as operators can autonomously perceive if an operation was executed correctly. For example, in step 2 of the proposed case study, the notification that "the assembly device is clamped" does not require a visual asset, as the operator's senses can discern it. However, there are scenarios where notification may benefit from a visual representation. For instance, in step 6 of the proposed case study, the instruction "make sure to align bores" requires a clear notification. In such cases, the focus group determined that a screen-fixed 2D text without color coding and animation is enough (Fig. 9).

Fig. 9 Example of displaying the information type NOTIFICATION through a text (highlighted in yellow) screen-fixed, without a color coding and static.

ORDER. To specify this information type, which refers, for example, to an assembly sequence, the focus group suggested using a world-fixed text, either in 2D or 3D format. The text should not incorporate color coding and should be animated in a manner that ensures it remains in front of the device camera, even if the camera moves in relation to the real world (Fig. 10). However, the focus group expressed that this information is often closely linked to the WAY TO in work instructions. In such cases, the animation of the product models may implicitly convey the order information as well. Therefore, including text as a separate visual asset for specifying the order information might be redundant. As a result, the second most favored visual asset for conveying the ORDER was the AM, specifically in the form of animated arrows pointing from one real component to the next in the assembly sequence. Nevertheless, the focus group acknowledged that the effectiveness of the AM depends on the proximity of the components. If two components are situated far apart, the information conveyed by the AM might be challenging to comprehend.

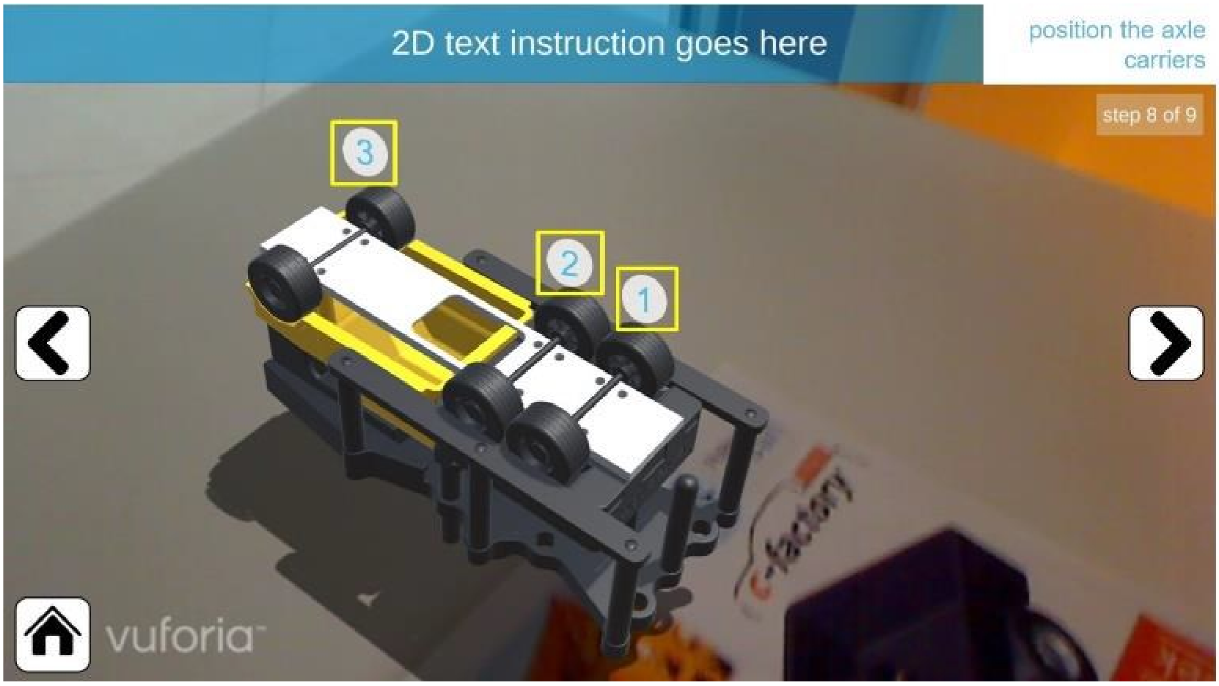


Fig. 10 Example of displaying the information type ORDER through an animated world-fixed text (highlighted in yellow) without a color coding.

ORIENTATION. To convey this information type, the focus group determined that a static world-fixed product model without color coding was the optimal solution (Fig. 11). It provides a preview of the desired orientation a component should have during the assembly process. The focus group members expressed similar considerations to those made for the WAY TO. However, it was important to note that the two information types could not be merged, as the WAY TO encompasses more details, such as the specific way to grasp or use a particular tool.

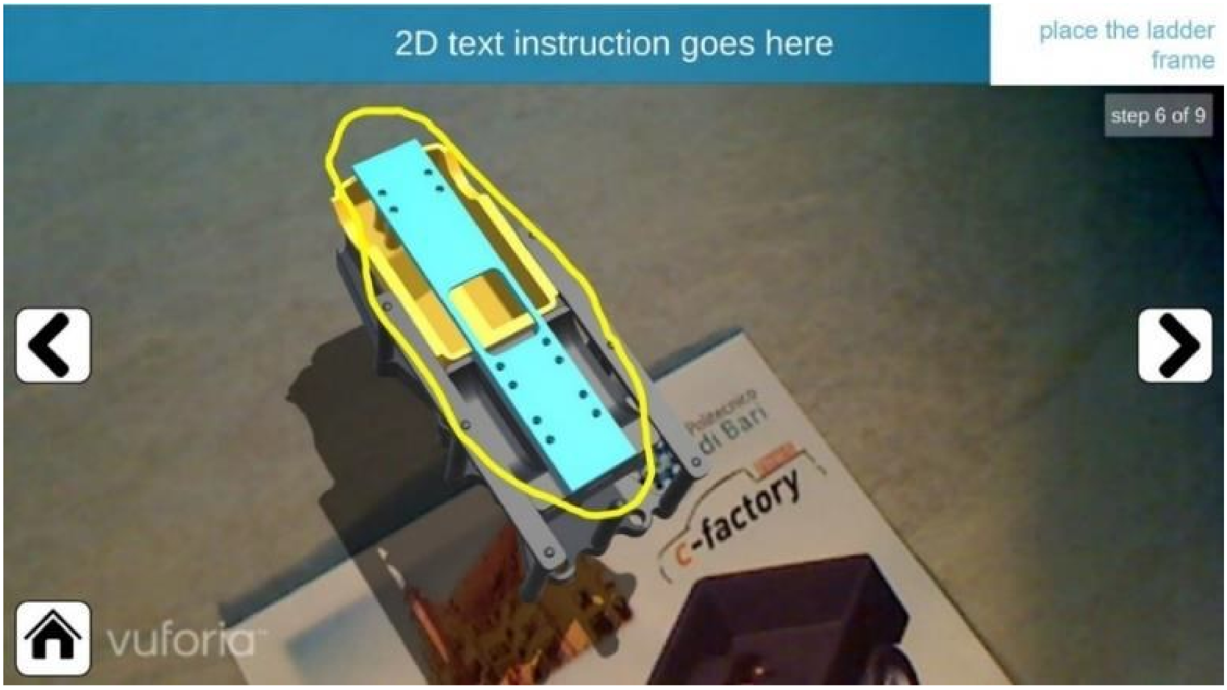


Fig. 11 Example of displaying the information type ORIENTATION through a static world-fixed product model (highlighted in yellow) without a color coding.

4.1.3. Lessons learned

To answer our second research question ("*Which virtual content is recommended for each information type?*"), we engaged a focus group of experts in the design of ARTD. Through their insights, we could restrict the possible choices of virtual content. The focus group was asked to propose a candidate visual asset for each of the six information types, and their recommendations are summarized in Table 3.

Table 3 Visual assets proposed by the focus group for the six information types.

Information type	Visual asset	Properties		
		Frame of reference	**Color coding**	**Animation**
Identity	Drawing	Screen-fixed	Yes	No
Location	Auxiliary model	World-fixed	No	No
Way-to	Product model	World-fixed	No	Yes
Notification	Text	Screen-fixed	No	No
Order	Text	World-fixed	No	Yes
Orientation	Product model	World-fixed	No	Yes

Our study revealed that product models could not be generalized to all the instructions in an ARTD. While previous research has highlighted that AR interfaces based on product models are the most common [8], the most engaging [9], and reduce error rate [61], the literature also identified disadvantages associated with their use [11], [63]. In fact, in some cases, a product model may convey excessive information, surpassing what is necessary for correctly executing a task, as stated in our work [64]. This implies that alternative visual assets should be considered for certain information types.

Our study did not specifically address the combination of visual assets within a complete work instruction, as we focused primarily on analyzing individual information types and their corresponding visual assets. In addition, it should also be considered that not all information necessarily requires using a visual asset. In some cases, the information can be obtained through object affordance or the operator's experience and familiarity with the task, as pointed out in our work [64].

4.2. Application of the "minimal AR" authoring approach: a user study

In the following subsections, by detailing our user study, we can grasp how the "minimal AR" authoring approach can effectively support the conveyance of AR information. Specifically, our user study focuses on a common manufacturing task, such as assembling two parts.

4.2.1. Design of the experiment

The user study aims to test our hypothesis: compared to the minimal AR signifier, user performance (completion time, mental workload, errors) does not improve with those conveying additional information. To conduct this experiment, a LEGO-based assembly task was designed. Participants were tasked with assembling abstract shapes using LEGO Duplo bricks. Prior research [65]–[68] has shown that LEGO-based tasks effectively simulate industrial assembly tasks due to the similarities in interactions, such as picking and placing objects. The advantage of using LEGO bricks is the ease of modifying various task variables, including quantity, order, and shape. This allows for a gradual increase in complexity according to industrial manufacturing requirements. LEGO bricks were chosen for this experiment because they are small, lightweight, and safe, making them convenient for laboratory use, as noted by Yang et al. [47].

Furthermore, using LEGO bricks for the assembly ensured that participants had prior experience handling them, as most users were familiar. This fixed affordance meant that participants knew how to hold and manipulate the LEGO bricks, unlike complex industrial equipment assembly, which heavily relies on specific human capabilities. Additionally, to ensure unbiased results, participants were unaware of the specific task details and were instructed to assemble different abstract shapes for each condition tested.

To define the tasks, we considered a gradual increase in the amount of information. Our study focused on categorizing the different information types described in [9]. The required information needed for a LEGO-based assembly task includes the following:

- The brick destination *location*.
- The *identity* of the brick to pick, in terms of color and shape.
- The *orientation* of the brick along the axis normal to the assembly plate (this information is provided through the brick affordance along the other axes).
- The sequence *order* of the bricks in the assembly.

Afterward, four LEGO sets were created, each becoming progressively more complex. The varying levels of information required were achieved by using one to three different LEGO brick shapes and incorporating one to two colors (Fig. 12). In each LEGO set, there were a total of twelve bricks, and they were evenly distributed among the various types.

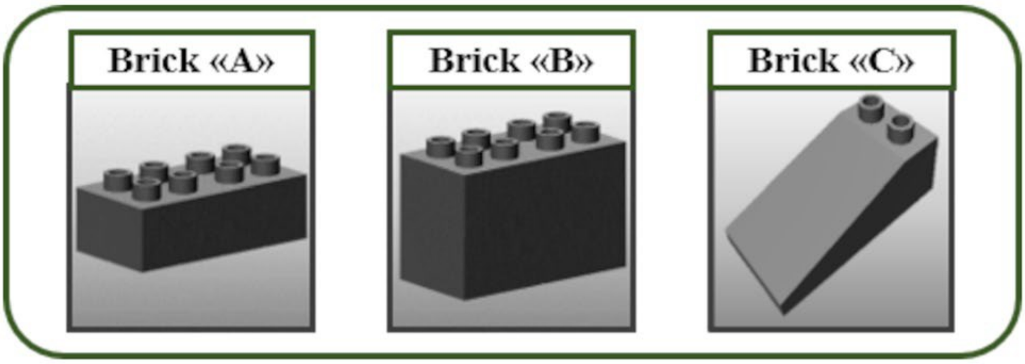
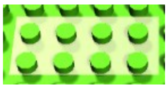
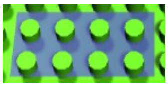
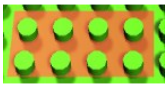


Fig. 12 LEGO Duplo brick shapes used in the experiment; colors used for each brick are blue and red.

We applied the "minimal AR" authoring approach to create AR signifiers for each LEGO set. This involved combining one or more visual assets to convey a set of minimal information. Four AR signifiers were designed using this approach (Table 4).

Table 4 AR signifiers designed for the experiment.

Signifier ID	Visual assets	Properties	Appearance	Information provided
1	auxiliary model	World-fixed; no color coding used; static;		Location
2	auxiliary model	World-fixed; color is the same of the brick to place; static;		Location Identity (color)
3	auxiliary model	World-fixed; color is the same of the brick to place; static;		Location Identity (color and shape)
	drawing	Screen-fixed; Color and shape are the same of the brick to place; static;		
4	product model	World-fixed; Color and shape are the same of the brick to place; animated;		Location Identity (color and shape) Orientation

The study did not dictate a specific selection of visual assets to convey the same information, allowing designers to choose according to their preferences or technical considerations. For example, if different visual assets were preferred for AMs instead of the filled rectangle used in the experiment, they were free to explore such alternatives. To evaluate the effectiveness of the design of AR signifiers, their performance was measured against an established baseline. It followed the traditional method of presenting LEGO brick assembly instructions, as described in [69]. This method includes drawings of exploded isometric views for each assembly step and a target drawing of the completed assembly. In the drawings, numbered balloons are used to provide the sequence of steps (Fig. 13). Conversely, the AR interface only displays instructions for one brick at a time during the assembly process.

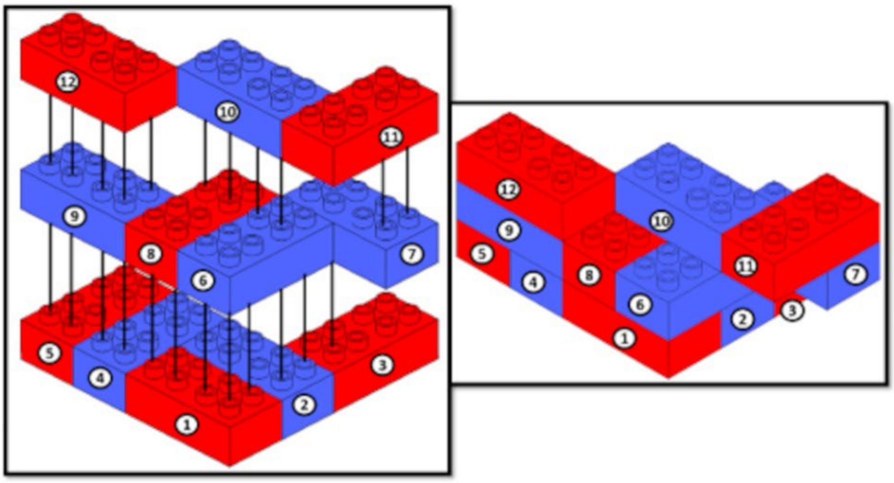


Fig. 13 Reproduction of a traditional signifier employed in LEGO manuals [69]*, used as baseline in the experiment.*

In this study, we examined the influence of independent variables on 4 LEGO sets, specifically looking at different signifiers (4 AR and 1 traditional used as a baseline). This led to a total of 20 different conditions (4 LEGO sets × 5 signifiers). For each LEGO set, 5 tests were created, each with a different signifier. However, we only conducted tests for 14 of these conditions (Table 5). This exclusion was because we disregarded conditions where the information provided by the AR signifier was insufficient to proceed. For instance, in the case of LEGO set 2, we excluded the condition with AR signifier 1 because users could not know if a blue or red LEGO brick had to be placed. We designed a within-subject experiment where each participant tested all the experimental conditions to gather data on user performance regarding completion time, mental workload, and errors.

Table 5 LEGO sets and signifiers tested in the experiment.

LEGO SET	**1**	**2**	**3**	**4**
Information needed	Location	Location Identity (color)	Location Identity (color) Identity (shape)	Location Identity (color) Identity (shape) Orientation
LEGO bricks used (Quantity)	A red (12)	A red (6) A blue (6)	A red (3) A blue (3) B red (3) B blue (3)	A red (2) A blue (2) B red (2) B blue (2) C red (2) C blue (2)
Minimal AR signifier	Signifier 1	Signifier 2	Signifier 3	Signifier 4
Other signifiers tested	Traditional Signifier 2 Signifier 3 Signifier 4	Traditional Signifier 3 Signifier 4	Traditional Signifier 4	Traditional

Regarding the mental workload evaluation, a dual-task paradigm was employed. Participants were asked to perform a secondary task, which in this study was a simple visual monitoring task. Their objective was to react promptly to a change in color, a technique previously used by Brunken et al. [70]. The rationale behind including an additional task is based on the concept that the capacity not used for the primary task can be directed to handle another task, as pointed out by Cegarra et al. [71]. Consequently, the performance in the secondary task serves as an indicator of the demands imposed by the primary task. The decision was made not to use workload assessment tools like the NASA Task Load Index (TLX) or SWAT (Subjective Assessment Technique). These tools would have required significantly more time to conduct the entire experiment for each participant due to the substantial number of conditions that needed to be tested.

4.2.2. The simulated AR interface

The experiments were conducted in a simulated AR environment. This technology, also known as indirect AR [72] or immersive virtual AR [73], has been widely used as a design tool in various fields like architecture, city planning, and industrial design [74]–[76]. Before the main experiment, a preliminary evaluation was conducted to compare the simulated AR environment with the true physical task. The evaluation involved a focus group of ten individuals (2 females, 24 to 48 years, mean = 35.7, SD = 9.42) with experience creating AR interfaces. The focus group discussion revealed minimal differences in how information was conveyed between the two modalities without affecting user performance. However, they noticed some distinctions in how users interacted with LEGO bricks in the simulated AR environment. To mimic the interaction between users' hands/eyes and physical LEGO bricks, we used interaction metaphors inspired by other user interfaces like Virtual Reality (VR) testing applications [77] and video game GUIs. These interaction metaphors covered actions such as (i) picking the requested brick, (ii) rotating a brick before placing it, (iii) assembling bricks on the plate or top of each other, and (iv) disassembling a brick that was wrongly placed. Additionally, (v) the point of view from which users observed the assembly was considered. The focus group agreed that these differences in interaction had little impact on the information conveyed by the AR signifiers, which were the focus of the study. However, they pointed out that being unable to change the point of view in the simulated AR environment could lead to some instances where information might remain hidden. In response to this feedback, the researchers decided to include a slider in the interface, allowing users to rotate the virtual LEGO plate along the vertical axis, thus giving them more control over their viewpoint.

The simulated AR application was created using Unity 3D Engine, importing 3D CAD models of LEGO bricks. These models were originally designed with Autodesk Inventor and then exported in the .obj file format. Within the Unity 3D environment, the interface for the simulated AR application was established, and specific scenes were set up for 14 different testing conditions. Users could place virtual LEGO bricks on a virtual 26x26 green LEGO Duplo plate, enabling them to construct various assemblies. A transparency shade was used to distinguish between the virtual LEGO bricks and the AR instructions. The virtual LEGO bricks were displayed without transparency (Fig. 14(1)). On the other hand, auxiliary and product models, which were part of the AR instructions, were rendered with a semi-transparent shade (alpha 150 on a scale from 0 to 255) (Fig. 14(2)), similar to previous works [47], [66], [78]. The simulation did not include virtual models of the picking bins, where the LEGO bricks would be stored. This decision was based on the approach taken in other studies [76], [78], aiming to keep the user's focus solely on the relevant information for analysis. Instead of showing the picking bins, pictures of the available LEGO bricks for a particular LEGO set were placed in the top-left corner of the interface (Fig. 14(3)).

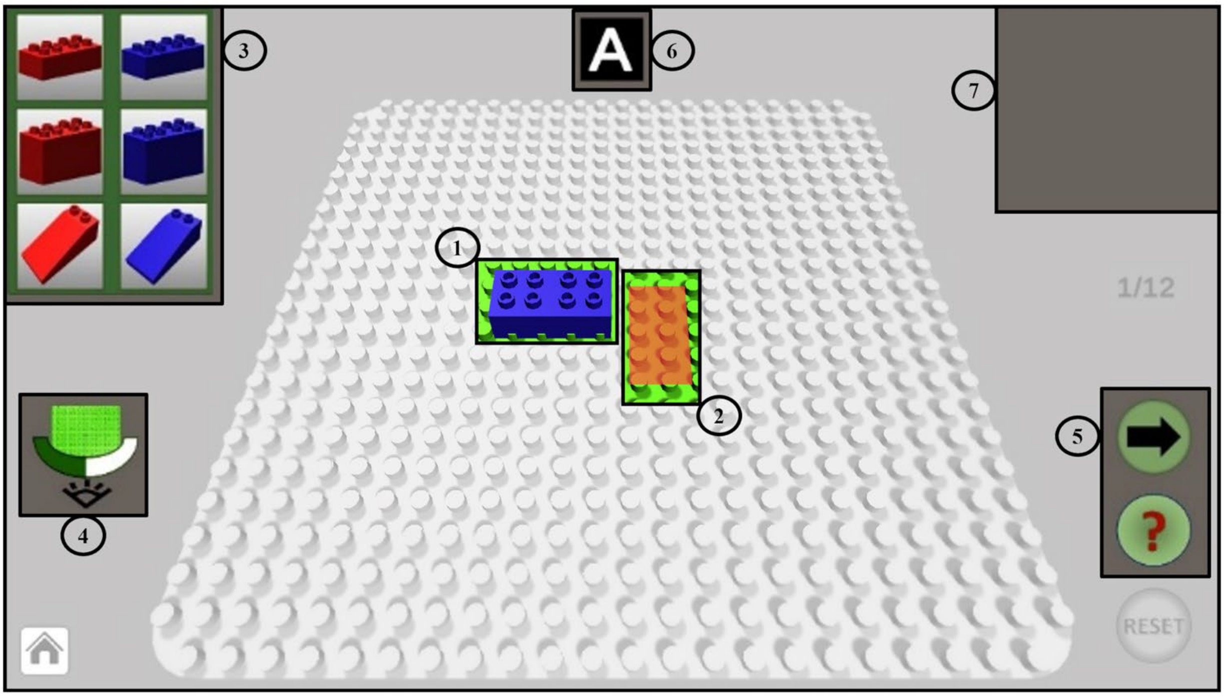


Fig. 14 Interface of the simulated AR application: a LEGO brick already placed (1), instruction provided using a world-fixed AM (2), GUI buttons with pictures of the LEGO bricks available for the assembly set (3), a slider to rotate the camera (4), GUI buttons to move to the following assembly step and to display the popup window with instructions (5), changing color letter for the dual task (6), the GUI area for the screen-fixed visual assets (7).

Users selected a brick in this system by clicking on its corresponding picture. Once chosen, a 3D preview of the brick appeared at the position of the mouse cursor on the GUI. To place the brick, users released it either on the green LEGO Duplo plate or on top of a previously placed LEGO brick. Additionally, they could rotate the brick around its vertical axis, clockwise or counterclockwise, using the "D" and "A" keys on the keyboard, respectively. To enhance the user experience, a slider was implemented on the GUI, allowing users to rotate the virtual camera along the vertical axis (Fig. 14(4)). This feature allowed them to change their point of view and find a more comfortable position for placing LEGO bricks within the virtual scene. The interface remained similar in the tests involving traditional signifiers, but the visual assets were omitted. Instead, users received assembly instructions displayed as drawings on their smartphones, placed next to the main screen where the application was running. All these design choices addressed technical challenges related to the software physical simulation, such as the unpredictability of virtual object behaviour, potential collisions, and attaching issues [67].

4.2.3. Procedure

The user study involved conducting 14 trials using the simulated AR application. In each session, the experimenter gathered general information about the participants and provided them with information about the test process and instructions on how to use the application. Subsequently, the users started a training session with the application to familiarize themselves with its features, which could last 1 to 5 minutes. During the initial training, a popup window was shown, explaining all the available interactions with the interface, and this window remained accessible throughout the trials if needed (Fig. 14(5)). The training scenario included examples of each AR signifier encountered during the experiment. Once the users felt adequately trained, the experimenter allowed them to initiate the experiment by specifying the trial ID to execute at various intervals.

A balanced Latin Square design was used for the task complexity and the signifier variable to reduce unwanted order effects, such as learning and fatigue, and minimize the carry-over effect that could impact the data collected. As a result of this design, there were 40 unique sequences in which the 14 trials were executed.

The main assembly task in each trial can be summarized as follows:

- Users click a button to start the trial.
- Users perform the assembly task, selecting and placing the twelve LEGO bricks according to the instructions.
- When users placed a LEGO brick, if the step was accomplished without errors, they could move to the following step by clicking a button in the GUI (Fig. 14(5)). Otherwise, the

experimenter said a mistake was made, and users had to repeat the step. An “undo” button allowed them to delete the last brick wrongly placed.

- At the end of the twelve steps, users pushed a button to end the trial.

Users were asked to perform an additional task to measure their mental workload during the main assembly task. Periodically, a signal was given in the form of a color change of a letter (from blue to white) [70] located at the top center of the interface (Fig. 14(6)). The users' objective was to respond to this signal as quickly as possible by pressing the "S" key on the keyboard. Once the users responded, the color of the letter would revert to blue, and the software recorded the time it took for them to react. The timing for the color changes was controlled by a predetermined sequence, with intervals between two changes varying from 8 to 18 seconds. As a result, multiple measurements of reaction times were collected within each experimental condition.

Data were collected for each trial during the experiment, including completion time, reaction time in the secondary task, and errors. The application automatically recorded times. Completion time was measured from when users initiated the trial by pressing a button to when they finished it by pressing another button. As for the reaction time in the secondary task, the average time from multiple measurements was calculated for each trial. Errors were manually identified by the experimenters, who also categorized the type of error, such as order, location, orientation, color, shape, or combinations of these. For each user, the total number of errors in each trial was collected for the analysis.

Forty individuals participated voluntarily in the study (9 females, 12 to 49 years old, mean = 24.6, SD = 5.08). Most of these participants were recruited from the local university, including students and staff members. To ensure consistency, all participants were confirmed not to be colorblind and had prior experience with LEGO assembly, averaging a familiarity rating of 4.1 (SD = 0.93, median = 4, min = 2, max = 5) on a 5-point Likert scale (1: Not at all familiar - 5: Extremely familiar). The users were evenly distributed regarding their familiarity with AR applications, with an average rating of 2.7 (SD = 1.48, median = 3, min = 1, max = 5). Each participant did an experiment that lasted approximately 40 minutes on average. The experimenters used the Microsoft Teams platform to communicate individually with each user. During the sessions, participants were requested to share their screens so the experimenters could observe the entire test.

4.2.4. Results

We collected data from each participant, including their completion time, mean reaction time, and error, which were classified as matched continuous variables. This is because we obtained data for each user under all 14 experimental conditions. We conducted separate analyses for different complexity levels (LEGO sets) to test our hypothesis. Specifically, we only compared the minimal and other AR signifiers for LEGO sets from 1 to 3.

The Shapiro-Wilk normality test (using the AS R94 algorithm) showed that the original data did not follow a normal distribution. However, the data was transformed into a normal distribution by applying a Box-Cox transformation with $\lambda = -0.5$ to completion time and reaction time. These normalized samples were then used for repeated measures ANOVA to compare more than 3 samples. Before conducting the ANOVA, the assumption of sphericity was checked using Mauchly's Test of Sphericity. If the assumption was violated, the Greenhouse-Geisser correction was applied to the ANOVA results. Post hoc tests were performed using the Bonferroni correction method. The Box-Cox transformation did not yield normal distributions for total errors. Therefore, nonparametric tests were used instead. In this case, the Friedman 2-way ANOVA was applied to compare more than 3 samples. The Wilcoxon ranks-sum test was used as a post hoc test, and adjustments were made using the Bonferroni correction method.

For all the LEGO sets, it was observed that the mean completion time differed statistically significantly among the conditions (Table 6). Further analysis using post hoc tests showed that AR signifiers did not lead to any statistically significant reduction in completion time compared to the signifier conveying minimal information. Instead, there was a significant increase in completion time when using signifiers 2 (22%) and 4 (14%) in LEGO set 1 (Fig. 15). This result supported our hypothesis regarding completion times for all LEGO sets. Additionally, the post hoc tests indicated that the traditional signifier caused a significant increase in completion time compared to all the AR signifiers ($p < 0.001$ for all pairs). This finding confirmed the effectiveness of all the proposed AR signifiers.

Table 6 Results of statistical analyses for completion time.

LEGO set	Repeated Measures ANOVA	Post hoc: Minimal Signifier vs.			
		Signifier 2	Signifier 3	Signifier 4	Traditional Signifier
1	F(3.29, 128.317)=91.109; p<0.001	p=0.003*	p=0.498	p<0.001*	p<0.001*
2	F(2.578, 100.541)=166.121; p<0.001	-	p=0.506	p=0.639	p<0.001*
3	F(2, 78)=277.962, p<0.001	-	-	p>0.999	p<0.001*

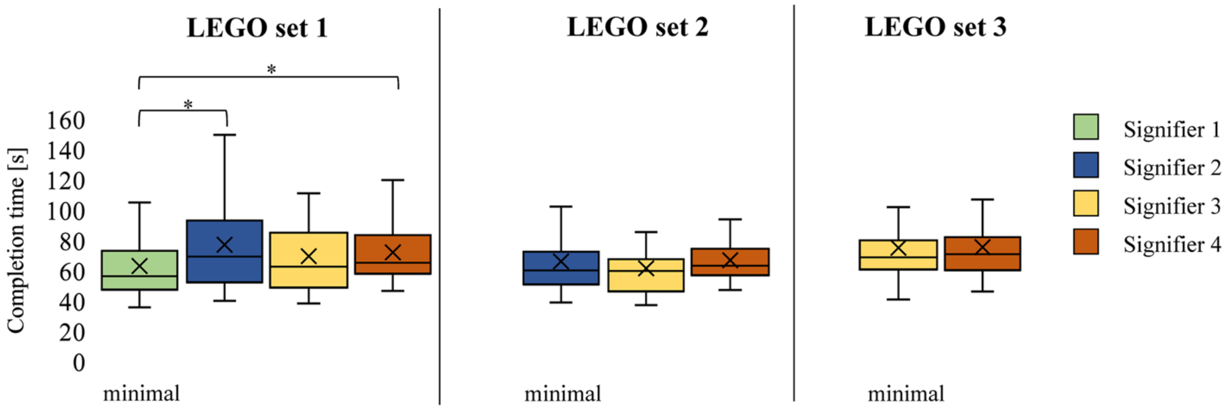


*Fig. 15 Task completion time (error bar: standard error, *: significant difference) for the three LEGO sets for which there is a comparison between the minimal and other AR signifiers.*

For all the LEGO sets, we found significant differences in the mean reaction time of dual tasks across all conditions (Table 7). Post hoc tests revealed no statistically significant decrease in reaction time when using any of the AR signifiers compared to the one with minimal information (Fig. 16). This outcome supported our initial hypothesis for all the LEGO sets, including users' mental workload. Furthermore, the post hoc tests also demonstrated that the traditional signifier significantly increased reaction time compared to all the AR signifiers (with $p < 0.001$ for all pairs). This finding confirmed the effectiveness and value of all the proposed AR signifiers.

Table 7 Results of statistical analyses for mental workload.

LEGO set	Repeated Measures ANOVA	Post hoc: Minimal Signifier vs.			
		Signifier 2	Signifier 3	Signifier 4	Traditional Signifier
1	F(4, 156)=23.078; p<0.001	p>0.999	p>0.999	p>0.999	p<0.001*
2	F(3, 117)=35.104; p<0.001	-	p>0.999	p=0.867	p<0.001*
3	F(1.7, 66.293)=34.009; p<0.001	-	-	p=0.132	p<0.001*

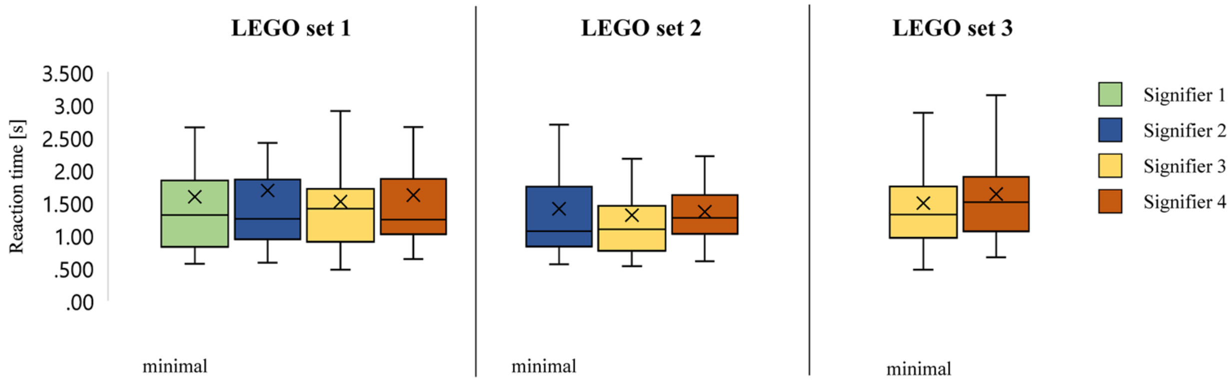


Fig. 16 Reaction time for the three LEGO sets for which there is a comparison between the minimal and other AR signifiers.

For all the LEGO sets, a statistically significant difference was observed in total errors (Table 8). Post hoc tests revealed that most AR signifiers did not result in a statistically significant reduction of errors compared to the one with minimal information, except for LEGO set 3. For this set, using signifier 4 significantly reduced errors (Fig. 17), even though signifier 3 already provided all the necessary information. The initial hypothesis regarding total errors was supported for all LEGO sets, except for set 3. Additionally, when comparing the traditional signifier to the AR signifiers, it was found that the traditional signifier caused a significant increase in errors compared to all AR signifiers ($p < 0.001$ for all pairs), except signifier 2 for LEGO set 1, which still showed a statistically significant increase in errors ($\alpha = 0.005$, $p = 0.028$).

Table 8 Results of statistical analyses for total errors.

LEGO set	Friedman 2-way ANOVA	Post hoc: Minimal Signifier vs.				Post hoc Sign. level
		Signifier 2	Signifier 3	Signifier 4	Traditional Signifier	
1	$\chi2(4)=40.784$; $p<0.001$	p=0.005	p=0.564	p=0.564	p<0.001*	α=0.005
2	$\chi2(3)=36.076$; $p<0.001$	-	p=0.564	p=0.317	p=0.001*	α=0.008
3	$\chi2(2)=36.646$; $p<0.001$	-	-	p=0.002*	p<0.001*	α=0.017

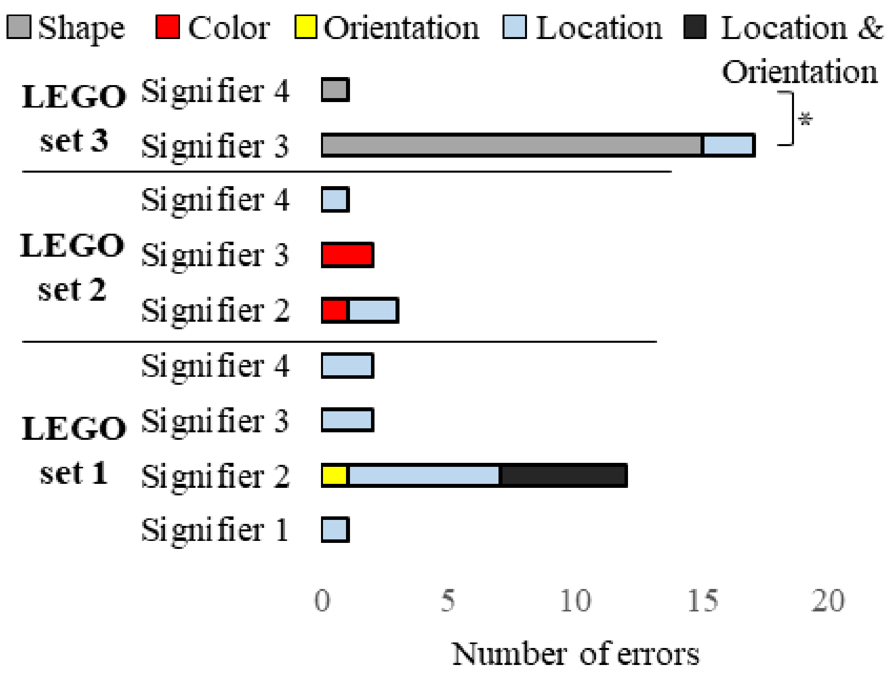


Fig. 17 Total errors (: significant difference) for the three LEGO sets for which there is a comparison between the minimal and other AR signifiers.*

4.3. Findings

The findings from the user study confirm our initial hypothesis that there is no benefit in using AR signifiers that convey more information than the minimal signifier. This pattern was observed across all the measured variables (completion time, mental workload, and errors) regardless of the complexity of the tasks and the increasing amount of information involved. We conducted the study using a LEGO-based assembly task to test the "minimal AR" authoring approach. Still, we believe that the results can also be applied to other manufacturing tasks, as mentioned in subsection 3.3. Then, based on these study results, we propose adopting the "minimal AR" authoring approach as a guideline for creating future ARTD in manufacturing. This approach can streamline the authoring process and prevent excessive efforts by selecting appropriate visual assets and avoiding information overload.

The "minimal AR" authoring approach can be summarized in the following steps:

1. Isolate the instructions to provide in the ARTD.
2. Divide the instruction into elemental tasks.
3. Analyze the elemental task to find the set of information needed to accomplish it.
4. Analyze the operating context (real object, user) to evaluate its affordance.
5. Define which information could be provided through the affordance.
6. Define which information must be provided through AR signifiers.

7. Consider all the AR signifiers that can provide the needed information.
8. Choose the AR signifier that minimizes the difference between the information provided and that needed (minimal AR signifier).

4.4. Design implications

We can extract some design implications from the findings of this study that can help understand how the "minimal AR" authoring approach can be used in various manufacturing tasks found in existing industrial AR research.

When a work instruction only includes the location of an object, it has been observed that employing an AM without specific colors is sufficient. This approach applies to various manufacturing tasks such as disassembly operations [2], inspection tasks [79], remotely supported maintenance [80], and point welding [81]. Henderson and Feiner [82] demonstrated a useful example of using AMs, using a combination of arrows and highlighting effects to indicate the location.

The findings of our experiment on location align with Radkowski et al. [11], who also did not find any significant differences in errors between an AR interface with product models and one with AMs for locating tasks. When considering the application of AR in an industrial application, it is easy to think of a CAD model overlaid on real equipment, as observed in numerous prototypes [26], [83]. This observation is confirmed by Gattullo et al. [9], who discovered that potential AR technical writers would prefer using product models over other visual assets for conveying every information type without considering design issues. Consequently, in many cases, product models are employed to browse real objects within the environment, as seen in [84]. Our study revealed that a product model is only necessary when conveying information about the orientation of an object, as demonstrated in [85]. For simpler information, taking design issues into account, using a product model becomes disadvantageous. This is due to their high sensitivity to registration accuracy, requiring meticulous management of depth cues (such as occlusion, shading, and shadows) to maintain visual coherence with real objects. Additionally, product models demand a higher level of effort for modeling compared to AMs. Furthermore, product models may not always be available, particularly when dealing with old equipment or requiring permissions for their use. Their creation also necessitates expertise in 3D modeling, with the complexity of geometry and animation directly impacting the modeling effort. On the other hand, AMs can be selected from an existing list of models or easily created when needed.

When work instructions become more complex, additional signifiers may be required with AMs to help operators correctly identify the actual objects they refer to. This is particularly evident in assembly operations involving numerous objects [86] or in maintenance tasks requiring specific tools [82]. To address this, it is recommended to use visual assets such as drawings or photographs to support object identification. These visual assets can be complemented with AMs that provide information about the object location. Our research indicates that no other AR signifiers are necessary when the task with a particular object is unambiguous. For instance, if the task involves inserting a pin into a hole, providing an animation of the product model of the pin demonstrating the insertion process would be unnecessary and redundant for the operator.

In some scenarios, the possible choices of products to be assembled are limited. Our study demonstrated that color coding can be effectively used in such cases. By assigning specific meanings to different colors of visual assets, we can convey information about the assembly process. Our experiment used two colors for the AM, resembling the approach taken with the LEGO bricks to insert. In manufacturing applications, various colors could represent different objects involved in a task. For instance, in Uva et al. [2], participants were required to assemble the intake and exhaust camshafts of a motorbike engine. They used colored AMs to distinguish their locations without relying on additional information, such as drawings or photographs, to indicate their shapes. However, there is a limitation to the effectiveness of color coding. If the number of objects placed at a specific point in an assembly becomes too high, the color coding method may lose its efficiency. This is because the operator would need to remember more associations between colors and objects, making managing and processing the information more challenging.

In assembly set 3, an unexpected outcome was noticed when combining a drawing with an AM to represent the minimal AR signifier. Surprisingly, besides the product model does not provide further information, it reduces errors. Specifically, as depicted in Figure 17, 88% of the errors related to the minimal AR signifier were attributed to wrong shapes. Upon analyzing the experiment recordings, it was observed that some users did not pay enough attention to the shape information provided by the drawing in the top-right corner of the interface. A fixed drawing on the screen was employed to minimize any occlusions in the real scene [63]. However, it was found that users mainly relied on the AM, which remained fixed in the world and was directly displayed in the working area. This result is useful for considering an arrangement of visual assets that does not need divided attention. For instance, in future AR interfaces, a situated visualization could be adopted, where all visual assets are conveniently placed around the AM, and the drawing is integrated as a label anchored to the AM itself.

Chapter 5. The "minimal AR" authoring approach in a real maintenance scenario[4]

The previous Chapter describes applying our "minimal AR" authoring approach in two different user studies, obtaining important insights. In particular, the findings confirm our hypothesis that there is no benefit in using AR signifiers that convey more information than the minimal signifier. Therefore, our approach can streamline the ARTD authoring process and prevent excessive efforts by selecting appropriate visual assets and avoiding information overload. However, these considerations were made under controlled experimental conditions and fixed affordance. Therefore, this Chapter aims to demonstrate that the identified guidelines are valid in a real industrial context. A user study has been conducted, including real maintenance workers with the same field experience, thus fixing affordance. They used and evaluated an ARTD designed according to our "minimal AR" authoring approach.

5.1. Worker technology acceptance of a Mixed Reality Technical Documentation

An important factor that must be considered when designing AR instructions in a real industrial context is the worker technology acceptance, meaning how industrial workers may effectively behave and respond when using Mixed Reality Technical Documentation (MRTD), for example, through smart glasses while carrying out real industrial maintenance tasks. Mixed Reality (MR) has emerged as a highly promising technology in the manufacturing industry over the past few decades [87], [88]. Specifically, it offers significant advantages in maintenance tasks by aiding workers in handling intricate, worker-focused processes and reducing cognitive strain through visual information, streamlined document analysis, and improved decision-making during operations [87].

MR technology facilitates smooth and natural interactions between the physical and virtual worlds [89]. It empowers workers to interact with objects using a tangible interface, seamlessly shifting between real and virtual environments. In simpler terms, MR not only lets operators take

[4] The results of the study described in this Section were published in the following paper:
E. Laviola, S. Romano, M. Gattullo, and A. E. Uva, "Evaluating the Worker Technology Acceptance of a Mixed Reality Technical Documentation," in Virtual Reality and Mixed Reality, 2023.

advantage of the benefits offered by AR, which overlays virtual information onto real-world objects to provide visual cues and simplified step-by-step guidance, but it also allows them to customize and interact with a non-referenced AR interface tailored to their individual needs [90].

Despite the proven benefits of MR technology, its usage in real industrial contexts is not yet widespread, even though workers generally acknowledge its advantages [91]. This limited adoption has been attributed to various reasons in previous studies. One major factor is the lack of knowledge about which specific technology is most suitable for certain tasks [92]. Additionally, shortcomings have been identified, such as the absence of optimal features (e.g., content authoring, interaction capabilities, hardware capabilities, and ergonomic design) in the technology used for specific tasks [93]. However, recent rapid technological advancements have addressed these barriers [94]. In fact, wireless commodity smart glasses, like the highly advanced OST HMD Microsoft HoloLens 2 [95], have become commercially available. These smart glasses offer a hands-free work experience, eliminating the need to hold devices like smartphones or tablets. Due to their commercial availability and unique features [90], [94], companies are increasingly interested in adopting them, and reliable guidelines for their implementation are being developed. As a result, the existing literature often focuses on finding the best ways to convey maintenance instructions through MRTD [64], [11] and designing effective MR interfaces to enhance workers' performance [96]. However, one critical aspect often overlooked in the literature is industrial workers' learning skills and acceptance when introduced to this new MR technology.

Although MR technology was introduced several decades ago, its applications for maintenance tasks using smart glasses have mostly been limited to laboratory research. For instance, Aransyah et al. [97] developed a HoloLens application to optimize pipeline maintenance procedures, focusing on improved inspection speed and accuracy. However, their testing was limited to a focus group in a controlled laboratory environment without considering how real operators might react to this technology. Similarly, Schlagowski et al. [98] developed a prototype for analyzing user needs in a real industrial scenario, but their focus remained on enhancing user performance. Only a few studies have involved workers in real industrial settings, and even those primarily concentrated on performance and ergonomic aspects rather than exploring operators' attitudes towards the technology [99], [100]. There is even less research concerning the acceptability of this new technology among operators. For example, in [101], a usability test was conducted to assess the reliability of a predictive maintenance system. While workers understood the useful features of the MR application, they required training to comprehend its usage fully. Therefore, there is a need for further studies to understand operators' attitudes towards this technology.

Understanding why individuals embrace information technologies is crucial, as it can greatly impact the design, evaluation, and predict how users will react to a novel technology [102]. Neglecting these factors may lead to unforeseen outcomes, such as the counterintuitive finding that paper-based TD might outperform MR technology in experimental studies [103].

Our work [104] explores how industrial operators respond to using MRTD developed following our "minimal AR" guidelines and using smart glasses in a real industrial maintenance scenario rather than in a controlled laboratory environment. We are particularly interested in understanding the acceptance of experienced technicians towards MR technology, especially when it involves unfamiliar interactions, rather than solely evaluating its impact on their performance during routine maintenance tasks. We selected the Microsoft HoloLens 2 [105] as HMD to facilitate our investigation due to its commercial availability and unique features currently attracting companies. Therefore, the following research question is investigated: "*In an industrial maintenance context, how much is the workers' acceptance towards a MRTD developed with our minimal AR guidelines and using a HMD in their daily work life?*". To answer this research question, we conducted a user study using the TAM, which helps identify the factors influencing the adoption of a new system [106]. The case study involved the maintenance of a real machine, where operators were accustomed to working under fixed conditions, such as lighting and equipment positions. After initial training, real workers were asked to interact with the GUI of a developed MRTD for a gas-fired radiant heating system, performing navigation tasks within the menu and a disassembly task. We then compared their task completion time with a focus group of MR experts to evaluate the impact of training on workers' learning skills, especially concerning interactions. Additionally, we collected feedback from the workers to gain deeper insights into their opinions regarding using MRTD. Another contribution of our study is to validate the "minimal AR" guidelines previously established through laboratory tests in a real industrial setting. These guidelines include ensuring a user-friendly interface [94], improving information readability with an OST display [107], and optimizing MRTD authoring [64].

5.2. Mixed Reality Technical Documentation design

For our case study, we selected a real maintenance scenario of the company Upgrading Services SpA. Our research project focused on creating a MRTD for a gas-fired radiant heating system (Fig. 18). To achieve this, we were granted access to the real machine and its maintenance manual, which were available in PDF format and the native language of the company.

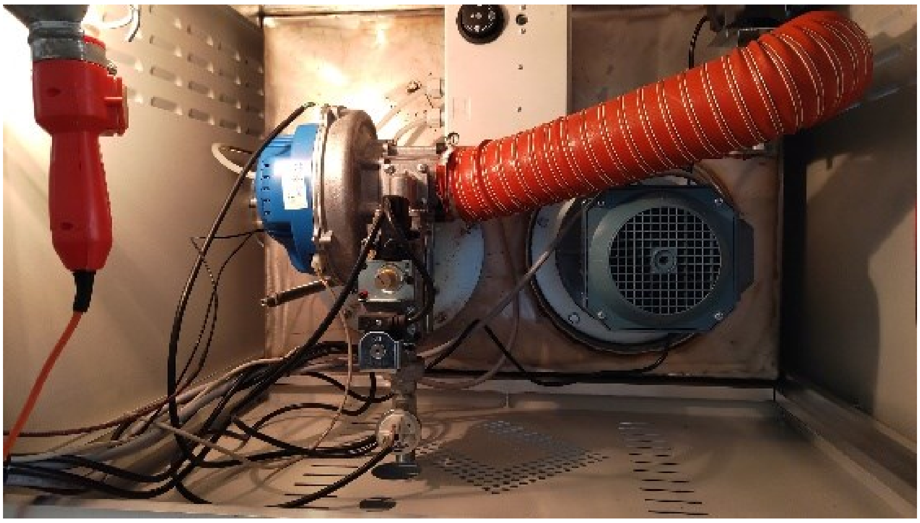


Fig. 18 Machine (gas-fired radiant heating system) the designed MRTD refers to.

Once we comprehended the functionality of the machine, we collaborated with company managers and experienced technicians to identify the primary user needs, listed in Table 9. These needs served as the foundation for defining the content of the MRTD in which we categorized the sections into two types: "AR scenes" that offer added value through AR in maintenance operations, and "MR scenes" that are essential for providing a complete manual to align with UN-005 but may not benefit from AR. The MRTD was specifically designed for this use case, focusing on the gas-fired radiant heating system. However, the framework we developed is versatile and can be easily adapted for other scenarios by replacing the pages of the new manual for the MR section. After tailoring the content according to the specific task requirements for the AR section, the visual assets used for conveying the same information can be replicated across different scenarios using the “minimal AR” approach described in Chapter 3.

Table 9 User needs deriving from the brainstorming with company managers and technicians.

ID	User Need	Source
UN-001	Identify the machine components	Technicians
UN-002	Localize the machine components	Technicians
UN-003	Understand the functionality of the machine components	Technicians
UN-004	Facilitate understanding of elementary maintenance operations	Technicians
UN-005	Provide all the information available in the PDF manual without exclusion	Technicians, managers
UN-006	Intuitive and easy-to-use device	Technicians, managers
UN-007	Be hands-free during maintenance tasks	Technicians, managers

To comply with UN-006 and UN-007, we selected the Microsoft HoloLens 2 [105]. To implement the MRTD, we used the Unity 3D Engine [108], along with the MRTK packages [109] and Vuforia Engine [110] for tracking. To ensure seamless functionality of the "AR scenes" content, we designed an image target and strategically placed it inside the right panel of the machine. To create the "AR scenes," we imported the complete 3D CAD model of the equipment into Unity.

The "MR scenes" in our MRTD refer to sections where there is no need to register information on the real components involved in the maintenance tasks. These scenes require a navigation interface to read information or select the corresponding "AR scene" to view. The purpose of these "MR scenes" is to serve the following functions:

- A login to provide a secure entry point for the operator to access the MRTD.
- Instructions for accessing "AR scenes" to offer guidance on accessing and interacting with the augmented content using natural features.
- Access to content that does not rely on augmented scenes, including information such as warnings, technical data, wiring diagrams, and calibration tables.

Fig. 19 provides an example of one of these "MR scenes" in the MRTD.

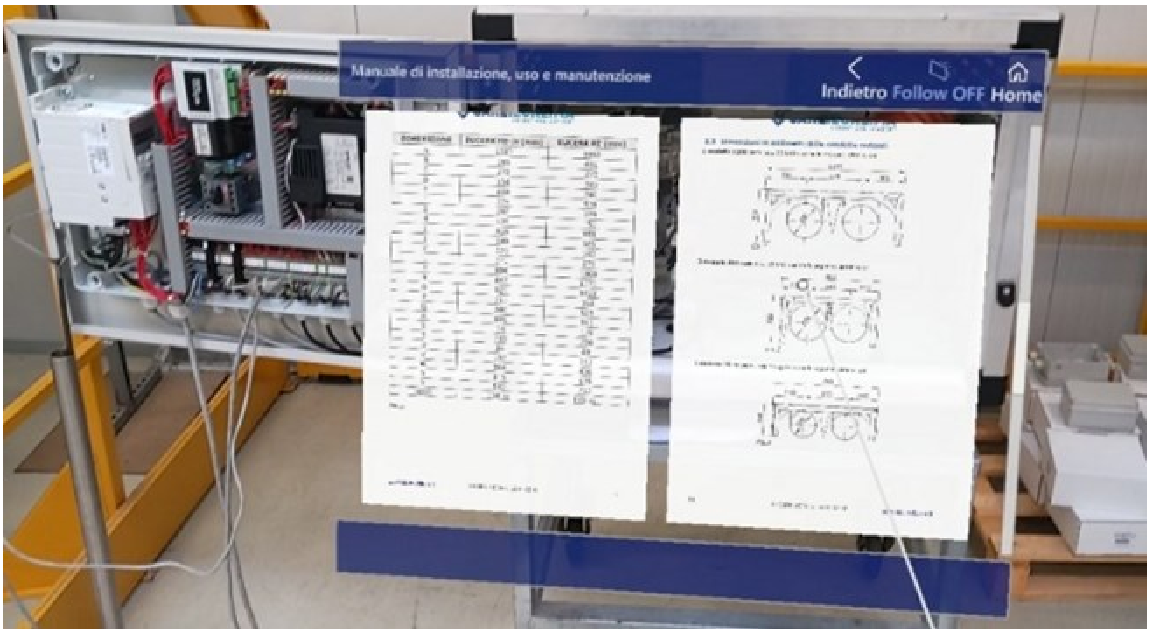


Fig. 19 Example of a "MR scene" of the developed MRTD with machine technical data.

Concerning the "AR scenes," we categorized them as those that necessitated the registration of information on the real components involved in the maintenance tasks. During our analysis, we identified three sections in the manual where AR technology could provide significant added value. Each of these sections represents an "AR scene" with specific features:

1. The first "AR scene" focuses on gas flow adjustment and includes features such as:
    - Identifying the components of the flow regulator.
    - Understanding the functionalities of these components for calibration.
2. The second "AR scene" deals with the adjustment of the mixing damper and encompasses features like:
    - Identifying the components of the mixing damper.
    - Understanding the correct positioning of these components for calibration.
3. The third "AR scene" is dedicated to the disassembly of thermal unit components and comprises features such as:

- Identifying the components of the burner, electrodes, and recirculation motor.
- Removing these components.

Fig. 20 provides an example of one of these "AR scenes" in the MRTD.

Fig. 20 Example of an "AR scene" of the developed MRTD about gas flow adjustment.

In creating the AR content for the MRTD, we adhered to the "minimal AR" authoring approach aimed at optimizing the visual assets used to communicate work instructions in manufacturing scenarios. Based on this approach, we considered the following guidelines for selecting appropriate visual assets:

- For work instructions that only involve identifying a component, we used a drawing.
- When the work instruction required identifying the component and its location, we used an AM.
- In cases where the work instruction needed to convey not only the identity and location of the component but also its orientation, we employed product models.

In accordance with the recommendations outlined in [94], the GUI of the developed MRTD was designed to adhere to essential principles that ensure user-friendliness: affordance, feedback, consistency, non-destructive operations (e.g., undo), discoverability, scalability, and reliability. We have described the specific design choices and strategies employed in the forthcoming paragraphs to implement these features effectively.

Both "AR scenes" and "MR scenes" were given an interface layout split into three distinct sections. In each section, buttons were incorporated with icons proposed by Microsoft (see Fig. 19 and Fig. 20). The first section of the interface showcases the title of the scene alongside essential buttons. These buttons allow users to perform various functions, including an "undo" button to revert to the previous section, a "Home" button to return to the main interface, and a button to enable/disable a mode where the interface can follow the user's movements, preventing it from getting out of view. The second section displays relevant textual content that comprehensively supports the user. This includes detailed instructions for each step in "AR scenes" to guide users through the process. Moreover, for "MR scenes," this section includes additional information such as technical data, wiring diagrams, and calibration tables. Finally, the third section is exclusively reserved for "AR scenes" with a toolbar containing buttons to facilitate filtering AR information based on different calibration configurations and component identification. It also allows users to switch between different steps or stages within the "AR scenes."

In both "AR scenes" and "MR scenes," we have incorporated a user-friendly feature that allows users to personalize the location, orientation, and size of the interface. This customization is achieved through anchors positioned along the contour of the interface, which is inspired by the suggestions in the MRTK packages. By adopting this approach, we aimed to maximize user comfort and cater to individual work area preferences. Even during experiments, users can adjust the interface to suit their needs. To enhance text readability, we have taken specific measures. Text elements, including labels and buttons, are presented with high-contrast colors, such as black or white, against a blue background [21]. Furthermore, we have implemented a visual feedback mechanism for buttons. When a user presses a button, it slightly lightens in color, providing immediate feedback and indicating that the button has been pressed.

Users are provided standard interaction options on the HoloLens, including the near pointer gesture, far pointer gesture, and voice commands. When using voice commands, users can use the names of the buttons, which are displayed and readable beneath their respective icons. These interaction methods operate independently of each other without any interference. As a result, we have deliberately allowed users the freedom to choose the interaction method they prefer.

5.3. Material and Methods

To answer the research question of this work ("*In an industrial maintenance context, how much is the workers' acceptance towards a MRTD developed with our minimal AR guidelines and using a HMD in their daily work life?*"), we conducted an experiment targeting expert technicians routinely engaged in maintenance tasks, particularly focusing on a selected case study machine. Due to the specific focus of this research on a particular group of individuals, the number of participants and the sample assortment provided by the company were limited. We managed to recruit a total of 7 participants (7 males, 25-47 years old, mean = 35, SD = 8). Each participant had more than 5 years of experience in the maintenance field. To assess their familiarity with MR technology and smart glasses, we used a 7-point Likert rating scale, where 1 indicated "Not at all familiar," and 7 indicated "Extremely familiar." On average, the participants rated their familiarity with MR as 2.0 (SD = 1.31, Median = 1, Min = 1, Max = 4). Similarly, the average familiarity rating for smart glasses was 1.86 (SD = 1.12, Median = 1, Min = 1, Max = 4).

Each user participating in the trial underwent two distinct parts: first, they engaged in a session where they used the developed MRTD on a gas-fired radiant heating system, and then they completed a subjective questionnaire. In the initial phase, users received training, during which they had the opportunity to try out all the interactions available on the HoloLens using a similar interface model as presented in the MRTD. Subsequently, they were assigned specific tasks to perform, which are listed as follows:

1. Log in to the MRTD (T1).
2. Read and comprehend the instructions regarding the function of the image target (T2).
3. Locate a specific section within the included PDF (T3).
4. Access the first "AR scene" and explore all the features of the buttons (T4).
5. Access the second "AR scene" and explore all the features of the buttons (T5).
6. Access the third "AR scene," manually perform step 1, and examine all the AR information related to the following steps (T6).

To assess the impact of training on the workers' learning and interaction skills, their task completion time was measured and compared to that of a focus group comprising 4 MR experts (2 females, 26-34 years old, mean = 29, SD = 3). The experts rated their familiarity with MR on the 7-point Likert scale, with an average score of 6.5 (SD = 0.50, Median = 6.50, Min = 6, Max = 7). Their familiarity with smart glasses was rated 6.0 on average (SD = 0.71, Median = 6.50, Min = 6, Max = 7). Unlike the workers, the MR experts had no prior experience with the machine used in the experiment.

In the second phase of the experiment, the operators were requested to complete a subjective questionnaire, which consisted of four sections: demographic characteristics, TAM factors, HoloLens interaction feedback, and recommendations on how to improve the MRTD. 29 selected TAM questions measured user acceptance. Participants were asked to rate their responses using a 7-point Likert scale, where 1 indicated "Strongly disagree," and 7 indicated "Strongly agree." These questions were adapted from validated questionnaires used in previous research [111]–[114] and were modified to suit the specific purpose of our study. As regards the

HoloLens interaction feedback, participants were asked to rate their experiences with the different interaction methods using a 7-point Likert scale. Specifically, the near pointer gesture received an average rating of 4.29 (SD = 2.19, Median = 5, Min = 1, Max = 7). The far pointer gesture received an average rating of 5.71 (SD = 1.03, Median = 6, Min = 4, Max =7). The voice command interaction received an average rating of 3 (SD = 2.27, Median = 2, Min = 1, Max = 7).

Before conducting the experiment, we ensured all users did not have color blindness. Additionally, participants were allowed to wear their eyeglasses while using the HoloLens. The duration of the experiment for each participant averaged around 40 minutes. We collected two data types for each participant throughout the experiment: errors manually checked by the experimenters and participants' feedback.

5.4. Results

In Table 10, we have provided the TAM questionnaire statements and descriptive statistics for each statement. We assessed the construct validity of six factors, namely Device Interaction (DI), Device Ergonomics (DE), Perceived Ease of Use (PEU), Perceived Usefulness (PU), Attitude Towards using the system (AT), and Intention to Use (IU). To measure each factor, the TAM questionnaire included three to eight items. The factor analysis was conducted, and a factor analysis score of at least 0.6 was considered satisfactory for survey validity [115], which was surpassed by almost all factors. This indicates the strong validity and reliability of the TAM survey. Furthermore, we assessed the internal consistency of items within each factor using Cronbach's Alpha reliability analysis. A Cronbach's alpha value of 0.7 or higher is generally considered acceptable for reliability [116]. In our study, all Cronbach's alpha values were above 0.7, demonstrating high internal consistency and reliability among the items. The mean values for each factor fell between 5.36 and 6.14. The standard deviation ranged from 0.89 to 1.86.

Table 10 Descriptive statistics of TAM factors for the MRTD acceptability with the HoloLens.

Factors	ID	Item	Factor analysis	Cronbach's alfa	Mean (SD)
Device Interaction	DI 1	HoloLens interaction is not frustrating	0.733	0.704	5.67 (1.34)
	DI 2	HoloLens interaction is fun to use	0.610		
	DI 3	HoloLens interaction is simple	0.907		
	DI 4	Doing the procedure by HoloLens is simple	0.413		

	DI 5	The MRTD would gain value adding audio/video	0.667		
	DI 6	Overall, the MRTD interaction is easy and simple	0.907		
Device Ergonomics	DE 1	The HoloLens is comfortable	0.684	0.891	5.43 (1.86)
	DE 2	The virtual element rendering is clear	0.966		
	DE 3	I am not tired due to the HoloLens after the tasks	0.739		
	DE 4	The HoloLens did not cause me any side effects	0.832		
	DE 5	The HoloLens did not strain my eyes	0.819		
Perceived Ease of Use	PEU 1	Learning how to use MRTD is easy for me	0.784	0.703	5.36 (1.49)
	PEU 2	Remember how to do the task using the MRTD is easy for me	0.384		
	PEU 3	I was able to access all the features of the MRTD without the help of an experienced technician	0.403		
	PEU 4	I was able to understand the entire MRTD system without prior knowledge of MR	0.572		
Perceived Usefulness	PU 1	It is easy to get from the MRTD what I want	0.633	0.914	5.88 (1.28)
	PU 2	Using MRTD improves my performance	0.727		
	PU 3	With the MRTD, I learn maintenance tasks faster	0.441		
	PU 4	With the MRTD, I learn maintenance tasks easier	0.383		
	PU 5	The MRTD supports me in individual understanding of more complex instructions	0.988		
	PU 6	The MRTD interface is easy for me to follow	0.691		
	PU 7	The association between real component and virtual element in the “AR scene” is good	0.318		
	PU 8	Overall, I find the MRTD useful	0.958		
Attitude Towards using the system	AT 1	Using the MRTD raises my interest in my work	0.828	0.852	6.14 (0.89)
	AT 2	I was able to understand what MR technology is using the MRTD	0.987		
	AT 3	Using the MRTD increases my motivation in performing the task	0.626		
Intention to Use	IU 1	Assuming I have access to the MRTD, I intend to use it in my daily work	0.884	0.939	5.90

IU 2	If I had to choose between traditional TD and the MRTD, I would choose the MRTD	0.973	(1.31)
IU 3	Assuming I have access to the MRTD, I feel able to consciously interact with MR technology	0.817	

Fig. 21 compares task completion times between workers and MR experts from the focus group. Except for training and tasks T4 and T6, no significant differences were observed in completion times between the two groups.

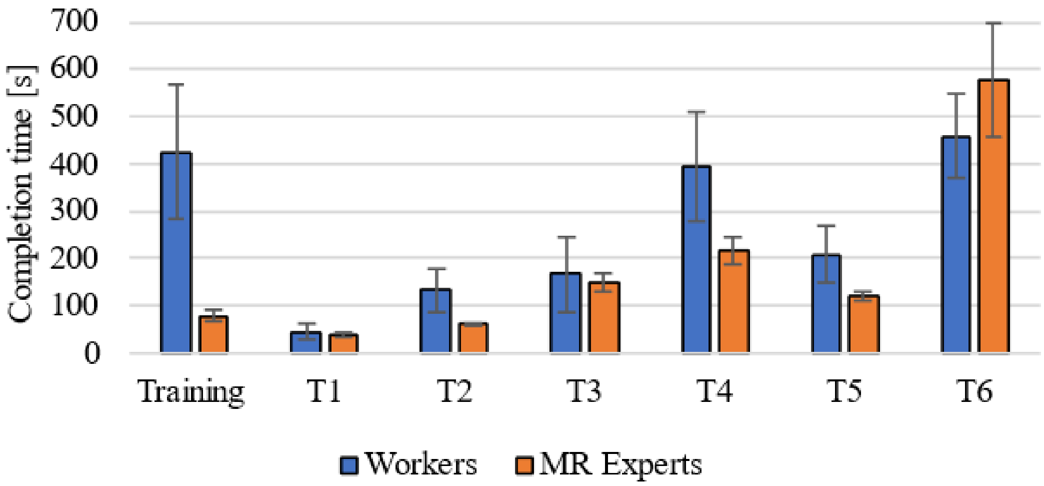


Fig. 21 Comparison of task completion times between workers and MR experts.

5.5. Discussion

This study represents one of the few attempts to explore the workers' acceptance of a MRTD using a Microsoft HoloLens in a real maintenance scenario. The findings from this investigation effectively addressed our research question. We demonstrated that the "minimal AR" guidelines can be effectively used for the MRTD authoring and we verified that workers generally embraced the MR technology when applied to maintenance tasks with a HMD. During the interviews, all the operators strongly approved of integrating this innovative technology into their daily work routines. According to their feedback, the MRTD facilitated rapid and easy access to TD information through AR content. This streamlined access helped them identify components more efficiently and better comprehend the task. Moreover, users noted a significant reduction in errors during task execution, to the extent that even inexperienced maintenance workers could successfully perform tasks with the MRTD. Furthermore, the workers appreciated the added value of the MRTD, which allowed the possibility of updating the content with additional information that might not have been initially available in traditional manuals.

All participants stressed the significance of training, especially for those unfamiliar with the HoloLens and MR technology. Despite the workers' limited prior exposure to MR and the HoloLens, there were no signs of negative attitudes towards the MRTD. This positive reception was attributed to the clarity of the instructions on how the image target functioned, enabling the workers to easily understand how to frame the marker for the "AR scenes." Notably, the workers' training times were understandably longer than those of the MR experts. However, the completion times for most tasks were comparable between the two groups. The main differences in completion times were observed in tasks T4 and T6. In task T4, workers took longer because it was the first task involving an "AR scene." On the other hand, MR experts encountered challenges in task T6 because they lacked experience in manually disassembling the machine. An interesting finding pertained to the standard deviation. While the workers exhibited varying completion times for each task, the MR experts consistently performed at the same speed. This result indicates that with practice towards MR technology, every operator can achieve high-performance times by overcoming any initial difficulties.

The workers successfully completed all tasks but demonstrated varying preferences for different types of interactions. The most favored interaction method among the users was the far pointer gesture, whereas the least preferred one was the voice command. The finding highlights that while voice command may be considered an "immediate" interaction by workers, it may be unreliable in specific situations. For instance, the voice command might not be accurately recognized in noisy environments, leading to potential inefficiencies. Additionally, the voice command functionality may become compromised if the user is wearing a facial protective system, like a mask.

According to the feedback from all users, the device ergonomics of the HoloLens was generally well-received during maintenance tasks. Users acknowledged the advantages of wearing the HoloLens, which gave them valuable benefits during their work. They particularly appreciated the flexibility to raise the transparent visor of the HoloLens when they needed to perform tasks without the interference of virtual elements that could obscure their view of real components. However, it is worth noting that two users who wore eyeglasses reported experiencing slight discomfort while using the HoloLens because the device was pressing on them.

During the study, participants offered feedback and recommendations for enhancing the MRTD in future developments. One of the suggestions related to interface readability. Although the option to enlarge the interface was available, users recommended having a large font size for text and wiring diagrams as the default setting. This would improve visibility and make it easier for users to interact with the content. Additionally, participants proposed the inclusion of illustrative videos alongside the provided PDF documentation. While not deemed essential, these videos could be helpful supplementary resources, providing visual demonstrations and further clarification for various tasks. Another important recommendation was incorporating an auxiliary voice feature to assist inexperienced users. This feature would guide users and help them locate specific features or elements if they encounter doubts during their interactions with the MRTD.

5.6. Lessons learned

As a result of the research focusing on a specific group of participants and the limited sample provided by the company, certain constraints were present. However, despite these limitations, the study yields valuable insights for adopting MR technology in companies. One notable finding is that while interaction through the HMD could benefit from technological advancements, MR offers significant advantages that make its implementation worthwhile and deserving of encouragement. The study emphasized the importance of conducting a thorough user needs analysis. For instance, the decision not to include all "AR scenes" in the manual was made based on the understanding that only those providing substantial added value should be incorporated. Additionally, the study highlighted the success of choices related to readability with good contrast and the “minimal AR” authoring approach. We achieved positive results by avoiding the use of attractive visual assets that may introduce excessive information. Workers encountered no difficulties while performing tasks and provided no negative comments, indicating a successful and well-optimized MR experience.

Therefore, this study allowed us to demonstrate that our “minimal AR” authoring approach positively influences user performance while performing a maintenance task in a real industrial context. Therefore, our next step has been to start optimizing the visual asset properties for each information type to maximize performance, as described in the following Chapter. This analysis will enable the creation of more comprehensive standards that can guide technical writers in developing ARTD.

Chapter 6. Visual asset optimization

This Chapter addresses the third issue of this dissertation: how to optimize the visual asset properties according to the information type and the working context. Selecting appropriate visual assets for ARTD is a complex process, particularly in a real industrial field. This complexity arises because AR depends on many factors that differ depending on the specific working context. Therefore, a modeling issue occurs. In fact, once the most suitable visual asset for a specific information type has been identified by applying the "minimal AR" authoring approach (e.g., an AM for a locating task), it is necessary to understand its optimized properties (shape, frame of reference, color coding, and animation) according to the working context. This Chapter addresses visual asset optimization based on factors which must be considered simultaneously, including the surroundings, lighting conditions, and blind areas. In our subsequent work [117], we focused on the information type LOCATION for which AMs were previously confirmed as the most suitable visual assets to convey this information [118].

6.1. Analysis of LOCATION as information type[5]

The localization of components is a critical aspect of procedural tasks like maintenance, assembly, and training, and it is a piece of information where AR cues can offer significant benefits. However, in complex machines where multiple objects are closely situated, there can be challenges in correctly attributing AR localization cues to the intended components, even if they are in-view and not occluded by other objects. To address this issue, the main goal of this work is to propose optimized AR visual cues that can enhance the accuracy and reliability of component localization in such scenarios.

The usage of AMs in AR applications offers advantages in terms of simplifying the authoring process. AMs can be readily accessed from a standard library, regardless of the specific machine or equipment used. Furthermore, they do not necessitate precise alignment with the real component, which enhances flexibility during the authoring phase. In contrast, although considered more visually appealing to end users [9], product models can provide an excess of information in addition to the fact that they typically require CAD models of each machine

[5] The results of the study described in this Section were published in the following paper:
S. Romano, E. Laviola, M. Gattullo, M. Fiorentino, and A. E. Uva, “More Arrows in the Quiver: Investigating the Use of Auxiliary Models to Localize In-view Components with Augmented Reality,” 2023 IEEE Int. Symp. Mix. Augment. Real., vol. PP, pp. 1–11, 2023, doi: 10.1109/TVCG.2023.3320229.

component that may not always be available. Additionally, they necessitate accurate alignment and overlapping with the real component [119].

However, using AMs in complex machines can introduce challenges in accurately identifying the correct component. AMs that employ pointing elements like arrows may create ambiguity as they can indicate multiple components. At the same time, AMs that define a spatial region like a sphere may encompass more than one component. In this work, we focused on addressing ambiguity arising from the localization of nearby, in-view, and not occluded components. To mitigate this ambiguity, we explored the possibility of designing optimized AMs based on the components shape. While AMs are commonly used in industrial AR prototypes, a notable lack of research is dedicated to designing AMs specifically for locating components in complex machines. Thus, we formulated the following research question: "*is it possible to design optimized auxiliary models for locating in-view not occluded components according to their shape?*"

6.1.1. Related Work

We analyzed in separate subsections the design variables of AMs considered in this research: frame of reference, shape, color, and animation. Lastly, we synthesized the main insights from existing literature on AMs design for localization tasks.

FRAME OF REFERENCE. The localization information can be conveyed through visual cues coded in the egocentric or exocentric frame of reference. Egocentric cues, such as arrows, halos, or CAD models, are aligned with the viewer's body axes and provide direct indications of the position of the target object relative to the viewer. On the other hand, exocentric cues, such as an overview map, compass, or world-in-miniature, provide information about the position of the target object in relation to other objects in the surrounding environment. Markov-Vetter et al. [120] found that egocentric visual cues result in faster and more reliable object localization. For this reason, AMs in existing AR interfaces are almost always in the egocentric frame of reference, as confirmed by Gattullo et al. [8].

SHAPE. The shape of an AM is the main property that influences how users process the localization information. There are two main categories of AM shapes: *pointing* AMs which indicate the specific location of the component to be localized, and *delimiting* AMs which define a region in space where the component is located. These categories can further be classified into 2D and 3D shapes.

Arrows are the main example of *pointing* AMs. 2D arrows simplify the creation of AR instructions [119]. However, they may not be easily recognizable from all viewing angles due to perspective effects unless they rotate consistently with the camera. This is why 3D arrows are more commonly used for localization tasks. Li et al. [121] proposed a taxonomy for the visual design of 3D arrow models. They suggested using prismatic arrows to indicate installation positions and draw attention to specific objects. This arrow shape is frequently employed to indicate components that need to be selected for maintenance operations [122]. Additionally, cylindrical arrows are also used in various scenarios. For instance, in [11], they indicate the points where screws should be fixed, while in [123], they guide users on where to place new parts during an assembly sequence. Arrows are also extensively used to direct users to points of interest not directly within the field of view of the device cameras, as demonstrated in [124] and [82]. Consequently, arrows in AR interfaces can convey different information, potentially confusing users. Lavric et al. [119] highlighted this concern by distinguishing between a blue vertical 2D arrow used for localization tasks and an orange horizontal 2D arrow with a distinct shape used to suggest to operators how to turn their head to reach the assembly area.

2D *delimiting* AMs are simple shapes drawn on a 3D plane aligned with one of the component surfaces, typically the one facing the user. These shapes can be categorized as filled or outlined [125]. Filled shapes may occlude the user's view as they cover a significant portion of the real environment. As a result, they are primarily used in SAR applications to emphasize the surface of a component that needs to be manipulated [2], [126]. However, for HMDs or HHDs applications, transparency is employed to mitigate occlusion-related issues [86], [127]. On the other hand, outline shapes are designed to avoid occlusion risks. They are slightly larger than the boundaries of the component to compensate for tracking inaccuracies [87]. The most used outline shapes include rectangles [128], circles [6], [122], multiple concentric circles [129], [130], and crosshairs [131], [132]. In remote collaboration applications, free-hand sketches are often employed [133], [134].

3D *delimiting* AMs are positioned to enclose the entire component that needs to be located. Compared to 2D shapes, they are less commonly employed in the literature. In an AR assembly procedure, Blattgerste et al. [131] used a simple cuboid that matched the size and color of the real object instead of employing a more detailed 3D model. Similarly, Renner and Pfeiffer [129] used a cuboid as a visual cue, comparing it with a 3D spline path originating in front of the user and extending to the target location. As the 2D filled AMs, the 3D ones can also occlude the user's view, particularly when dealing with large objects. To address this issue, transparency is employed [135], [136], or the 3D figure is rendered in wireframe mode [137].

COLOR. Red is widely used in localization tasks due to its high visibility to stand out against various backgrounds. It serves as an effective means to capture attention and highlight important information [138]. For example, Obermair et al. [122] use red circles and arrows to indicate screws and other components needing removal in a maintenance procedure. Schwerdtfeger and Klinker [139] compared three different types of red-colored cues for conveying order-picking information. Other colors commonly used for object localization in IAR interfaces include green and yellow. One possible reason for their usage is that these colors are less prevalent in industrial environments, reducing the likelihood of objects being mistaken for other elements. For example, Radkowski et al. [128] employed a green frame to highlight the part that requires assembly in an AR assembly procedure. Similarly, Webel et al. [140] used a yellow highlight to provide spatial information about the current step in an AR maintenance procedure.

Using colored AMs allows the conveyance of additional information through color coding, where different colors are associated with specific meanings. Typically, red is associated with hazard or error, yellow with caution or warning, and green with success or completion. For instance, in a study by Funk et al. [141], green was used to indicate the position of the next picking bin and where to assemble the selected part, while red was used to signal picking from the wrong bin in a manual assembly workplace. In another example, Gruenefeld et al. [142] employed a color gradient from blue to red to encode the distance between a 2D AM and the physical object it refers to. They applied the cold and warm metaphor commonly seen in heatmaps, where red represents proximity and blue represents greater distance.

When designing colors for AR interfaces, it is important to consider the perception of colors specific to the device used. OST HMDs and SAR systems have limitations on the range of colors that can be effectively used. In an early research study involving old-generation HMDs, Thomas et al. [143] recommended avoiding certain colors such as cyan, orange, magenta, pink, and red. In SAR systems, dark colors, like blue, produce minimal light and are challenging to recognize [144]. Similarly, with OST devices, dark colors tend to disappear against light backgrounds visually [145]. Another perceptual challenge with OST devices is color distortion, specifically a shift in hue due to the blending of AR graphics with the background texture under varying lighting conditions [146]. Merenda et al. [147] conducted a study to evaluate user performance and color perception when interface elements were presented against different backgrounds. They discovered that blue, green, and yellow performed more consistently and reliably than other commonly recognizable colors. Based on these findings, Ping et al. [148] specifically selected these three colors to investigate their effect on depth perception in AR highlighting. Their study revealed that distance estimation errors were lower with green and yellow AMs compared to blue.

ANIMATION. Animations can be useful in directing the user's attention to an object, particularly when it is occluded by another object [149]. One common animation technique is blinking. For example, Volmer et al. [150] suggested employing a blinking annotation at intervals of 300 milliseconds to provide information about the precise location of the next task. Another example is that of blinking rectangles used as an animation technique by Buttner et al. [151] to recreate the attention funnel effect in spatial AR for picking tasks.

Animations can also be applied to arrows to enhance their effectiveness. One animation technique involves animating arrows back and forth along their axis, as shown in [152]. Additionally, the "fade" and "wipe" animations are widely used as infographics to guide users' attention and improve engagement [153]. The fading animation gradually hides the AM once the user has located the object it refers to. However, these techniques were primarily proposed for out-of-view objects. For instance, in [82], a 3D arrow pointing to the target object gradually fades to full transparency when the user orients toward the target. Bonanni et al. [154] introduced a technique where the brightness of the AM used to locate a handle is modulated to draw the user's attention to it. A combination of directional animation and fading can create wipe animations for in-view objects, where the AMs gradually fade in and out along a specific direction [155].

In IAR interfaces, animated cues are commonly employed, especially to provide operators with a preview of the task they need to accomplish. However, AMs used for locating information are typically static. The animation for the LOCATION does not offer additional information beyond what is already conveyed by the static version of the AM.

SUMMARY. Based on the analysis of the literature, several conclusions can be drawn:

- AMs are typically displayed in the egocentric frame of reference.
- There is no recommended shape for AMs. Although 3D arrows are commonly used, their association with other information types can be confusing.
- There is no single recommended color for AMs. Red, yellow, and green are the most used colors.
- Static AMs are generally preferred over animated ones.

Overall, there is a gap in understanding how AMs can effectively locate objects in AR interfaces for procedural tasks. Specifically, there is a lack of studies addressing whether a generic shape of AM, such as an arrow, can be applied for all types of components or if its choice is influenced by the shape of the component being located.

6.1.2. Material and Methods

A first user study (US 1) was carried out to examine whether user preferences regarding AMs for a localization task can be affected by the shape of the localized component. The study focused on three AM properties: shape, color, and animation. This study did not consider position and orientation, as existing research clearly indicates that AMs should be presented within the egocentric frame of reference [8], [120].

In our experiment, we used the model of a car engine (Fig. 22) as a case study. The Microsoft HoloLens 2 [105] was selected as the device. We used the Mixed Reality Toolkit (MRTK) [109], Unity 3D Engine [108], and Vuforia Engine [110] to develop the application. To align the virtual content with the physical machine, we created an image target and positioned it on a smooth surface beneath the car engine. We designed a Graphical User Interface (GUI) to allow users to choose the AM design properties. It was positioned on the ceiling and could be accessed through the HoloLens. We implemented eye-tracking technology to streamline the interaction process and minimize potential biases, allowing users to interact with the GUI more efficiently.

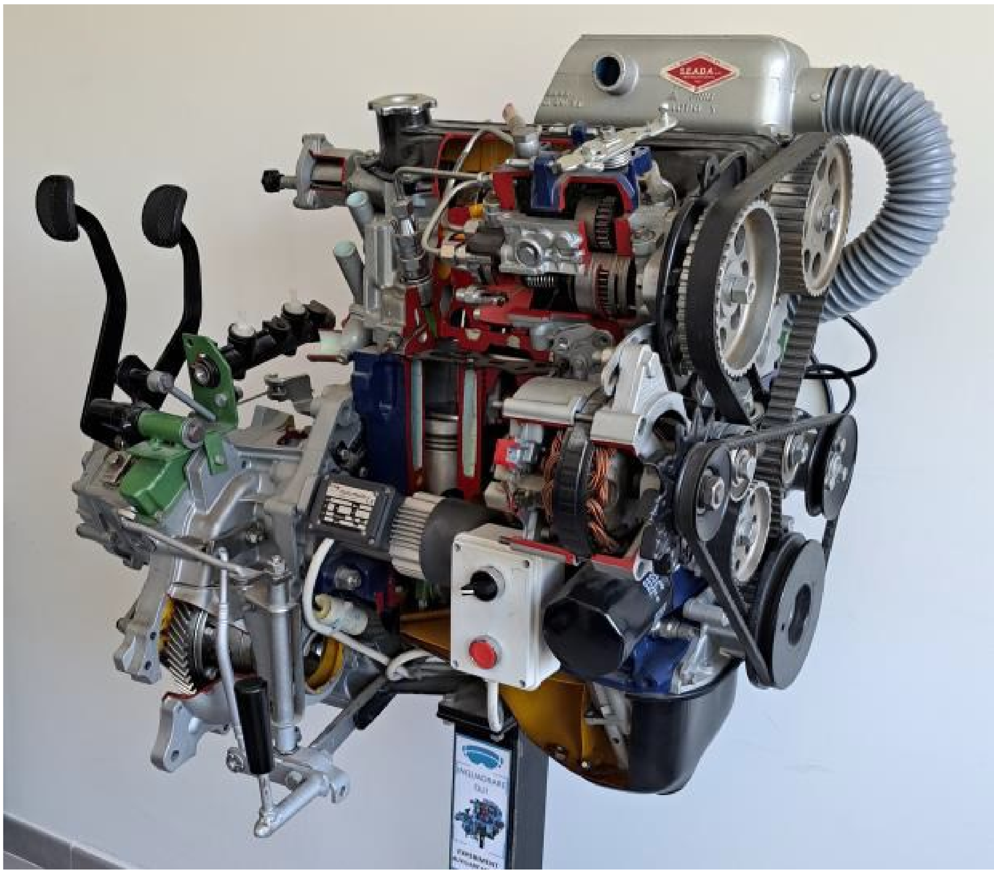


Fig. 22 Machine (car engine) used for the experiments.

To examine how the shape of components influences the properties of AMs, it is fundamental to modify this parameter during the study. Existing literature often describes the shape of an object using its bounding box [156], [157]. To achieve this, we used a CAD tool to calculate the Object-Oriented Bounding Box (OOBB) [158] for each component, i.e., a cuboid that encompasses the entire object (Fig. 23). The center point of the bounding box aligns with the centroid of the component, while its orientation is determined by leveraging symmetry or functional axes of the component, such as the rotation axes of pulleys, motors, and screws. In cases where it was not feasible to define such axes, we referred to the primary axes of the overall machine.

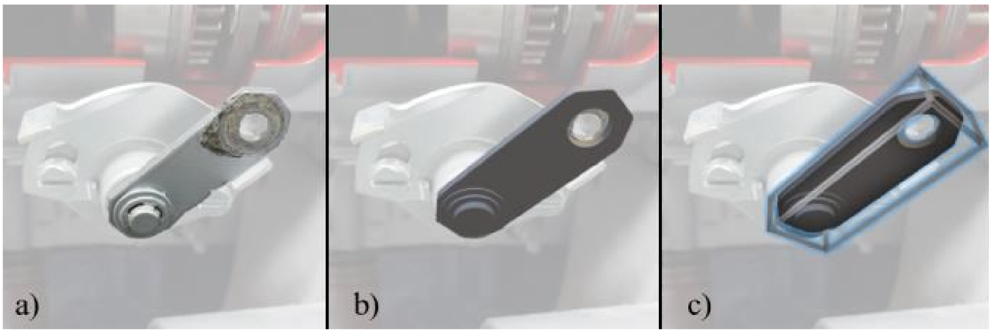
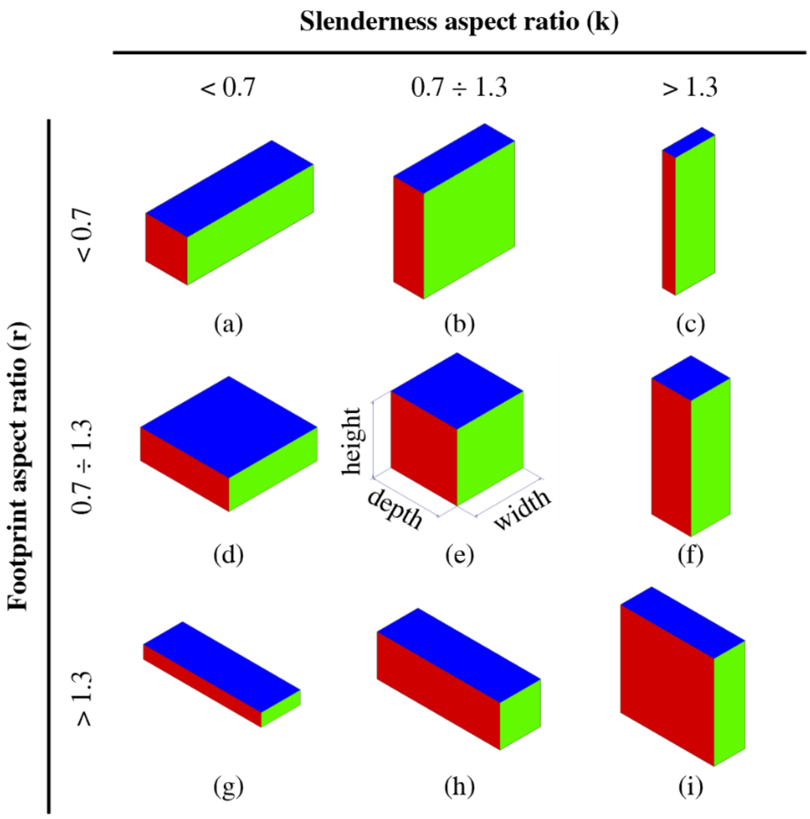


Fig. 23 For each object (a), exploiting its CAD model (b), we computed its bounding box (c) to define the shape.

Then, we proceeded to classify the shapes of the bounding boxes using the parameters "footprint aspect ratio" (r = depth/width) and "slenderness aspect ratio" (k = height/width), as defined in [159]. A perfect cube corresponds to r = 1 and k = 1. However, in the case of many industrial components, the bounding boxes may have r and k values close to 1, indicating a shape similar to a cube but not exactly perfect. To address this, we conducted a preliminary analysis of the bounding boxes for 40 different component shapes to determine which could be considered cubes. Based on this analysis, we established a tolerance of ±0.3 for r and k, resulting in three intervals for each parameter. Consequently, this led to formulating nine conditions, as presented in Table 11. Among the conditions, "e" represents a *regular cuboid* since the footprint and slenderness values are close to 1. Conditions "a," "f," and "h" have the same aspect ratio for the bounding box but differ in orientation, so they were grouped together as an *elongated cuboid*. Similarly, conditions "b," "d," and "i" were categorized as a *square plate*, while conditions "c" and "g" were classified as a *rectangular plate*.

Table 11 The bounding boxes for each condition according to the parameters (r, k) to define the component shape.



Additionally, we considered a second parameter to define the shape of a component, which is related to the bounding box. In fact, the same aspect ratio of a bounding box can be achieved either by a shape that occupies a significant portion of it or by a shape that covers it less uniformly or regularly. We introduced a geometric parameter known as the volume percentage to differentiate between these two conditions. It is calculated by comparing the volume of the component (including its cavities) to the volume of the bounding box (V = Vc/Vb). A preliminary evaluation of 40 different component shapes showed us that components with $V < 0.2$ could be classified as *irregular*. For these irregular components, the OOBB does not accurately represent their real shape. Consequently, we did not differentiate them based on the aspect ratio of the bounding box, and we considered them as a separate fifth category in this study. Following the established component categories, we selected 2 components with distinct sizes from the machine engine for each category. These components were denoted as "A" and "B" within their respective categories. Consequently, a total of 10 components were included in the localization task. Table 12 provides an overview of the components chosen from the car engine.

Table 12 Components chosen for the US 1 and ranked in each category according to the parameters (r, k, and V).

Component	Shape category				
	Regular cuboid	**Elongated cuboid**	**Square plate**	**Rectangular plate**	**Irregular**
A	 k=1.00 r=1.13 V=0.79	 k=2.00 r=0.95 V=0.97	 k=0.15 r=1.00 V=0.86	 k=2.78 r=0.11 V=0.94	 V=0.12
B	 k=1.00 r=1.00 V=0.75	 k=13.33 r=1.00 V=0.79	 k=0.23 r=1.00 V=0.34	 k=1.50 r=0.19 V=0.92	 V=0.13

Regarding the AM shapes, we discovered that Paint 3D provides the most extensive selection of 2D and 3D standard geometries. However, we excluded certain shapes unsuitable for a localization task, as they are typically used for different purposes (cross, tick, heart, and cloud). On the other hand, based on our literature review, we also incorporated 3D *pointing* AMs, specifically the prismatic and cylindrical arrows. The selected 36 AM shapes were categorized as 3D shapes, 2D outline shapes, and 2D filled shapes, following the classification outlined in [125] (Fig. 24).

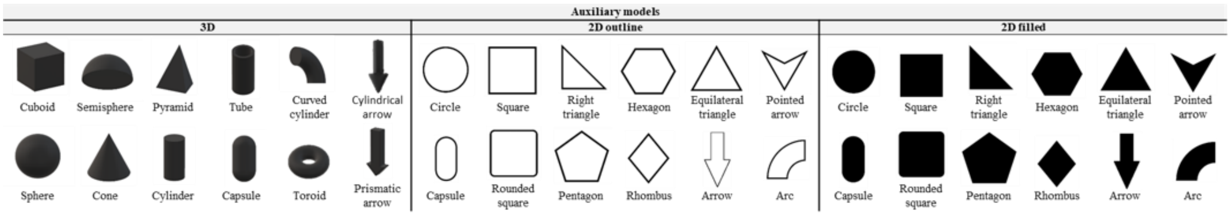


Fig. 24 The AM shapes selected for the US 1 and ranked as 3D shapes, 2D outline shapes, and 2D filled shapes.

In the AR environment, we designed all the AMs by configuring the Unity transform properties, which include position, rotation, and scale. To determine the AMs positioning, we used the centroid of the bounding box for each component as a reference point. Distinct Unity transform properties were applied to *pointing* and *delimiting* AMs. For *delimiting* AMs, we superimposed them onto the components by aligning the centroid of the AM with the centroid of the bounding box. Their scale was set to the minimum value necessary to encompass the entire bounding box of each component. *Pointing* AMs, on the other hand, were positioned above the bounding box by directing them towards the centroid of the bounding box, with a downward orientation. Their scale was set in accordance with the dimensions of the bounding box. Both *pointing* and *delimiting* 2D AMs were oriented by overlaying them on the principal plane of the bounding box. This plane was determined by identifying the normal vector most closely approximating the user's viewing direction.

To determine the colors, we adopted Berlin and Kay's method of categorizing basic color terms based on their evolutionary process consisting of seven stages [160]. In this way, we defined eleven "basic color categories." However, we excluded gray and black from our color options since they did not display well on an OST display, as noted in [147]. From the remaining nine colors, we converted the centroid positions from the Munsell color system to sRGB coordinates. As a result, the selected colors for the experiments were white (RGB = 255, 255, 255), green (61, 145, 89), blue (22, 128, 162), yellow (237, 192, 44), red (174, 42, 50), purple (123, 87, 142), orange (226, 127, 45), brown (140, 94, 46), and pink (214, 124, 130). All AMs were designed with an opacity value of alpha = 1, ensuring that they appeared fully opaque and provided optimal visibility with the maximum graphic output of the HoloLens device.

In terms of animation, besides static AMs, we selected two commonly used animations to attract the user's attention. These animations include the "blinking" effect, as seen in [150], [154], and the "wipe" effect [155].

45 unpaid participants (11 females, 20-34 years old, mean = 24.7, SD = 3.01) were recruited from our university and local companies. They were 7 bachelor's and 24 master's degree students in engineering, 8 Ph.D. students in mechanical engineering, and 6 employed engineers. All participants had normal or corrected-to-normal vision, and none had any color vision impairments. Prior experience with AR was assessed using a 7-point Likert scale, and the average familiarity rating was 3.29 (SD = 2.10, Min = 1, Max = 7). During the experiment, participants were allowed to wear their eyeglasses while using the HoloLens device. The study required no specific previous working experience from users. The average duration of the entire experiment, including the final questionnaire, was approximately 45 minutes per participant.

We employed a balanced Latin Square design to establish the order in which each user would localize the 10 components. The process of choosing the AM properties through the AR GUI for each component can be summarized as follows (Fig. 25):

- Users were initially provided with information about the component they needed to localize. This information was conveyed through a photograph displayed in the AR GUI.
- Users could explore and evaluate all the proposed AM shapes rendered using the Unity Default Material Shader. These shapes were presented without any animations.

Subsequently, users selected the AM shape they deemed most suitable for localizing the component.

- Once users made their selection, they confirmed it, and the GUI gave them the option to choose the color for the AM shape. Users could test the selected AM shape with various available colors. They then selected the color they believed to be the most appropriate. It is worth noting that the experimenter advised users to choose the color pragmatically rather than based on personal preference.
- After confirming the color selection, the GUI allowed users to choose the animation for the AM shape. Users tested the selected shape and color with all the available animations and chose the animation condition they considered most suitable for localization.

Users' motivations regarding the selected AM properties were collected at the end of each choice.

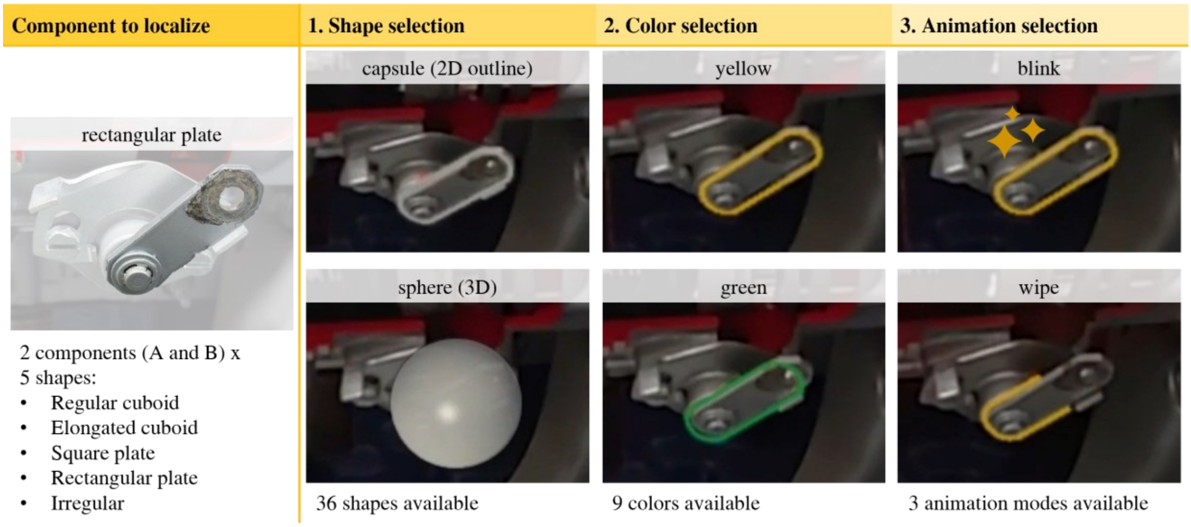


Fig. 25 Procedure for US1: users chose the AM properties for 10 components of 5 different shape categories. They first selected the shape of the AMs among 36 possible available shapes; then, users chose the color and the animation for the selected AM shape. We reported only some examples of the possible choices.

6.1.3. Results

Fig. 26 provides a visualization of the frequency of AM shapes and their corresponding components. Our observations showed that no users proposed a single AM shape for every component. Users chose 7.3 (SD = 1.5) different AM shapes on average. To explore further, we investigated whether it would be possible to propose specific AM shapes based on the five categories of component shapes, considering user preferences.

		Regular cuboid A	Regular cuboid B	Rectang. plate A	Rectang. plate B	Square plate A	Square plate B	Elongated cuboid A	Elongated cuboid B	Irregular A	Irregular B
3D	Cuboid		2%		2%			2%		2%	
3D	Semisphere					7%				7%	2%
3D	Pyramid									2%	
3D	Tube	11%	2%		2%		4%	29%	22%	7%	2%
3D	Curved cylinder								2%		4%
3D	Cylindrical arrow	4%	2%	4%	2%	9%		9%			9%
3D	Sphere	4%	2%			2%				4%	7%
3D	Cone									2%	
3D	Cylinder	62%	13%			4%	7%	20%	22%	29%	
3D	Capsule		2%	9%				11%	9%		
3D	Toroid		7%			7%	22%				2%
3D	Prismatic arrow		11%	2%		16%		4%		4%	16%
2D outline	Circle		11%	4%			51%	4%	2%	13%	11%
2D outline	Square	2%			18%			2%			
2D outline	Right triangle										
2D outline	Hexagon		16%			16%	4%			9%	
2D outline	Equilateral triangle				2%					2%	
2D outline	Pointed arrow		2%	4%		2%					11%
2D outline	Capsule			53%	7%			9%	27%		
2D outline	Rounded square	7%			24%			2%			
2D outline	Pentagon		2%							2%	2%
2D outline	Rhombus										4%
2D outline	Arrow		4%	2%			2%		2%		11%
2D outline	Arc										4%
2D filled	Circle					2%	9%	2%			
2D filled	Square				13%					2%	
2D filled	Right triangle										
2D filled	Hexagon		13%			24%				7%	
2D filled	Equilateral triangle									4%	
2D filled	Pointed arrow	4%	2%	2%	7%	2%					
2D filled	Capsule			18%	4%			2%	11%		
2D filled	Rounded square				16%			2%			
2D filled	Pentagon					2%					
2D filled	Rhombus										
2D filled	Arrow	4%	7%		2%	7%			2%	2%	7%
2D filled	Arc										7%

0 … 65%

Fig. 26 Frequency with which users chose an AM shape for each component. The heatmap chart helps show the trends among AM shape proposals; a darker color indicates a high number of proposals, and a white box represents no proposals.

Before conducting the experiment, we hypothesized that a 3D sphere, cuboid, or cylinder would be the most preferred AM shapes for the *regular cuboid* category because these shapes would provide uniform coverage in all three directions. We observed that the most chosen AM shape for component A was the *3D cylinder*, selected by 62% of users. However, for component B (a nut), there was greater uncertainty in the AM shape selection, probably due to its small dimensions, causing it to appear more like a point with a reduction of the component shape effect. Nevertheless, the 3D cylinder was still one of the highly rated proposals for component B. Users justified their choice of the 3D cylinder for both components by stating that it closely approximated the shape of the respective components. This observation can be generalized to other industrial components in the regular cuboid category, as cylinders are commonly found in machine components such as shafts, pins, and screws compared to other shapes.

In the case of the *elongated cuboid* category, our initial hypothesis was that 3D shapes with one dimension significantly greater than the other two (such as a cylinder, capsule, or elongated cuboid) would be the most preferred for this category. We found that for component A, the *3D tube* was selected by 29% of users, followed closely by the *3D cylinder* at 20%. These two proposals showed a high level of agreement compared to other shapes, indicating a preference for shapes that resemble elongated forms. However, for component B, there was a higher level of uncertainty in the selection of the AM shape. This can be attributed to the thin nature of component B, which was a thin lever. The thinness of the component caused its shape to appear more like a line, diminishing the impact of the component shape. Users justified their choice by stating that these 3D shapes (cylinder and tube) encapsulated the components well. They mentioned that the advantage of the tube shape was the ability to see the component inside it.

In the case of the *rectangular plate* and *square plate* categories, our hypothesis before the experiment was that 2D outline shapes would be the most preferred. This hypothesis was based on the idea that outline shapes can effectively highlight the component without occluding it. For the *rectangular plate* component A, there was a high agreement (53%) for the *2D capsule* outline shape compared to the filled shape (18%). Similarly, for the *rectangular plate* component B, there was an agreement for the *2D square* shape with both segmented (31% overall, 18% for the outline one, and 13% for the filled one) and rounded edges (40% overall, 24% for the outline one and 16% for the filled one). For the *square plate* category, the *2D hexagon* shape was favored for component A, with a high agreement for both the outline shape (16%) and the filled one (24%). In the case of *square plate* component B, there was a strong agreement (51%) for the *2D outline circle* shape. Users justified their choices by noting that, given the small thickness of these components, a 3D model was unnecessary. Then, among the available 2D models, they selected shapes that best represented the boundary of the component. The preference was generally for the outline versions of the shapes, as they did not occlude the component. These results support our initial hypothesis and provide valuable insights regarding the shape of the AM. These findings can be generalized to other industrial components within these categories.

In the case of the *irregular* category of components, our initial hypothesis was that uniform shapes such as the 3D sphere or cuboid would have been the most preferred because they could encompass the component regardless of its specific form. However, the experiment results showed a high level of disagreement among the proposed shapes for components A and B. For component A, the *3D cylinder* was the most preferred shape, selected by 29% of users. Additionally, the *2D outline circle* received a preference of 13%. For component B, there was even greater variability in the preferred shapes. The *3D cylindrical arrow*, *2D outline circle*, *2D outline prismatic arrow*, and *2D outline cylindrical arrow* were all equally favored, each selected by 11% of users. Users expressed difficulty finding a suitable AM shape for this category of components. Based on the survey results, we concluded that these findings may not be easily generalizable, and further research is required before providing specific guidelines for AM selection in relation to irregular components.

From the data analysis on color preferences, it was observed that five users consistently proposed the same color independently of the component shape. In four cases, that color was yellow. Fig. 27 displays the frequency of AM colors collected during the study. Interestingly, yellow emerged as the most chosen color for all the component categories. On average, it was selected by 34% of users, with a maximum frequency of 44% for component B in the *regular cuboid* category and a minimum frequency of 22% for component A in the *rectangular plate* category. Other colors, such as green (average of 13%), red (13%), and blue (12%), were also preferred, but to a lesser extent than yellow. This result aligns with numerous studies in the literature [147], [148], where yellow and green are favored in perception with OST devices. These colors are known to provide more consistent and reliable recognition compared to other colors. One possible reason for the preference for yellow is its strong contrast with most of the engine components in the proposed case study.

	Regular cuboid		Rectangular plate		Square plate		Elongated cuboid		Irregular	
	A	B	A	B	A	B	A	B	A	B
Orange	13%	9%	20%	4%	7%	16%	7%	9%	9%	4%
Blue	9%	9%	9%	16%	7%	11%	13%	16%	18%	16%
Yellow	36%	44%	22%	40%	40%	24%	31%	36%	29%	36%
Brown	4%	2%	11%	7%	4%		2%	7%	2%	
Pink	2%		11%		2%	2%		2%	2%	2%
Red	11%	9%	11%	9%	16%	18%	20%	9%	16%	11%
Green	7%	18%	7%	18%	13%	18%	13%	9%	9%	18%
Purple	9%	2%	2%		2%		4%	7%	7%	7%
White	9%	7%	7%	7%	9%	11%	9%	7%	9%	7%

0													65%

Fig. 27 Frequency with which participants chose an AM color for each component.

Fig. 28 presents the data collected regarding the frequency of AM animations. The blinking effect is the most preferred animation in all categories, except for elongated cuboid component A, where a static AM was favored with a similar frequency. On average, the blinking effect was selected by 47% of users. Following this, the static animation was chosen by 30% of users, and the wipe animation was preferred by 23%. Users provided feedback indicating that the blinking effect captured more attention than the other proposed animations. Additionally, they noted that blinking 3D AMs allowed the component to remain visible during the animation. Finally, users who preferred a static AM expressed that animation was unnecessary, especially for larger-scale components.

	Regular cuboid		Rectangular plate		Square plate		Elongated cuboid		Irregular	
	A	B	A	B	A	B	A	B	A	B
Wipe	24%	16%	24%	20%	20%	18%	27%	24%	33%	22%
Blink	49%	58%	56%	42%	49%	47%	36%	47%	36%	53%
None	27%	27%	20%	38%	31%	36%	38%	29%	31%	24%

0												65%

Fig. 28 Frequency with which participants chose an AM animation for each component.

Design Guidelines. Based on the findings obtained from US 1, we have formulated a set of recommendations for the design of AMs:

- When dealing with a component that has a regular cuboid bounding box (*regular*), it is advised to use a 3D cylinder as the AM.
- For components with an elongated cuboid bounding box (*elongated*), employing a 3D tube or cylinder as the AM is recommended.
- In the case of a component with a plate-shaped bounding box (*plate*), it is suggested to use a 2D AM that mimics the shape of the component boundary.
- Regardless of the component shape, it is advised to employ yellow AMs with a blinking animation.

It is important to note that these guidelines apply only to components that uniformly cover the bounding box ($V > 0.2$). Additional research is necessary for components with more irregular shapes to establish appropriate design guidelines.

6.1.4. Validation of the guidelines

To verify the findings obtained in US 1, we designed a second user study (US 2) in which we conducted an objective assessment. Specifically, we compared the AMs created based on the guidelines derived from US 1 with those designed according to what is evident in the literature and what is available using a commercial AR authoring platform.

In our study, we implemented a within-subject user study with two independent variables: component shape (*regular*, *elongated*, and *plate*) and AM shape (*baseline* and *optimized*). As a result, we examined six different experimental conditions. For each one, we conducted four replications selecting four distinct components corresponding to each of the three component shapes, for a total of 12 components. The case study focused on the same engine as US 1.

The AM shape used as a baseline was a 3D prismatic arrow widely used in similar tasks, as documented in the literature. These 3D prismatic arrows are also commonly employed as visual cues in commercial AR authoring tools like Microsoft Dynamics 365 Guides and Microsoft Dynamics 365 Remote Assist. Following our guidelines, the shape of optimized AMs was tailored to the component shape. For the *regular* category, we suggested using a *3D cylinder*, a *3D tube* for the *elongated* category, and a *2D outline* shape that follows the component boundary for the *plate* category. Based on the outcomes of US 1, we decided to use a yellow color (RGB = 237, 192, 44) for the proposed AMs and animated them with a blinking effect. The yellow color was also applied to the 3D arrows, aligning with the findings of US 1, as no definitive color recommendations were found in the available literature. However, concerning animation, we observed that most interfaces employ static AMs. Therefore, we found more appropriate to use a static 3D arrow as a baseline. This choice was further validated by examining commercial AR authoring tools, such as Microsoft Dynamics 365 Remote Assist, which allows color customization but lacks animation options. Regarding the position, orientation, and scale of our AM proposals, we set the Unity transform properties in line with those used in US 1.

We compared performance (localization time and recognition accuracy) and user experience (ease of localization, clarity of localization, and enjoyment) in the localization task. This analysis involved using both baseline and optimized AMs for the three different component shapes. Before the experiment, we formulated the following hypotheses:

H1. The localization time achieved with optimized AMs will be significantly lower compared to the baseline for all three component shapes.

H2. The recognition accuracy obtained with optimized AMs will be significantly higher compared to the baseline for all three component shapes.

H3. The user experience associated with optimized AMs will be significantly higher compared to the baseline for all three component shapes.

We enlisted a total of 24 users (8 females, 21-32 years old, mean = 24.1, SD = 2.40) from our university and local companies to participate in our study. They were distinct from the participants of US 1. They were 7 undergraduate students and 11 graduate students pursuing engineering degrees. Additionally, there were 5 Ph.D. students specializing in mechanical engineering and 1 employed engineer. All participants had either normal or corrected-to-normal vision, and none had any color vision impairments. During the study, participants were provided with the HoloLens and were allowed to wear their eyeglasses if needed. It is worth noting that no prior working experience was required from the participants. The participants rated their familiarity with AR on a 7-point Likert scale, with the average familiarity level being 3.21 (SD = 2.20, Min = 1, Max = 7). On average, each participant spent approximately 25 minutes engaged in the experiment.

For each participant, we developed a randomly generated predetermined sequence. The localization task for each component involved the following steps:

- Users started the localization task by pressing the "Start" button on the GUI.
- Users identified the component using the AM provided.
- Users pressed the "Finish" button to confirm their selection once they felt confident in their identification.

During the test, the application automatically captured and stored the localization time for each AM in an online spreadsheet. The localization time was measured from when participants pressed the "Start" button to initiate the task until they pressed the "Finish" button to conclude the task. As part of the task, participants were also required to touch the localized component physically, allowing the experimenter to record any potential errors that occurred manually.

After the experiment, participants were asked to complete a subjective questionnaire comprising three statements related to ease of localization, clarity of localization, and enjoyment. Participants were asked to rate each statement on a 7-point Likert scale, ranging from 1 (strongly disagree) to 7 (strongly agree). The statements included:

i. "It allows me to easily localize the component."

ii. "It allows me to localize the component without ambiguity or confusion regarding nearby components."
iii. "I enjoyed using it to localize a component."

Participants rated the optimized and baseline AMs for each component category.

The recorded data from each participant, including localization time, recognition accuracy, and user experience rating, were categorized as continuous variables since data was collected for each user under all experimental conditions. To determine the normality of the original data, the Shapiro-Wilk normality test (using the AS R94 algorithm) was conducted, revealing that the data did not follow a normal distribution. Therefore, the Wilcoxon ranks-sum test, a nonparametric test for comparing two samples, was used for each data analysis. The summarized results of the statistical analysis can be found in Table 13.

Table 13 Results of statistical analyses for the data measured in US 2. The asterisks indicate statistically significant different conditions.

Measurements		Optimized auxiliary model vs. baseline		
		Regular	Elongated	Plate
Objective	Localization time	Z=-1.739 p=0.082	Z=-1.432 p=0.152	Z=-1.538 p=0.124
	Recognition accuracy	Z=-3.000 p<0.001*	Z=-5.000 p<0.001*	Z=-5.831 p<0.001*
Subjective	Ease of localization	Z=-3.609 p<0.001*	Z=-3.871 p<0.001*	Z=-3.927 p<0.001*
	Clarity of localization	Z=-4.321 p<0.001*	Z=-4.230 p<0.001*	Z=-4.143 p<0.001*
	Enjoyment	Z=-3.609 p<0.001*	Z=-3.964 p<0.001*	Z=-3.742 p<0.001*

Regarding the localization time, no statistically significant difference was found between the proposed design of AMs and the baseline across all experimental conditions. Consequently, we rejected the hypothesis H1.

In terms of recognition accuracy, we assessed the error rate for each experimental condition as:

$$\text{ER\%} = \frac{\text{N. errors}}{(\text{N. participants}) * (\text{N. components})} * 100 \qquad (1)$$

We observed a statistically significant reduction in error rate with the proposed design compared to the baseline across all experimental conditions. Fig. 29 presents the data collected, considering the errors from all four replications of each condition. These findings support the confirmation of hypothesis H2.

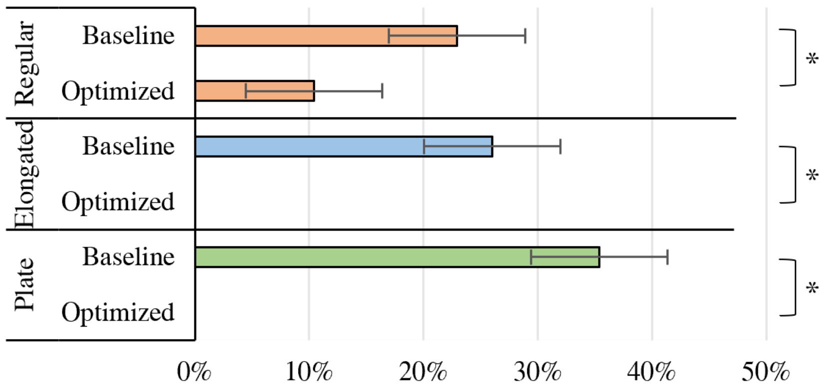


Fig. 29 Error rate for all the experimental conditions. The asterisks indicate statistically significant different conditions.

Regarding the subjective measurements, we gathered data on the ease of localization, clarity of localization, and enjoyment of use. For each of these aspects, we identified a statistically significant improvement in user experience with the proposed design compared to the baseline across all experimental conditions. Fig. 30 displays the comprehensive data collected from all four replications of each condition. These results provide evidence for confirming hypothesis H3.

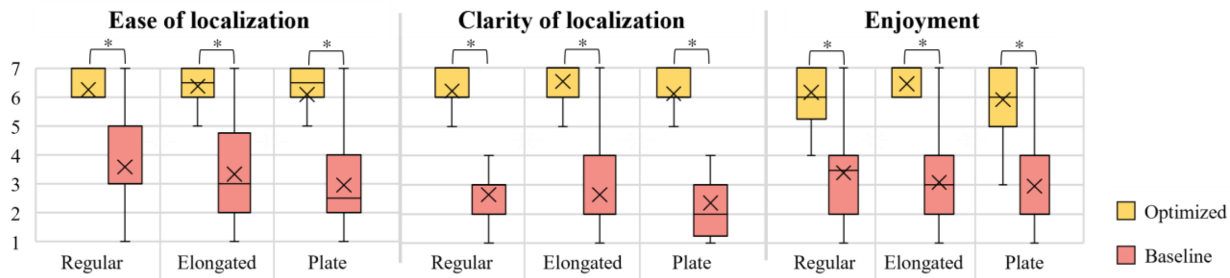


Fig. 30 Median scores for subjective measurements in terms of ease of localization, clarity of localization, and enjoyment for all the experimental conditions. The asterisks indicate statistically significant different conditions.

The results of US 2 validated the guidelines formulated in US 1 regarding both performance and user experience.

In terms of user performance, although we did not find a statistically significant difference in localization time, we observed a significant improvement in recognition accuracy when using optimized AMs compared to the baseline. This improvement is because the AMs designed according to our guidelines effectively isolate the correct component, minimizing any confusion with nearby ones. In contrast, 3D arrows make it difficult to distinguish the target component from others in close proximity or the same line of sight. Another factor that contributed to the increased recognition accuracy with optimized AMs was the inclusion of a blinking animation.

This animation proved crucial in helping users identify the correct component, as it eliminated occlusion issues that can occur when using a static AM.

The findings from analyzing the results on recognition accuracy were further supported by evaluating the user experience. The subjective measurements (ease of localization, clarity of localization, and enjoyment of use) received significantly higher ratings when using our optimized AMs compared to the baseline across all component shapes. Users preferred using the optimized AMs because they enabled easy and unambiguous recognition of the target component without confusion with nearby components. Additionally, from a hedonic perspective, users reported that the optimized AMs made the task more engaging and enjoyable.

6.1.5. Lessons learned

The results of this study provide a positive answer to our research question: "*Is it possible to design optimized auxiliary models for locating in-view, not occluded components according to their shape?*" US 1 demonstrated that there is no single recommended shape of AM for different component shapes to localize. However, using yellow color and blinking animation proved beneficial and can be applied universally. Based on the findings of US 1, we proposed guidelines for designing optimized AMs to locate in-view and not occluded components with regular shapes in complex machines. These components can be effectively represented by their bounding box, and the recommended shape of the AM depends on the aspect ratio of the bounding box. US 2 confirmed that using the optimized AMs led to improved performance and user experience compared to generic 3D arrows. Although our case study focused on complex industrial equipment, the results can apply to other industrial scenarios (such as electrical panels) and other fields involving localization tasks where objects in close proximity can be easily confused when using 3D arrows.

Previous studies have demonstrated that peripheral cueing, which involves salient spatial cues at the relevant position, leads to faster attentional shifts compared to central cueing, which involves symbolic cues like arrows. This is because central cueing requires additional time to interpret the symbolic cue [10]. However, the pure localization task addressed in our study is relatively simple and quick to accomplish. Therefore, we did not observe a significant improvement in localization time when using optimized AMs. Nevertheless, users reported that localizing components using the optimized AMs was easier compared to using 3D arrows. The most significant result of this study regards the accuracy of the localization task. By using *delimiting* AMs with a shape geometrically similar to that of the component, there was less ambiguity in recognizing the target component compared to using arrows. In fact, arrows indicate a point in space but, in complex machines, this point may correspond to multiple components close to each other. Moreover, arrows in AR interfaces serve various purposes beyond localizing in-view components, such as indicating out-of-view objects [82], [124], providing navigation directions [161], or functioning as interaction buttons [113]. Consequently, operators may misinterpret the information associated with the arrow.

An interesting finding in this study concerns the use of animations. While they do not provide additional information for the localization task, the results revealed that the blinking animation can have several beneficial effects. Firstly, it helps to capture users' attention towards the components, enhancing their focus on the task. Additionally, the blinking animation assists in overcoming occlusion issues that can arise with 3D AMs, offering a valuable alternative to techniques such as transparency [135], [136] and wireframe rendering [137] used in prior studies. Lastly, including animations improves operators' overall enjoyment when performing procedural tasks.

Yellow color is also beneficial in directing user attention compared to other colors. Yellow contrasts effectively with machine components, which are typically metallic and tend to have dominant gray and dark colors.

These findings regarding animations and colors align with the research presented in [162], which suggests that salient attributes such as movement and color can effectively serve as peripheral cues, improving attention guidance. Moreover, our study demonstrated that using colored and animated AMs enhances operators' enjoyment. This result carries significant importance from the perspective of Industry 5.0, as it emphasizes the well-being of industrial workers as a central aspect of the production process [163].

It should be considered that this study focused specifically on the localization of in-view and not occluded objects. In fact, there have been previous studies that have addressed techniques for drawing user attention to out-of-view objects, such as using 3D arrows [124], [142], pulsing 3D halos [124], or a 2D bar [164]. These visualization methods can be easily integrated with the AMs designed based on our guidelines. When an object is out-of-view, one of these visualization methods can direct user attention to the object. Once the object becomes in-view, an optimized AM can locate it precisely. However, there is still a possibility that, even though the object is in-view, it may be partially or completely hidden by other objects. In such cases, specific rendering techniques like alpha blending [165], [166] can be applied to the AMs to allow users to understand the component the AM refers to. An alternative solution to localize components in these blind areas could be to use exogenous cues, such as presenting a side-by-side 3D model or a virtual mirror, as proposed in our work [167] presented in the following subsection.

6.2. Augmented Reality Technical Documentation in blind areas[6]

Another important factor that must be considered when designing AR instructions in a real industrial context is the presence of blind areas [168], i.e., areas in which machine components block workers' line of sight. Important information may be lost when components are not clearly visible. As a result, workers may find it challenging to understand what they need to do, leading to longer completion times, reduced accuracy, and increased mental effort. This research aims to compare different solutions to provide AR instructions in blind areas, measuring user performance in accomplishing a real maintenance task.

AR procedural instructions are typically presented as 3D models [8] to guide users in tasks like assembling, removing, or inspecting components. However, a challenge arises when these 3D models are spatially registered (in-situ content [123], [169]) in blind areas causing occlusion and depth conflicts. However, this approach is helpful because it automatically draws the user's attention to the relevant positions, providing spatial cues [10]. On the other hand, an alternative solution is side-by-side content [123], which is not directly aligned with the working area but displayed separately beside it. This method allows placing the AR instructions in non-blind regions but requires users to constantly switch their focus between the instructions and the working area. Previous studies [123], [170] comparing in-situ and side-by-side AR instructions for assembly tasks found that in-situ instructions performed better. However, maintenance tasks differ from assembly tasks as they often require working on fully assembled machines with components that may be completely hidden during the task. In fact, they involve restoring the performance of a system, equipment, or product [171]. Besides assembly, maintenance tasks involve more tasks, such as recognizing and removing the target component from the equipment [172]. In the literature, no specific studies are comparing in-situ and side-by-side AR content designed for conveying maintenance instructions in blind areas.

[6] The results of the study described in this Section were published in the following papers:
- E. Laviola, M. Gattullo, A. Evangelista, M. Fiorentino, and A. E. Uva, "In-situ or side-by-side? A user study on augmented reality maintenance instructions in blind areas," Comput. Ind., vol. 144, no. June 2022, p. 103795, 2023, doi: 10.1016/j.compind.2022.103795.
- E. Laviola, S. Romano, and M. Gattullo, "Virtual Mirror for Maintenance: an Augmented Reality Support Tool," in Design Tools and Methods in Industrial Engineering, 2023.

Addressing the occlusion between virtual and real objects becomes crucial to ensure proper comprehension of maintenance instructions in areas where the view is obstructed. The occlusion problem has two main effects on information understanding. Firstly, when there is partial occlusion between the in-situ instructions and real objects, it becomes challenging for operators to determine the relative depth of both virtual and real objects. Secondly, if real components entirely hide the in-situ instructions due to being in a blind area, operators may struggle to identify which component they are meant for. These two issues are referred to as "the order problem" and "the X-ray vision problem," respectively, as defined by Macedo and Apolinario [173]. When designing an AR interface for maintenance, it is essential to propose presentation methods that address both problems simultaneously. In some cases, specific instructions may relate to components in blind areas, while others do not. Macedo and Apolinario found that using a geometric virtual representation of real objects, often called a "phantom model," is a popular strategy to solve the occlusion order problem. However, this approach alone is insufficient to resolve the X-ray vision problem. In fact, in-situ instructions in blind areas would be completely hidden by the phantom model. Therefore, the X-ray vision problem can be solved by controlling the alpha blending between real and virtual content. Additionally, combining alpha blending with other perceptual cues such as color, edges, and texture further enhances the sense of depth without sacrificing the visibility of the virtual objects [148], [165].

Despite numerous proposed approaches, dealing with occlusions with in-situ AR instructions remains an unresolved problem [36]. Furthermore, the impact of these techniques on a user's ability to perform procedural tasks in a real industrial context is not yet fully understood [123], [168], [170], [174]. Hence, we aim to address the research question: "*How does the information provided through AR content affect user performance for maintenance tasks in blind areas?*" To answer this research question, we developed and assessed three AR presentation modes that facilitate manual maintenance procedures involving components located in blind areas. The three presentation modes are as follows: 3D "in-situ" rendered with X-ray technique (Fig. 31a), "3D side-by-side" on a CAD model replicating the blind area (Fig. 31b), "2D side-by-side" on a virtual mirror (Fig. 31c).

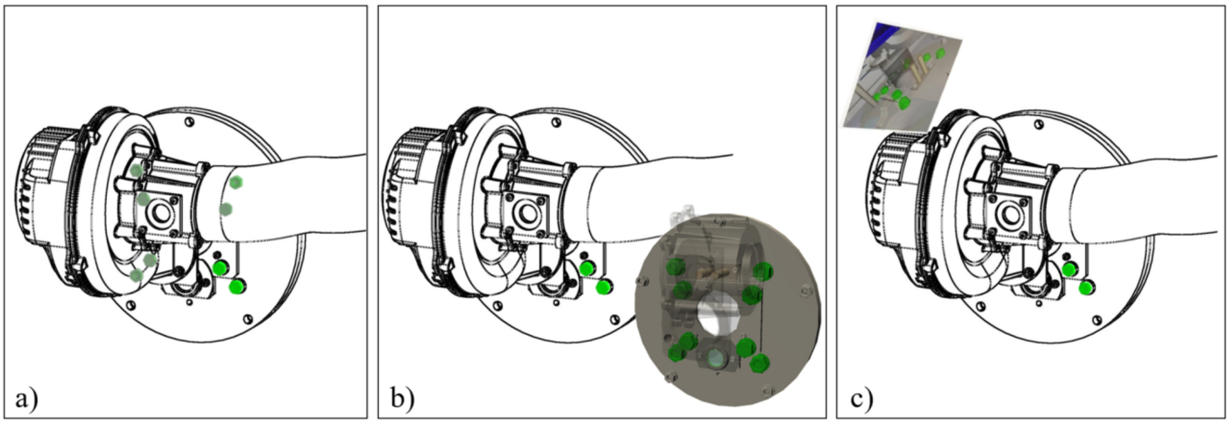


Fig. 31 The AR presentation modes evaluated for maintenance instruction in blind areas: (a) "in-situ" X-ray 3D models, (b) "3D side-by-side" with 3D models of the working area, (c) "2D side-by-side" with a virtual mirror for hidden components.

We conducted a user study where we measured user performance for the three presentation modes in terms of completion time, accuracy, and cognitive load. Participants were instructed to perform a bolt-disassembly task derived from the maintenance manual of a gas-fired radiant heating system used for winter climate control in large areas such as industrial premises, warehouses, hangars, and railway workshops. User performance with the three proposed presentation modes was compared with that obtained by presenting the disassembly instruction on a drawing taken from the PDF documentation of this machinery.

6.2.1. Related work

The problem of displaying instructions in blind areas for IAR interfaces is a topic that remains unresolved in the existing research. Many studies have focused on developing algorithms and technical frameworks to address two main challenges: the "order problem" and the "X-ray vision problem." These issues are addressed separately in the literature, as pointed out by Macedo and Apolinario [173]. To tackle the order problem, various approaches have been explored. Breen et al. [175] developed model-based and depth-based methods to calculate interactive occlusions, which deal with determining the correct order of virtual objects in relation to the real scene. Berger [176] introduced a contour-based approach involving a 2D computation to check the position of virtual objects. On the other hand, other researchers have worked on a 3D reconstruction method, where the depth maps of the real scene are combined with a 3D point cloud alignment. This technique, proposed by [177], handles mutual occlusions between real and virtual objects. Several other algorithms have been proposed to address the order problem by performing depth computation for all pixels [178] and combining previously suggested approaches [179]. Regarding objects that are completely hidden or occluded in blind areas, multiple techniques have been developed to render these hidden objects, both with and without photo-realism. Macedo and Apolinario [173] presented various techniques, including alpha blending, which adjusts the object appearance based on human transparency perception as studied in [165], and intensity settings applied to the object color [166]. Furthermore, Kalkofen et al. [180] investigated a method that extracts hidden object edges to visualize them. Additionally, Bajura et al. [181] and Sandor et al. [182] proposed virtual windows and saliency maps as other methods to visualize hidden objects.

Despite the numerous methods developed to address the order and X-ray vision problems in IAR scenarios, there is a notable lack of studies investigating how these proposed techniques impact user performance during procedural tasks in a real industrial context. Only a few research papers have delved into the relevance of occlusion handling for information comprehension, specifically in industrial AR scenarios, and even these studies were confined to assembly tasks that solely tackled the order problem. For instance, Macallister et al. [36] conducted a study comparing three different in-situ visualization techniques that displayed partially occluded parts during an assembly task. While the authors provided the same in-situ assembly instructions for both occlusion-handling approaches (using X-ray augmentation) and those without, they did not find statistically significant differences in completion time and errors across the various presentation modes. However, there remains a significant need for further research to confirm these findings in various other industrial contexts, such as maintenance tasks.

An alternative approach to dealing with occlusions with in-situ AR instructions in blind areas is using side-by-side content. Two studies, one by Khuong et al. [170] and another by Blattgerste et al. [123], compared in-situ AR instructions with side-by-side 3D AR content in a Lego Duplo assembly scenario, yielding somewhat conflicting results. Due to improved registration accuracy, Khuong et al. demonstrated that side-by-side AR instructions performed better than the in-situ overlay. On the other hand, Blattgerste et al. found that in-situ AR instructions outperformed the side-by-side variant in terms of both errors and completion time. At the same time, there was no significant difference in cognitive load related to viewpoint change. However, it is important to note that both conclusions pertained to an assembly scenario where the X-ray vision problem did not affect the AR instructions. Chu and Ko [174] developed an interface for assembling components completely hidden from the operator's view. In a simple desktop computer assembly task, they found no significant difference in performance (time, accuracy, cognitive load) between a side-by-side 2D animation and in-situ 3D instructions. Feng et al. [168] proposed showing the target position of the assembly along with the hand movement for a bolt-assembly in a blind area. They conveyed this information through both a side-by-side video and in-situ virtual information, where the X-ray vision problem was addressed using gradient transparency in the area around the fixation point. Their results revealed no significant difference between side-by-side and in-situ instructions regarding completion time, accuracy, and usability. However, it was found that side-by-side content led to a significantly higher cognitive load compared to the in-situ approach.

A virtual mirror is a type of side-by-side content that offers an extra perspective of the working area. While real mirrors are commonly used in maintenance operations in blind areas, the adoption of virtual mirrors in AR maintenance interfaces has been relatively limited, despite being widely used in medical applications. Navab et al. [183] and Bichlmeier et al. [184] introduced the concept of a virtual mirror in augmented laparoscopic surgery to address the challenge of perceiving 3D depth and shape when rendering virtual volume data, such as organs and blood vessels. More recently, Martin-Gomez et al. [185] proposed an "augmented mirror" that reflects both the virtual and real content of an AR scene. Augmented mirrors offer an additional viewpoint, making them versatile for various applications, including visualization, exploration, and alignment. However, it is worth noting that using real mirrors presents its own set of challenges. Occlusion issues can arise when a real-world object is positioned between the observer and the mirror or between the mirror and a virtual object, hindering a clear view and potentially impacting the effectiveness of the virtual mirror technique.

From the analysis of the state of the art, it can be stated that:

- Many works focus on the creation of algorithms for occlusion management, trying to solve only “the order problem” [175]–[179] or “the X-ray vision problem” [165], [166], [173], [180]–[182].
- Few studies explore in-situ AR content for occlusion handling, and they aim to solve only “the order problem” [36].
- Few studies explore side-by-side AR content for occlusion handling, and they aim to solve only “the X-ray vision problem” [183]–[185].
- It is still not clear which is the best solution to solve both “the order problem” and “the X-ray vision problem” for occlusion handling and how they affect user performance [123], [168], [170], [174].

Based on the previously discussed studies, there is a research gap in displaying AR instructions for components in blind areas. Additionally, most of these studies have focused on assembly tasks, which are less sensitive to the occlusion problem compared to maintenance tasks. Therefore, the primary contribution of our work is to address this research gap by comparing user performance using different AR presentation modes for conveying instructions during maintenance tasks carried out in blind areas. Unlike previous works, our proposed presentation modes, both in-situ and side-by-side, were intentionally designed to simultaneously tackle both the "order problem" and the "X-ray vision problem."

6.2.2. Material and Methods

To assess the three different AR presentation modes to convey maintenance instructions in blind areas, we conducted a user study. We based it on a real maintenance scenario provided by Upgrading Services SpA, a company involved in a research project to develop an AR maintenance manual for a gas-fired radiant heating system. For this purpose, they gave us the real machine and its PDF maintenance manual. From the provided manual, we selected a maintenance task involving identifying and removing screws necessary to disassemble the appliance burner for cleaning (Fig. 32).

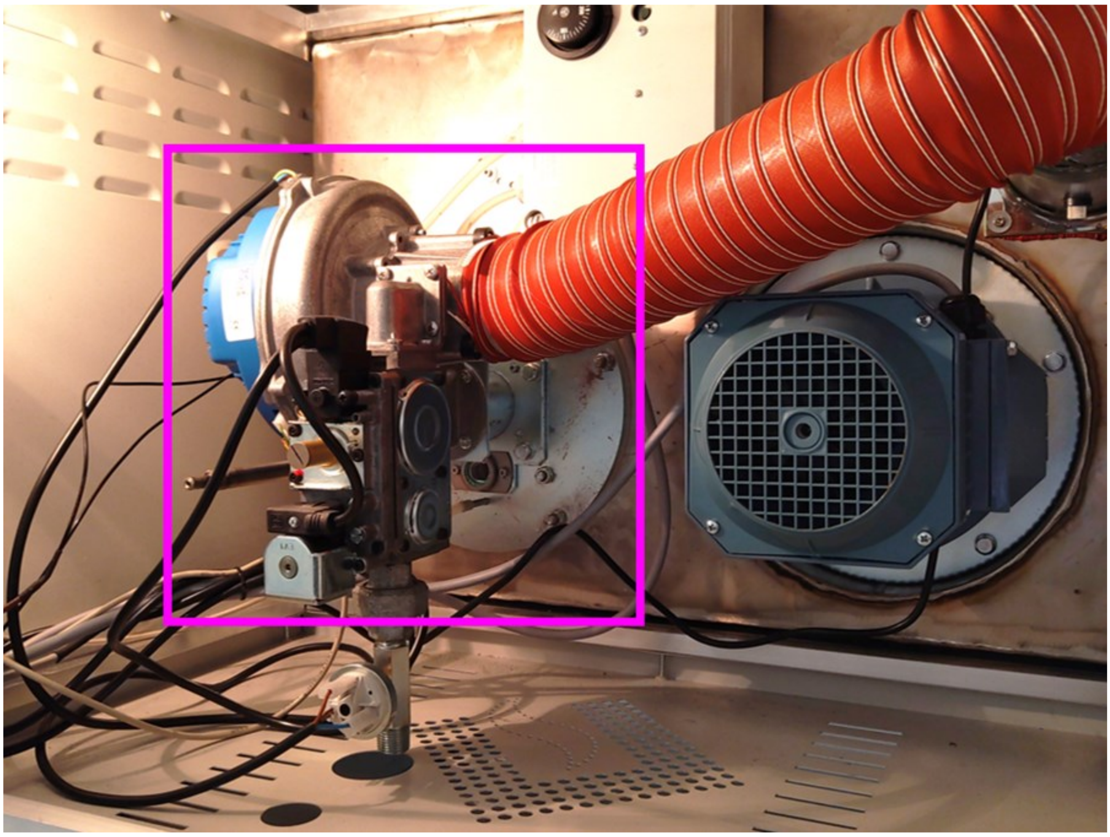


Fig. 32 Machine (gas-fired radiant heating system) used for the experiment; the blind area for the maintenance task is highlighted.

We divided this main task into eight steps, each requiring the identification of one screw. This case study represents our target scenario because the screws involved are in a blind area behind the burner. As a result, operators may encounter difficulties locating some of these screws during the task. If they mistakenly remove the wrong screw, it can lead to safety problems as the component might not be secured properly. Additionally, if the operators realize their mistake, they will waste time disassembling and reassembling the equipment.

The AR interface was displayed using the Microsoft HoloLens 2 [105], an OST device. To build this interface, we used the Unity 3D Engine [108] along with the MRTK [109] and Vuforia Engine [110] for tracking. An image target was designed and placed inside the right panel of the gas-fired radiant heating system. Leveraging the 3D CAD model of the entire equipment in Unity, we created the AR scenes to be displayed on the HoloLens. To facilitate interaction with the AR content, we integrated HoloLens voice input, allowing users to give commands directly using their voice. To streamline the user experience, we implemented simple voice commands such as "test 1" and "step 2". This enabled users to select a specific experimental test or step without relying on gestures or physical interactions.

In the "in-situ" presentation mode (Fig. 33a), we used an animated CAD model of the screw spatially aligned with the real-world counterpart to provide information on which screw to remove. This approach has previously been used in similar scenarios [8], [186]–[188]. To achieve this, we inserted the 3D CAD model of the burner into the AR scene using a DepthMask Unity shader. The rendering order of occluded virtual components was managed by employing two cameras with different depth properties and culling masks. Each screw was associated with two overlapping CAD models assigned to distinct depth layers, differing only in appearance but not in geometry. To address the X-ray vision issue, we implemented an alpha blending technique inspired by Livingston et al. [166]. This technique involved varying the opacity and intensity to make the virtual objects visible without causing difficulties in perceiving transparency changes on the OST device. Using this rendering approach, all CAD models of the screws were visible from every perspective. In cases where a burner component obstructed the view of the screw, we represented the screw with a desaturated green semi-transparent material (HSV: 120°, 0.50, 1, alpha 0.5). Conversely, if a screw was not occluded and fully visible, we used an opaque green material with full intensity (HSV: 120°, 1, 1, alpha 1). This way, the 3D models in the blind areas were rendered differently from those visible, allowing workers to view AR instructions in blind areas and effectively address depth conflicts with real objects.

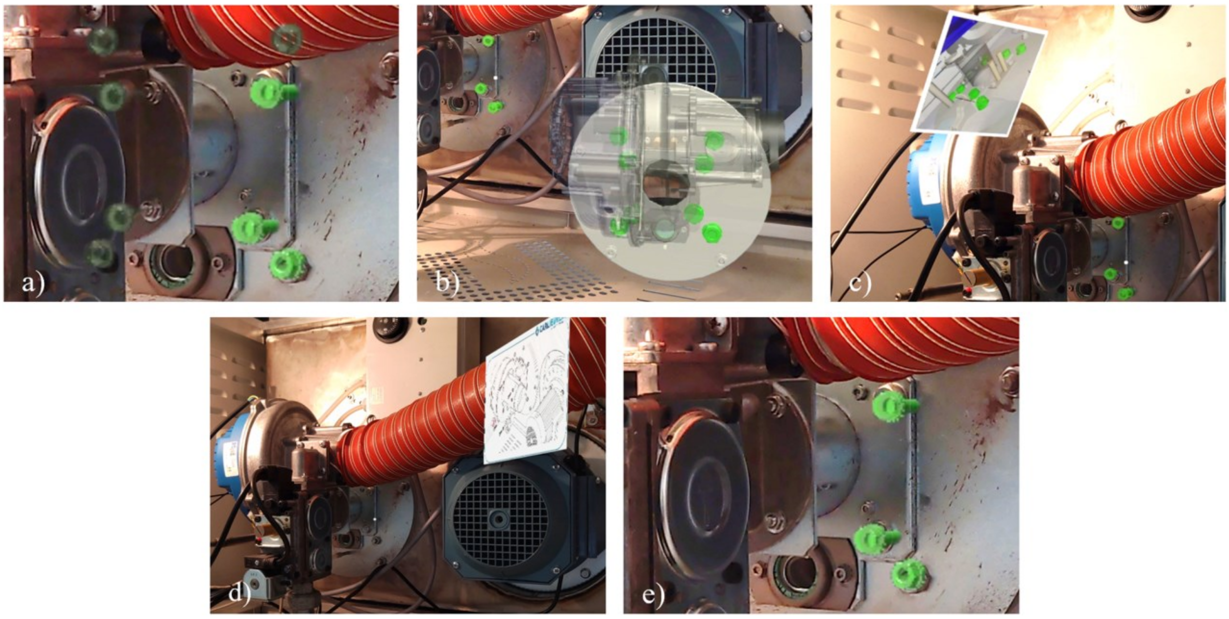


Fig. 33 The AR interface with the presentation modes evaluated in the user study: (a) "in-situ," (b) "3D side-by-side," (c) "2D side-by-side," (d) "baseline," (e) "control." All the virtual screws are shown together only for clarity, but one screw per time was displayed in the user study.

We compared displaying in-situ instructions using two types of side-by-side content: 3D and 2D. Both side-by-side approaches aimed to address the X-ray vision problem by providing workers with an additional perspective on the blind area. In both the designed side-by-side AR presentation modes, in-situ 3D instructions were included using phantom rendering. This means that the 3D models of the screws were always rendered with an opaque green material and matched the real screws. However, the virtual screws might not always be visible from certain angles due to occlusion by a burner component placed in front of them. By employing this approach, workers could benefit from in-situ instructions for visible components while relying solely on side-by-side instructions for components in the blind area.

In the "3D side-by-side" presentation mode (Fig. 33b), the instructions are presented using a 3D visual asset, a CAD model representing the blind area. This approach has been used previously for assembly purposes by [123], [170]. The side-by-side 3D model includes a scaled-down version of the burner area rendered with a semi-transparent shade. Additionally, animated opaque green 3D models of the screws are displayed on this model. As a result, the operator can observe the blind area with all the screws covered by the burner, providing a comprehensive view of the components involved. The CAD model is placed on the equipment floor in a comfortable zone that does not interfere with the operator's main task. The specific location of this model is fixed and determined based on the specific case study.

In the "2D side-by-side" presentation mode (Fig. 33c), the instructions are displayed on a 2D visual asset, a plane resembling a virtual mirror. It is strategically positioned to enable workers to view components and related instructions in the blind area. The idea for this approach originated from real-world industrial contexts where physical mirrors are used, for example, in car repair scenarios where technicians cannot easily access certain areas under the engine. While virtual mirrors have been commonly used for medical purposes, their application in AR maintenance interfaces has been relatively limited. Hence, it remained uncertain whether they could serve as a valid solution for providing AR instructions in blind areas. The presentation mode combines both in-situ 3D models of the screws and a side-by-side virtual mirror. The rationale for using in-situ content and its rendering is similar to the "3D side-by-side" mode. The virtual mirror is designed as a world-fixed virtual plane, displaying a portion of the 3D CAD model of the equipment along with animated CAD models of the screws. The goal was to replicate the experience closely users would have with a real mirror while leveraging the advantages of AR, as previously suggested by [174], [183], [184]. The virtual mirror is a square shape measuring 200 mm × 200 mm, similar to commercially available ones, with clearly visible borders to aid in perceiving its orientation [174]. Its position and orientation are carefully chosen to ensure that all the screws involved in the task are visible. Transparency is used to render the CAD model of the equipment within the mirror, making it possible to view screws that might be occluded by it. Moreover, the content within the virtual mirror responds to the user's head movements, providing a more realistic impression of interacting with a real mirror. To assess the effectiveness of the virtual mirror design, a preliminary evaluation was conducted by comparing it with a real mirror of the same dimensions placed in the same position in our work [189]. The performance of 20 users was measured in a screw localization task. The preliminary evaluation results showed no statistically significant differences between the virtual mirror and the real one in terms of completion time, accuracy, and cognitive load, indicating that the virtual mirror was as effective as its physical counterpart.

The effectiveness of the three proposed presentation modes was assessed by comparing them to a fourth AR mode referred to as the "baseline," which served as the reference point. In the "baseline" mode, a world-fixed drawing (Fig. 33d) was displayed, resembling how instructions for the task are traditionally presented in PDF documentation for the machinery. This drawing represents an explosion diagram, a common method used in traditional documentation for presenting disassembly instructions. The information about the screw to identify is provided using a numbered balloon for one screw at a time. The drawing is positioned close to the main task area but is outside the operator's field of action. Thus, with the HoloLens, we simulated an operator that watches this information on a monitor. Additionally, a fifth AR presentation mode, the "control" mode, was designed and used as a control mode. It solely displays in-situ 3D models of the screws using phantom rendering (Fig. 33e). Evaluating user performance with this "control" presentation mode was necessary to demonstrate that this content alone is insufficient for correctly accomplishing the chosen disassembly task.

The gas-fired radiant heating system was provided by the company that made the equipment available in our laboratory for the experiment before its installation in a plant. We set up the equipment at a height similar to that of the maintenance procedure on the installed equipment. We placed an incandescent lamp to enhance visibility during maintenance, just like operators usually do. We ensured that this lighting condition allowed all the AR presentation modes to be correctly visible with the HoloLens.

We conducted independent test trials to evaluate and compare each AR presentation mode, including the "baseline" and "control" ones. These trials involved the identification of eight screws located behind the burner area of the gas-fired radiant heating system. By designing a simple task for all participants, regardless of their backgrounds, we addressed our research question and gained valuable insights into how the choice of AR content impacts user performance in blind areas. To carry out the experiment, we followed a procedure proposed in [190], which required the installation of bolts in a blind area. The experiment employed a within-subjects design, meaning each participant's data was collected under all five experimental conditions. To minimize any potential order effects or carry-over effects on the data, we employed a balanced Latin Square design to determine the order in which the AR presentation modes were shown to each user. Additionally, to avoid any bias from learning effects, we varied the disassembly order of the screws in each trial, even though the screws remained the same.

At the beginning of each session, the experimenter gathered basic information from the users and provided a detailed explanation of the experimental procedure. The users were also given instructions on how to use the AR application with the HoloLens. Subsequently, the users underwent a training session with the application to familiarize themselves with its functionality, which could last 1 to 3 minutes. During this training phase, the users were exposed to various scenarios, each representing a different AR presentation mode used throughout the experiment. This training familiarized the users with the application and helped them get comfortable using voice input on the HoloLens to proceed to the next steps. Once the users felt adequately trained and confident, the experimenter offered them the option to initiate the experiment by saying which specific AR presentation mode to test.

To begin each trial, users said "start test," after which an AR instruction for the first screw was displayed. The experimenter then instructed the users to locate the screw and physically touch it with the index finger of their hand once they had understood the provided information. This approach aimed to ensure that users' skills and dexterity in workshop operations did not influence their performance. During the trial, any errors made by the users were recorded by the experimenter, who had a clear view of the blind area due to the removal of a side panel from the equipment (Fig. 34). In the event of an error, the experimenter did not notify the users. If users could not recognize a screw, they could say "skip" and proceed to the next step. To access information about the subsequent screw, users pronounced the voice command "step n," where "n" corresponded to the ID step from 1 to 8. When all eight steps were completed, users could conclude the trial by saying "finish test." After each trial, users were asked to complete a NASA TLX questionnaire [191] to assess the perceived cognitive load associated with each specific AR presentation mode.

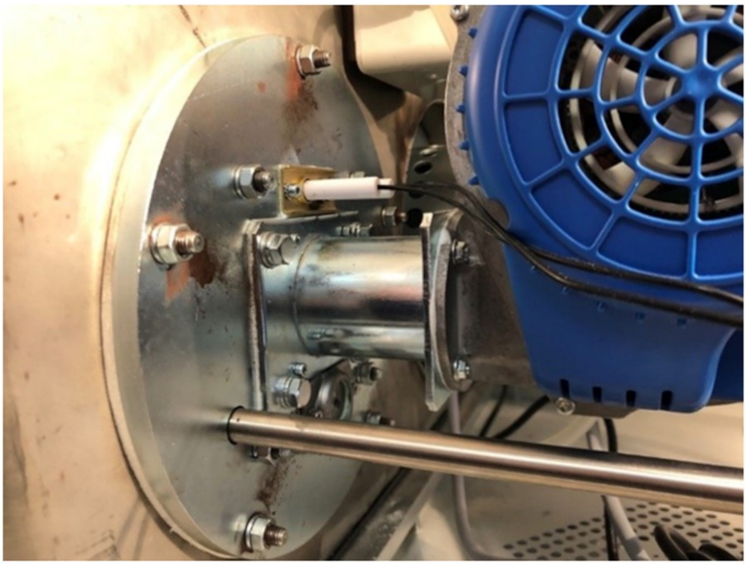


Fig. 34 The blind area of the bolt-disassembly task from the experimenter's point of view.

We recruited a total of 42 participants (6 females, 19-54 years old, mean = 24.7, SD = 6.57) from our university and local companies. As an incentive for participation, each participant received a 10€ coupon to spend in our university store. To ensure the participants' technical proficiency, we ensured that all of them had a background in the relevant field, making the sample representative of novices who could perform AR maintenance procedures. Among the participants, there were 16 bachelor's degree students, 18 master's degree students in engineering, 2 Ph.D. students, 2 post-doc research fellows in mechanical engineering, 1 freelancer, and 3 employed engineers. To ensure accurate perception, we verified that none of the participants were color-blind. On a 7-point Likert scale (1: Not at all familiar - 7: Extremely familiar), the average familiarity level with AR was 2.62 (SD = 1.81, Min = 1, Max = 7). Throughout the experiment, participants were allowed to wear their eyeglasses along with the HoloLens. The average duration of the experiment for each participant was approximately 50 minutes.

We collected data for each trial throughout the experiment, including completion time, accuracy, and cognitive load. The application automatically recorded the completion time for each trial. This time was measured from when users said "start test" to initiate the trial to when they said "finish test" to conclude it. During the execution of each trial, experimenters manually detected and recorded any errors made by the users. These errors were categorized into two types: wrong screws and screws not found. We calculated the weighted NASA TLX score for every trial to assess cognitive load and recorded it for each user. Additionally, we tracked the movement of the users' pupils using the built-in eye tracker of the HoloLens 2. The eye-tracking data was visualized through MRTK heatmaps displayed in the working area. Using a series of gradient-colored dots, these heatmaps helped identify areas that attracted more of the user's visual attention.

6.2.3. Results

We followed the approach outlined in [192] regarding the statistical analysis. We performed a one-way repeated measures analysis of variance (ANOVA) at a 95% confidence level to analyze the completion time data. However, the original time data were found not normally distributed according to the Shapiro-Wilk test. To address this, we applied a Box-Cox transformation which allowed us to obtain normal samples. We also verified that the assumption of sphericity was not violated. We conducted paired t-tests as post hoc tests with the Bonferroni correction to identify statistically significant differences between each couple of AR presentation modes. For the errors and cognitive load data, we used non-parametric tests. Specifically, we employed the Friedman test, an alternative to the one-way repeated measures ANOVA, at a 95% confidence level. Subsequently, Wilcoxon signed-ranked tests were used as post hoc tests, and the Bonferroni correction was applied to identify statistically significant differences between different AR presentation modes.

In our analysis, we discovered a highly significant difference in mean completion time (Fig. 35) among the different AR presentation modes ($F(4, 164) = 46.721$, $p < 0.001$). Upon conducting post hoc tests (Table 14), we observed that users performed notably better with both the "in-situ" and "3D side-by-side" modes when compared to the "baseline" mode. However, there was no significant difference in completion time between the "baseline" and "2D side-by-side". Additionally, we found a statistically significant reduction in completion time when using the "in-situ" presentation mode compared to both the "2D side-by-side" and "3D side-by-side."

Moreover, users were significantly faster when using the "3D side-by-side" mode in comparison to the "2D side-by-side." As expected, the "control" mode had the slowest completion times, significantly slower than both the "in-situ" and "3D side-by-side," while no significant differences were observed when compared to the "2D side-by-side."

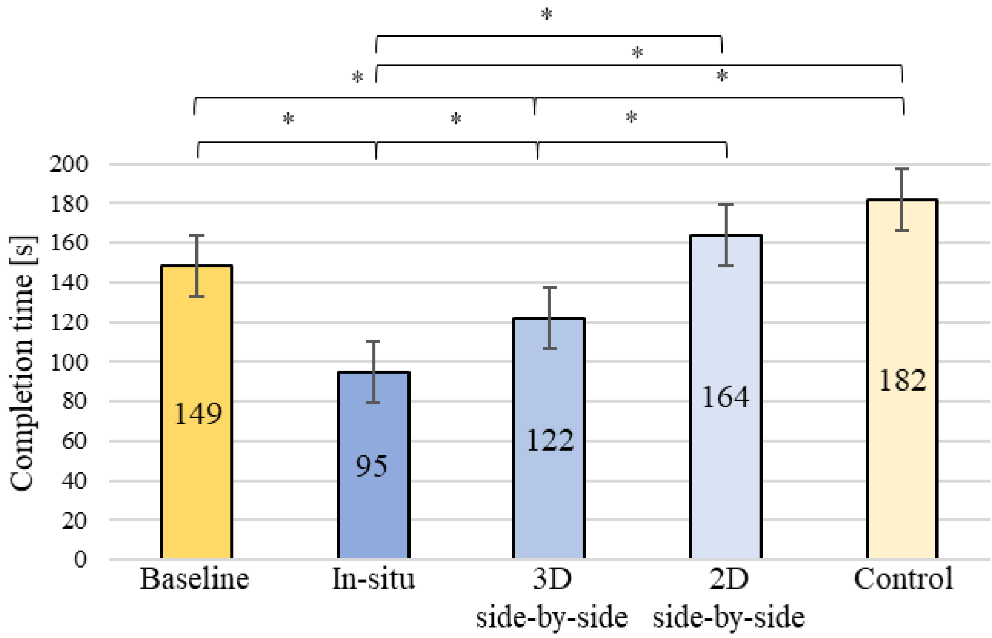


Fig. 35 Mean completion time for the AR presentation modes. The asterisks indicate a statistically significant difference between the two compared conditions.

Table 14 Results of the post hoc pairwise comparisons of the presentation modes for the completion time. The asterisks indicate a statistically significant difference between the two compared conditions.

	In-situ	3D side-by-side	2D side-by-side	Control
Baseline	p<0.001*	p=0.005*	p=0.974	p=0.053
In-situ	-	p<0.001*	p<0.001*	p<0.001*
3D side-by-side	-	-	p<0.001*	p<0.001*
2D side-by-side	-	-	-	p>0.999

We measured the error rate for each presentation mode as:

$$ER\% = \frac{\text{N. errors}}{(\text{N. screws})} * 100 \quad (2)$$

We observed a statistically significant difference in error rates (Fig. 36) among the various AR presentation modes ($\chi2(4) = 43.766$, $p < 0.001$). Subsequent post hoc tests (Table 15) revealed that users exhibited significantly higher accuracy when using both the "in-situ" and "3D side-by-side" compared to the "baseline." However, no significant differences in error rates were observed with the "2D side-by-side" compared to the "baseline." Furthermore, we noticed a statistically significant increase in error rates with the "2D side-by-side" presentation mode when compared to both the "in-situ" and "3D side-by-side" modes. However, there was no statistically significant difference between the "in-situ" and "3D side-by-side". As anticipated, the "control" mode resulted in significantly more errors than both the "in-situ" and "3D side-by-side," while no significant difference was observed when compared to the "2D side-by-side" mode. Notably, 67% of the "control" mode errors were attributed to screws not being found. This finding confirms that the disassembly task selected for this study accurately represents the issue of the blind area.

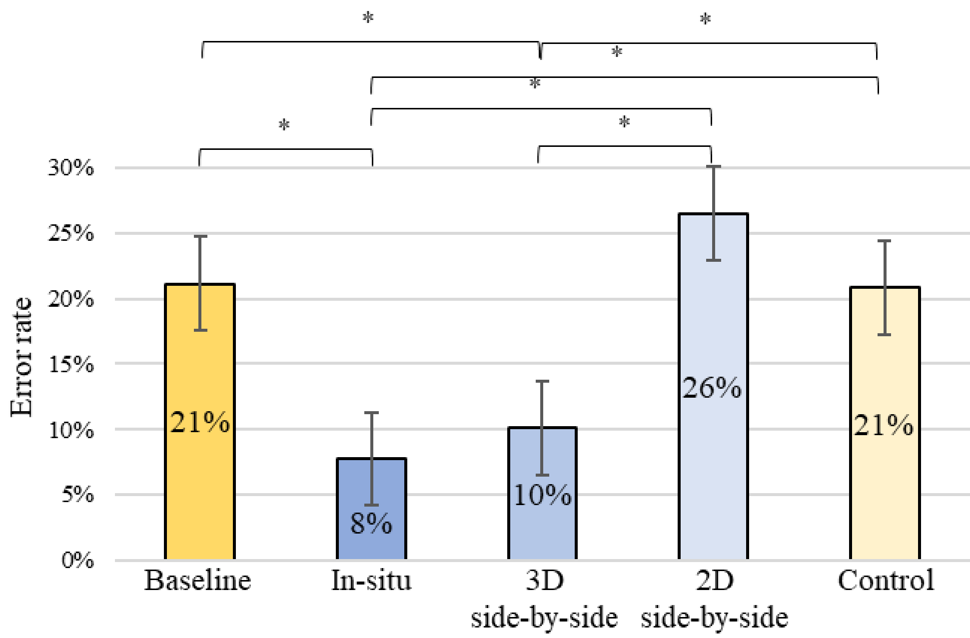


Fig. 36 Mean error rate for the AR presentation modes. The asterisks indicate a statistically significant difference between the two compared conditions.

Table 15 Results of the post hoc pairwise comparisons of the presentation modes for the error rate. The asterisks indicate a statistically significant difference between the two compared conditions.

	In-situ	3D side-by-side	2D side-by-side	Control
Baseline	p<0.001*	p=0.001*	p=0.174	p=0.979
In-situ	-	p=0.345	p<0.001*	p<0.001*
3D side-by-side	-	-	p<0.001*	p=0.002*
2D side-by-side	-	-	-	p=0.150

We identified a statistically significant difference in cognitive load (Fig. 37) among the different AR presentation modes ($\chi2(4) = 65.530$, $p < 0.001$). Subsequent post hoc tests (Table 16) revealed that there was a significant reduction in cognitive load with both the "in-situ" and "3D side-by-side" compared to the "baseline." However, no significant difference in cognitive load was observed when comparing the "2D side-by-side" to the "baseline." Furthermore, we found that the "2D side-by-side" presentation mode led to a statistically significant increase in the weighted NASA TLX score compared to both the "in-situ" and "3D side-by-side" modes. However, there was no statistically significant difference in cognitive load between the "in-situ" and "3D side-by-side"; they both resulted in a medium level of cognitive load [193]. As expected, the "control" mode induced a significantly higher cognitive load than both the "in-situ" and "3D side-by-side" modes, while no significant differences were observed when comparing it to the "2D side-by-side."

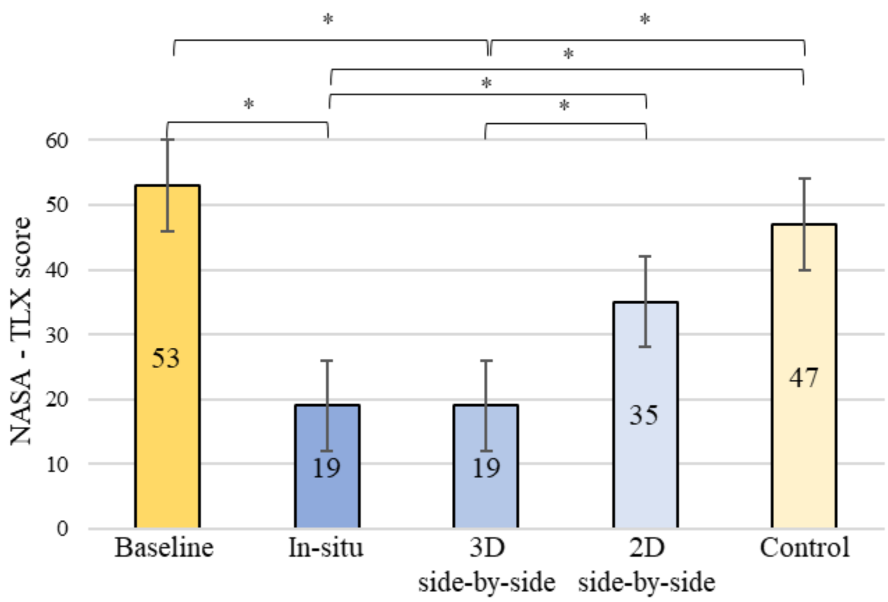


Fig. 37 Mean weighted NASA TLX scores for the AR presentation modes. The asterisks indicate a statistically significant difference between the two compared conditions.

Table 16 Results of the post hoc pairwise comparisons of the presentation modes for the weighted NASA TLX score. The asterisks indicate a statistically significant difference between the two compared conditions.

	In-situ	3D side-by-side	2D side-by-side	Control
Baseline	p<0.001*	p<0.001*	p=0.586	p=0.530
In-situ	-	p=0.115	p<0.001*	p<0.001*
3D side-by-side	-	-	p<0.001*	p<0.001*
2D side-by-side	-	-	-	p=0.086

6.2.4. Lessons learned

The user study provided valuable insights into our research question: "*How does the information provided through AR content affect user performance for maintenance tasks in blind areas?*" Results indicated that two of the three proposed techniques, "in-situ" and "3D side-by-side," effectively displayed AR maintenance instructions in blind areas. Both presentation modes showed better performance in terms of completion time, accuracy, and cognitive load compared to the "baseline," which involved using a drawing in PDF documentation of the equipment. Moreover, the "in-situ" mode outperformed the "3D side-by-side" in terms of completion time, making it a promising choice for future implementations of AR maintenance interfaces in blind areas. This finding is significant since, to the best of our knowledge, no prior study had directly compared the effectiveness of 3D side-by-side instructions with in-situ instructions for maintenance tasks in blind areas.

Prior studies have suggested employing alpha blending to enhance the visual coherence of an AR scene, providing a smooth transition between the visualization of visible and occluded objects [165], [166], [194], [195]. Our research builds upon this concept and provides new insights by showing that using alpha blending to render AR in-situ instructions does not negatively impact user performance during maintenance tasks. This finding is particularly significant because it was not previously demonstrated in a real industrial scenario. It highlights that workers can comprehend disassembly instructions even when they cannot match the virtual component visually with the corresponding real object.

In the case study conducted for this research, there were multiple screws within the user's field of view, positioned at different depths. Despite this complexity, the application of alpha blending proved effective in preventing users from becoming confused about which screw to remove. For future applications, it is possible to enhance the clarity of information further by employing varying levels of opacity and intensity to create distinct depth layers, as suggested by [166]. By implementing this approach, workers would find it easier to understand the instructions when dealing with the same type of component situated at different depths, a common occurrence when dealing with screws or similar items.

Another notable finding from this study is that integrating in-situ instructions displayed using phantom rendering alongside a side-by-side 3D model proves beneficial for maintenance tasks. The results confirmed that relying solely on phantom rendering ("control" mode) is inadequate for successfully completing a disassembly task. However, when instructions are provided on both phantom rendering and a side-by-side 3D model, there is a significant improvement in user performance. This observation contradicts the findings of a previous study [123] that focused on assembly tasks involving parts that others could partially hide and for which phantom rendering was helpful. The reason for this discrepancy likely lies in the presence of blind areas in maintenance tasks, which are not encountered in assembly procedures. In fact, in maintenance tasks, blind areas hinder the usefulness of phantom rendering because crucial information remains hidden, making it impractical for such scenarios.

The proposed "3D side-by-side" mode enables operators to use in-situ instructions when visible and switch to the side-by-side 3D model instructions for components located in blind areas. This approach aims to minimize the need for operators to move their heads between the working area and the location of the side-by-side content. However, analyzing the eye-tracking heatmaps (Fig. 38) for the "3D side-by-side" mode, it becomes evident that many users focus more on the side-by-side instructions than the in-situ instructions. The shift of attention from the side-by-side content back to the working area, along with the mental matching process between the 3D model displayed and the real burner might be the primary reasons why users were slower when using the "3D side-by-side" mode compared to the "in-situ" mode, as previously suggested in [126]. Despite these drawbacks, the study found no statistically significant differences in accuracy and cognitive load between the two modes.

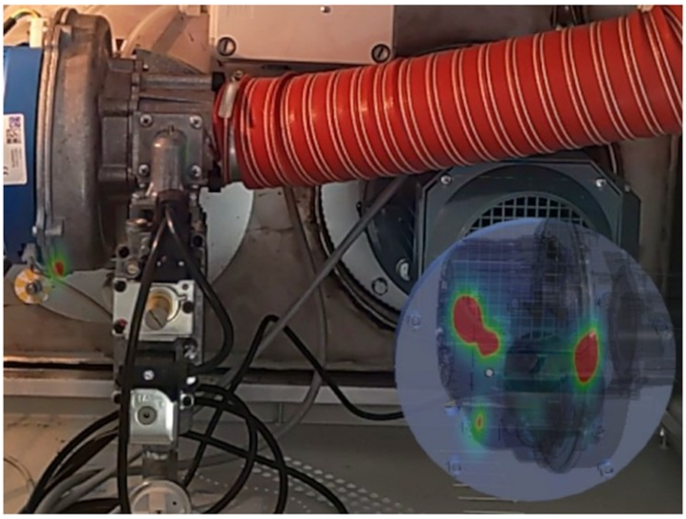


Fig. 38 Eye-tracking heatmap for a user in the “3D side-by-side” mode; red areas indicate the most attention capturing elements.

Using a virtual mirror ("2D side-by-side") to display hidden instructions was ineffective in the proposed scenario. User performance with this presentation mode was notably worse than the "in-situ" and "3D side-by-side" modes. Furthermore, there was no statistically significant improvement compared to the "baseline." The main reason for this outcome was the lack of hand-eye coordination. Users could not see their hand movements in the virtual mirror, leading to difficulties in performing the tasks accurately. To anticipate these challenges, a preliminary experiment was conducted by us [189] to compare user performance in a screw localization task with both the virtual mirror and a real one. This preliminary evaluation showed no statistically significant differences in user performance, leading to using the virtual mirror in the main experiment. However, it is essential to note that the users selected for the main experiment were not accustomed to using mirrors, unlike experienced workers who might be more familiar and skilled with mirror usage. Consequently, participants misunderstood the information regarding which screws to remove. Observations indicated that 42% of the errors associated with the virtual mirror resulted from this kind of misinterpretation. As a result, further research is necessary to propose additional visual cues within the virtual mirror, specifically to assist inexperienced workers with hand-eye coordination.

Similar to our study, Chu et al. [174] and Feng et al. [168] also focused on developing AR functions to assist with manual assembly tasks in blind areas. Although their comparison of in-situ instructions with side-by-side content differs from our approach, their findings partially align with ours, specifically concerning the "3D side-by-side" and "in-situ" conditions. They found no significant performance differences between side-by-side and in-situ AR instructions. However, unlike our results, they did not find a significant difference in completion time. This discrepancy could be attributed to the placement of their side-by-side content, which was much closer to the working area compared to our side-by-side 3D model. Interestingly, both studies proposed using virtual hands in their in-situ instructions to provide the operator with information about the relative position between his/her hand and the assembly interface.

Despite the potential reduction in cognitive load demonstrated by Feng et al. [168], using virtual hands for hand tracking in blind areas is not feasible in a real maintenance scenario, as proposed in our study. Feng et al. tracked the operator's hand by employing a Leap Motion sensor, but this approach has limitations. The sensor requires a cable connection with the computer and relies on an unobstructed line of sight between the sensor and the hand, which may not be practical in real maintenance environments. In the context of our work, providing hand tracking in blind areas is challenging due to potential obstacles and occlusions that naturally occur in maintenance settings. As a result, it becomes impractical to implement hand tracking solutions that rely on external sensors and unobstructed views.

The results presented in this study were obtained using OST HMDs, specifically the HoloLens 2. However, it is reasonable to assume that these findings can be generalized to VST displays, whether head-worn devices like the Varjo XR-3 or handheld devices like tablets. VST displays offer the advantage of better control over the transparency of virtual elements by using the video stream from the device camera, as detailed in [196]. In a VST display, the "in-situ" mode would likely exhibit even more effectiveness than in an OST display, as the virtual elements would not blend with real objects, avoiding potential changes in appearance due to varying lighting conditions. Nevertheless, despite these advantages, we excluded VST head-worn displays due to concerns related to latency and security issues, as indicated in [197]. Additionally, VST handheld devices were not considered as they would not leave the operator's hands free, which could potentially hinder operator performance, despite offering superior display quality.

Using OST HMDs in our study may limit the generalizability of the results to indoor working scenarios. In our case study, we recreated the working environment within our controlled laboratory setting, where we used an incandescent lamp to regulate the lighting conditions. However, in outdoor environments, natural lighting is variable, and the impact of the proposed visualization modes on user performance may differ for OST devices. Specifically, transparent virtual elements could become challenging to perceive in high lighting conditions, making the "in-situ" mode less suitable. Despite conducting the study in a laboratory, the selected case study is still highly representative of the target scenario. Gas-fired radiant heating systems are commonly installed inside industrial warehouses, often positioned on the roof (Fig. 39). We tried to replicate real working conditions as closely as possible, ensuring that the height relative to the users in the laboratory setup resembled the maintenance procedures performed on the real installed equipment.

Fig. 39 Example of installation of a gas-fired radiant heating system (in red) on the roof of an industrial warehouse.

Chapter 7. Development of authoring systems for technical documentation

This conclusive Chapter describes the systems we developed to simplify the authoring based on the previous guidelines identified by our "minimal AR" approach and the optimizations investigated for the information type LOCATION and the visual assets design for blind areas. The aim is to simplify the ARTD authoring process for developers.

7.1. Automating auxiliary model selection for LOCATION: a preliminary exploration[7]

The design guidelines presented in the study of subsection 6.1 could be implemented in AR authoring tools to automatically suggest AMs for conveying localization information. Using the CAD model or point cloud of a component makes it possible to determine its bounding box and calculate the shape factors (r, k, and V). Based on these values, the authoring tool can display the AM in the interface with the position, orientation, and scale automatically adjusted according to the bounding box. Such a tool would greatly reduce the effort required in the authoring of AR interfaces.

As a starting point for achieving this goal, in this work [198] we initially attempted to support designers through an automatic implementation process of AR interfaces in the industrial field, particularly for tasks related to component location and recognition. This study aimed to develop a strategy that enables designers to automatically generate diverse geometric shapes of CAD models quickly and easily. These shapes would be used as AMs in ARTD, specifically for the component location and recognition tasks.

Our previous work [117] confirmed that AMs have proven highly valuable in localizing components with different shapes and sizes. In this type of industrial activity, AMs are extensively used and are considered particularly suitable based on six identified parameters [63]. When compared to other models, there are several advantages to using AMs. For instance, they

[7] The results of the study described in this Section were published in the following paper: S. Romano, E. Laviola, and M. Gattullo, "ADAM: Automatic Development of Auxiliary Models. An Authoring Tool for Augmented Reality Technical Documentation," in Design Tools and Methods in Industrial Engineering, 2023.

effectively overcome the challenge of occlusion by employing simple geometrical shapes, which allows them to only overlap the relevant components without obstructing the operator's field of view. This significantly reduces risks for workers operating within a specific area. Additionally, AMs contribute to enhanced accuracy in component recognition and reduce communication time, i.e., the time refers to the duration between when the AM appears and when the operator correctly interprets the required action.

Expertise in CAD modeling is required to develop AMs to be imported into the AR application. Unlike other visual assets that need to be customized for each case study regarding shape or textual content, AMs can be reused multiple times, regardless of the context. This eliminates the need to repeatedly create them from scratch, wasting unnecessary time and effort. For these reasons, we have developed ADAM (Automatic Development of Auxiliary Models), a tool specifically designed to automatically create generic geometrical CAD models that can serve as AMs in AR interfaces. Its GUI allows designers to select a CAD model from a library. Then, they can specify the desired dimensions and download the model in a .obj file format, which is universally compatible with the software used for implementing AR interfaces.

7.1.1. Development of the authoring tool

The ADAM GUI (Fig. 40) has been developed using AutoIt and leverages Microsoft Excel and Autodesk Inventor Professional. When selecting the AMs to be included in the ADAM library, we considered the fundamental models offered by well-established modeling software. In our case, we opted for Paint3D, which is a native Windows software. We have chosen eight 3D AMs: the hollow cylinder, pentahedron, arrow, cylinder, cone, curved arrow, sphere, and ring. However, it is worth noting that designers can easily expand the library, allowing them to include additional 2D or 3D CAD models.

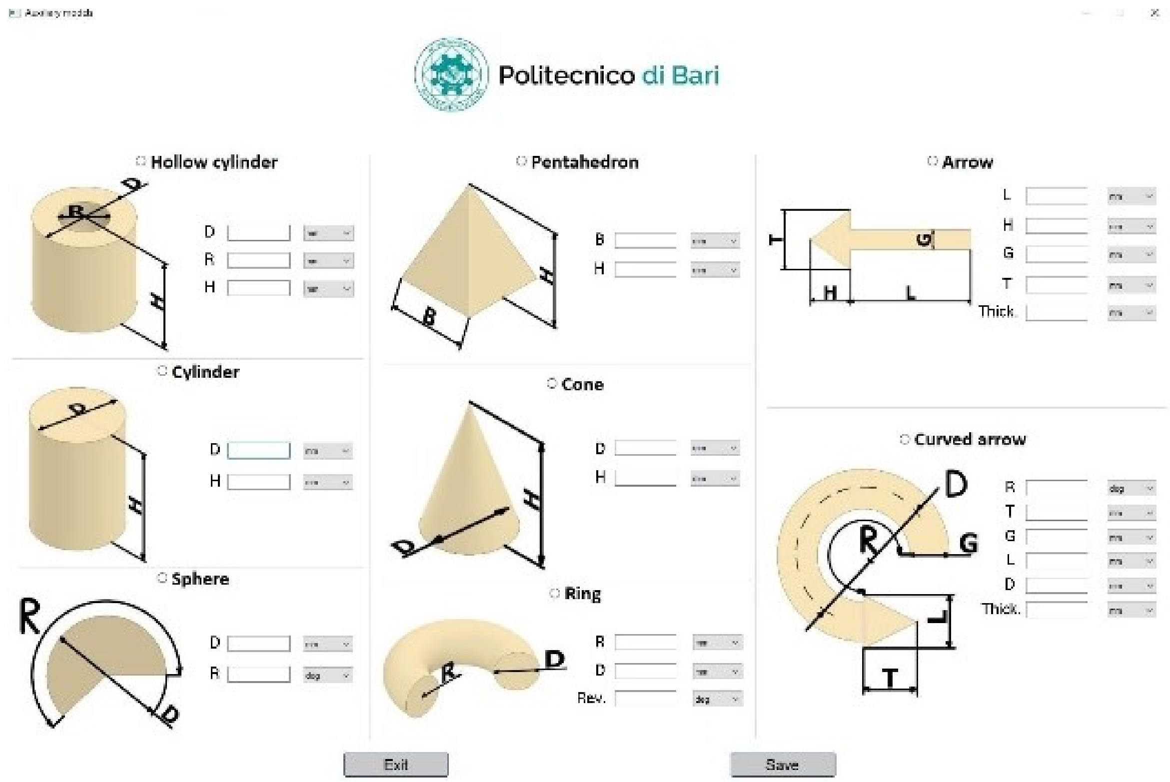


Fig. 40 ADAM GUI.

The input in the ADAM interface is processed and saved in an Excel spreadsheet connected to the Inventor parametric part files corresponding to each geometry. Within the ADAM tool, there is a primary directory named "Auxiliary Models," in which an executable file launches the application. Additionally, a subfolder called "Resources" contains all the necessary files for the application to work. These files include the CAD models in. ipt format, an Excel file, a script file, a folder containing images, and the project .ipj file. To ensure ADAM operates correctly, users need to have two specific software programs installed on their computer: Excel and Inventor. Even though users may not directly use these programs, they are required for the script to execute properly.

The GUI has been designed with Koda Form to be minimal and easy to use by any designer, even with limited experience in CAD modeling. Each CAD model is presented within its dedicated section, providing relevant information and functions. Specifically, a radio button allows users to select the desired model. A preview image of the model is displayed, accompanied by measurement indications to assist the designer. Input fields are provided for the designer to enter dimension values, with a dropdown menu for selecting the respective unit of measurement. Towards the bottom of the interface, there are two buttons: "Exit" to close the application and "Save" to download the CAD model in a .obj file format automatically.

To develop the tool, we created each CAD model in Inventor with parametric settings specifying dimensional parameters such as diameter, height, length, etc. To establish a connection between the parameters and an Excel spreadsheet, we used iLogic, an integrated module within the software. When users insert data in the ADAM GUI, the respective values automatically update the associated cells in Excel. The CAD model dynamically adjusts to reflect the updated dimensions through the established link between Excel and Inventor.

Using the iLogic module, we developed four functions for our tool. Firstly, an automated process updates the .ipt file based on the data linked to the Excel sheet. Secondly, we implemented a feature that allows users to save the desired model as a .obj file without overwriting the existing .ipt file. Furthermore, Microsoft Excel and Inventor operate in hidden mode, meaning they are not visible on the screen during the model creation process. Instead, users look at a progress bar that indicates the advancement of the script.

7.1.2. User evaluation

We conducted a case study on implementing an AR interface for Microsoft HoloLens 2 [105] to validate the ADAM tool. The study consisted of localizing components of various shapes and sizes on an engine using four AMs: hollow cylinder, cylinder, sphere, and ring. We set up a focus group of five expert designers (2 females, 28-40 years old, mean = 33.20, SD = 4.07) who were asked to implement the AR interface using the ADAM tool and the current authoring method. The first method involved using the ADAM tool to generate and download the AMs in the correct dimensions. These models were then placed on the respective components within the Unity Engine, the development platform for the AR interface. In contrast, the second method required the designers to manually model each AM from scratch using the CAD software Inventor. These models were subsequently imported into Unity, which had to be manually scaled to fit the corresponding components. To evaluate the effectiveness of both methods, we compared the time spent by the designers to complete the tasks and assessed the user experience. The latter was evaluated using the System Usability Scale (SUS) [199]. Additionally, subjective feedback was collected from the users to gain further insights into their experience.

Based on Fig. 41, the average time taken by the five designers using the current authoring method is around 12-13 minutes. However, when using the ADAM tool, the average time decreases significantly to approximately 2 minutes. This indicates a substantial time difference between the two methods, with our proposed tool allowing for an average time saving of 84%. These findings highlight the efficiency and time-saving benefits of using the ADAM tool compared to the traditional authoring method.

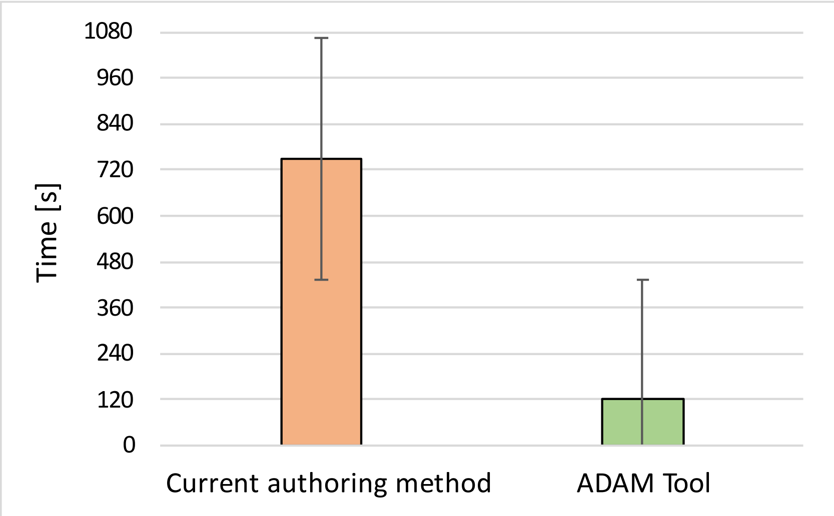


Fig. 41 Average task execution time of the two methods.

Table 17 displays the scores obtained from the SUS questionnaire. The mean SUS values for the five designers using the ADAM tool and the current authoring method are 86.50 and 55.50, respectively. The SUS score for the ADAM tool indicates excellent usability, as defined by [200]. On the other hand, the SUS score above 50 for the current authoring method suggests that it is considered "Ok" in terms of usability. These results imply that the expert designers expressed high satisfaction with the usability of the ADAM tool.

Table 17 SUS score for the compared methods.

Methods	N	Min	Max	Mean	S.D.
ADAM Tool	5	82.50	90.00	86.50	3.35
Current Authoring Method	5	50.00	60.00	55.50	3.71

For the evaluation, we deliberately selected experienced designers already familiar with CAD software as part of their regular work. The objective was to determine if these designers would be willing to replace the current method with our ADAM tool after testing it. The positive outcomes of the evaluation were further reinforced through personal interviews conducted after the experiment. During the interviews, all designers expressed their positive opinions about our tool. In summary, the designers' feedback indicated a strong appreciation for the speed and implemented functionalities of the ADAM tool. They found it a practical and effective support system for automatically generating AMs.

7.2. SMARTDOC: a guided authoring tool

Over the past two decades, several studies have confirmed that AR technology is highly effective in aiding industrial operators with tasks like assembly and maintenance [83]. Additionally, AR can enable less experienced workers to perform complex tasks. Given this context, many industrial companies seek to transition from their TTD to the ARTD, which involves creating virtual work instructions directly overlaid onto real products in 3D, along with customized visual assets to convey information effectively [3]. While this innovative form of TD has the potential to enhance operator performance, it also introduces new challenges during the content creation phase. These challenges mainly arise from the intricate requirements of industrial settings, such as integrating the real environment with a wider variety of visual assets. This complexity often hinders the adoption of AR in the industry [5]. Consequently, developing instructions for ARTD demands highly skilled developers. However, many companies still lack expertise in AR technology.

Several commercial platforms [201]–[203] have been developed to streamline the content creation process to provide digital work instructions to operators. However, these platforms are limited to presenting instructions in either 2D images or 3D models and do not leverage the capabilities of AR. Only a few commercial platforms, like Vuforia Work Instructions [201] and Microsoft Dynamics 365 Guides [202], can create AR instructions that are referenced to real components. In the literature, very few tools capable of automating the authoring process exist, as demonstrated in [203]. These solutions typically require the initial development of ARTD through a desktop interface before it can be used on devices like the Microsoft HoloLens 2. The absence of clear guidelines for assisting technical writers in creating ARTD highlights a gap in finding the best approach for generating user-friendly, efficient working instructions.

In this study, we introduce SMARTDOC (SiMplified authoring tool of Augmented Reality Technical DOCumentation) as our developed authoring tool. SMARTDOC serves as an innovative system to aid technical writers in creating ARTD and provides guidance to operators during their tasks. To the best of our knowledge, SMARTDOC represents the first standalone tool capable of creating ARTD directly from the user's device, in this case, the HoloLens 2, without relying on an external desktop authoring tool. The primary aim is to simplify the process of authoring ARTD for technical writers in the manufacturing industry.

7.2.1. Development of the authoring tool

The TD in SMARTDOC consists of multiple instructions, each containing specific information that needs to be conveyed. With SMARTDOC, users can insert one or more visual assets, complete with their properties (type, position, orientation, scale, color, and animation). This is accomplished through a guided step-by-step process, allowing users to create virtual elements that match the information type being conveyed and are properly aligned with the real components they refer to.

SMARTDOC was specifically designed for the Microsoft HoloLens 2 because it enables operators to have their hands free while carrying out tasks. The tool was developed using the Unity 3D Engine, incorporating the Mixed Reality Toolkit and Vuforia Engine for tracking. The process of registering virtual elements onto real components is facilitated by a customizable natural feature that can be placed in a specific area chosen by the user who is authoring the content.

Once the user launches the HoloLens application, he/she must select a save slot associated with a sequence of operations to be included in the ARTD using a hand menu (Fig. 42a). All saves are stored in a JSON file format, allowing the user to access previous step sequences for reference or to make modifications based on new requirements. Subsequently, the user can choose between two modes for using the tool: "authoring" or "viewer."

In "viewer" mode, users can only observe the step sequence that has been previously created, and they cannot make any changes to it. On the other hand, in "authoring" mode, users can edit one or more steps, depending on the complexity of the instruction they want to convey (Fig.42b). When a user clicks on the button corresponding to a particular step, a series of guided steps must be followed to insert all the virtual contents, along with their associated properties. During the authoring phase, each virtual element can be manipulated, and it is linked to the natural feature that must be positioned in a specific area beforehand.

Fig. 42 SMARTDOC hand menu with the save slots of each operations sequence (a), the step selection (b), and the textual instruction insertion (c).

The initial virtual element added to the scene is a textual instruction that should always be visible in an ARTD, as explained by Mourtzis et al. [204]. This instruction is presented on a colored label, creating a visual contrast with the real background, as depicted in Fig. 42c. The HoloLens device uses a virtual keyboard to achieve this. Subsequently, the user can choose a visual asset to include in the scene. More precisely, in line with the visual asset classification proposed in [8], the user can access a library of prefabricated AMs (e.g., cubes, cylinders, arrows) shown in Fig. 43a. Alternatively, they can upload their customized product models (Fig. 43b) or 2D images (Fig. 43c) in real-time to a designated folder within the application using the HoloLens Windows Device Portal website. To facilitate this, the HoloLens device must be connected to the same Wi-Fi network as the computer running the Windows Device Portal. The accepted file formats for these elements are obj for 3D models and png or jpeg for images. Initially, both AMs and product models are rendered using the Unity Default Material Shader.

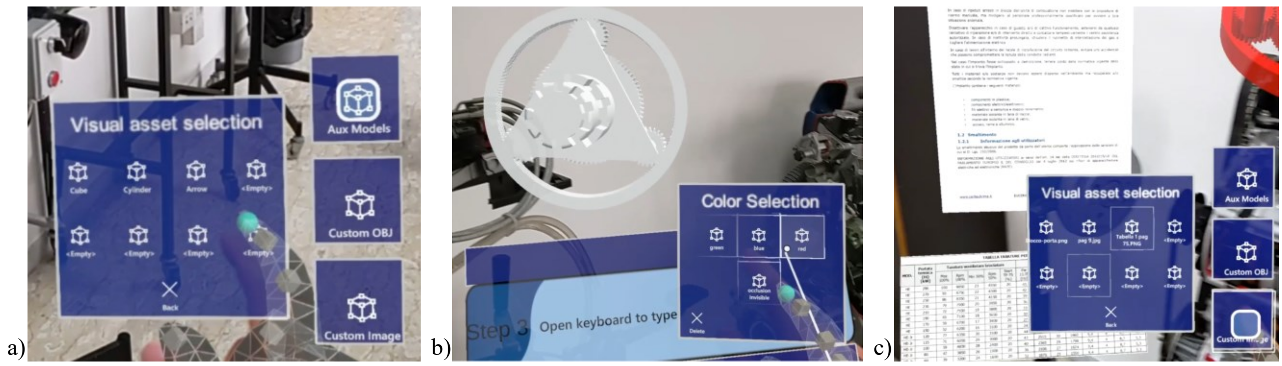


Fig. 43 The SMARTDOC visual asset selection: the AM menu (a), an example of uploaded product model (b), an example of uploaded 2D images (c).

Once the user has chosen a visual asset (excluding images), he/she has the option to select its color along with the opacity level, as illustrated in Fig. 44a. If there is a need to account for the obstruction caused by a physical component in relation to a virtual element, a Depth Mask shader can be applied to a 3D model representing the real component. Following this, the user can select an animation for the chosen visual asset, including options like static, translation, rotation, and blink, as shown in Fig. 44b. SMARTDOC offers a selection of commonly used animation types from existing applications. Furthermore, users can configure specific parameters for each animation, such as speed and direction of translation. These steps for inserting virtual content can be repeated for the same step or different ones within the application. To remove a virtual element, the user can simply pinch the corresponding visual asset for 7 seconds. Additionally, there is a dedicated button to delete all content associated with each step.

a) b)

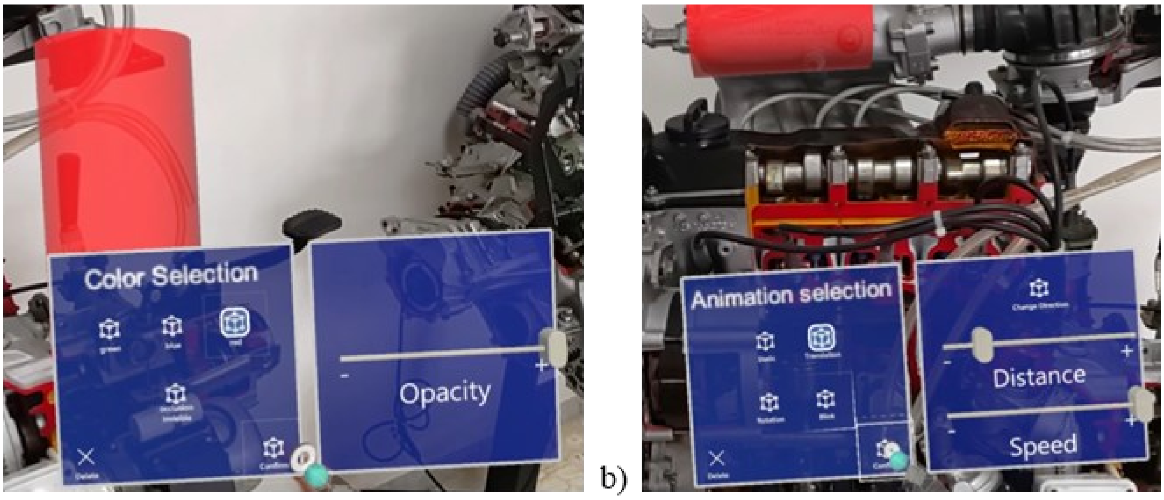


Fig. 44 Properties settings of the chosen visual asset: color selection (a), and animation selection (b).

7.2.2. Novelty compared to similar authoring tools

SMARTDOC stands out as a unique tool due to several distinctive features. Firstly, it is a standalone application, enabling users to create content for ARTD directly from their HoloLens device. Moreover, it offers the capability to upload 3D models and custom images directly through the Windows Device Portal website, eliminating the need for external software. This sets it apart from all other commercial platforms where such external software is typically required. Additionally, SMARTDOC simplifies the process of creating instructional steps with textual information directly on the HoloLens. This approach enhances usability by guiding users through the content creation process without necessitating the learning of new software tools. Another notable distinction is SMARTDOC capacity to handle occlusion effectively. It allows users to select a Depth Mask shader for any visual asset when necessary, ensuring that virtual objects interact realistically with real-world components.

Looking ahead to future developments, the focus will be on expanding the capabilities of this authoring tool. Plans include the integration of a machine learning system that can provide users with a list of suggested visual assets based on component properties and the specific task requirements outlined in a work instruction. This enhancement will further streamline the content creation process and enhance the user experience.

7.3. ARTD prototype in the design phase of the product lifecycle[8]

Numerous studies in the literature [83], [133], [205] have confirmed the effectiveness of AR in maintenance operations. The authoring of AR manuals involves determining how to provide instructions to operators using AR technology. To adopt a concurrent engineering approach, these manuals must be developed during the product design phase, parallelly to the product design. In fact, the ARTD project also has some implications on the product design, e.g., in finding solutions to enhance tracking capabilities. In certain situations, developers might not have physical access to the product, necessitating a reliable alternative to speed up the AR manual development process. In this work, we compared three solutions for displaying a demo version of an AR manual when the real product is not available, opting to replace it with its CAD model.

Traditionally, companies follow a sequential engineering process for maintenance, aiming to restore any functionality of a product during its lifecycle [206]. After the design phase, product manuals are created with textual instructions and drawings. However, the authoring of AR manuals is more complex due to various factors that must be considered simultaneously, such as the environment and lighting conditions. AR manuals offer different ways of displaying instructions, using visual assets such as AMs, product models, and drawings [8]. Compared to TTD, ARTD require a more complex design that is based on user needs analysis, leading to the specification of system requirements. A demo version is typically developed to test the effectiveness of the designed AR manual in an industrial field. This demo is presented to company managers for feedback and reviews. The AR demo manual can be directly displayed on the real product if the product already exists. However, for new products, the AR demo manual can only be shown on real prototypes after their production, which slows down the overall development process of the final application.

Most of the literature focuses on creating AR manuals for existing products, typically by transforming TTD into ARTD [207], [208]. Then, the research question of this work is: "*How to*

[8] The results of the study described in this Section were published in the following paper:
E. Laviola, M. Gattullo, and A. Evangelista, "Displaying Augmented Reality Manuals in the Design Phase of the Product Lifecycle," in Advances on Mechanics, Design Engineering and Manufacturing IV, Springer, Cham, 2022, pp. 1316–1326. doi: 10.1007/978-3-031-15928-2_115.

display a demo version of an AR manual in the design phase of a new product?” A comparative analysis of three distinct approaches has been designed to address this research question.

7.3.1. Prototypes

As a case study, we used an innovative semi-hermetic compressor of a local company that is still in development and has not been produced yet. The main purpose of the application is to assist workers in performing various tasks related to the installation, dismantling, and maintenance of the compressor. Additionally, the instructions provided by the application shall also be used to recognize and localize all the product components.

In a concurrent engineering approach, creating the AR manual prototype simultaneously with the product design is necessary. Consequently, a significant concern arises regarding how to display the compressor and integrate AR instructions effectively. As a result, three distinct demo versions of the manual were developed, all with the common objective of determining the most suitable technical solution. The aim is to allow technical writers to perform adequate analysis to select the appropriate visual assets for the final AR prototype once the real compressor becomes accessible.

We developed an application for a handheld device using Unity 3D Engine for each proposal, importing the 3D CAD model of the compressor provided by the company. It is essential to note that all three demo versions show the same content but offer different ways for users to interact with the 3D CAD model and access AR information referenced in the virtual compressor. Upon launching the application, an initial virtual menu appears with buttons that lead users to different sections following the structure of the digital manual provided by the company. For virtual scenes that do not require AR instructions, users can access information through primarily textual content, screen-fixed graphics, and illustrative drawings. However, in the AR interface, users can interact with various visual assets such as AMs, product models, drawings, and labels. These interactions are made possible through buttons strategically placed along the edges of the device screen. The placement ensures that these buttons do not obstruct the virtual elements displayed in AR while still allowing users to press them comfortably and maintain a stable grip on the device with both hands.

The three developed demo versions of the AR manual were named using the following terminology:

1) Augmented Reality (AR) prototype
2) Desktop Virtual Reality (DVR) prototype

3) Augmented Desktop Virtual Reality (ADVR) prototype

In the AR approach, Vuforia natural feature tracking was used to associate the CAD model of the compressor to a 2D image that represents a real drawing in scale 1:1 of the top-view of the CAD model (Fig. 45). Consequently, during the AR manual demonstration, the user's device displays the 3D model of the compressor in place of the real compressor. Furthermore, all the AR instructions are associated with this virtual 3D model.

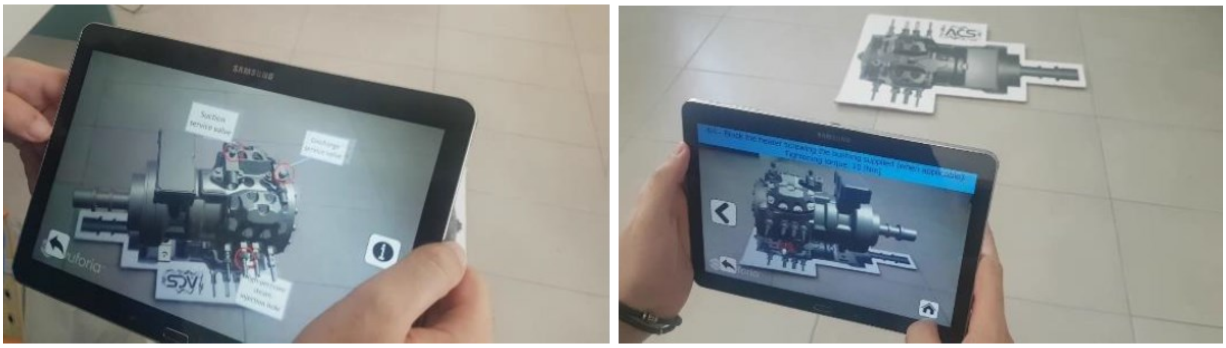


Fig. 45 User interaction example of the AR demo version: the handheld device frames the natural feature, and in this way, the AR content, consisting of the virtual compressor and the proposed visual assets, appears on the user's device.

The second approach involved developing a DVR prototype using a completely virtual environment with both the product and its instructions presented on the handheld device (Fig. 46). This approach was similar to the one suggested by [209]. Users could interact with the 3D model of the virtual compressor using the touch screen of the display. They could perform actions such as panning, rotating, and zooming the 3D model of the entire product, just like they would in any 3D CAD software. As a result, users could use GUI buttons to access and visualize instructions as virtual elements which replicate the information that would typically be seen through AR if the product were physically available.

Fig. 46 User interaction example of the DVR demo version: the handheld device shows the 3D model of the compressor and the AR content, consisting of the proposed visual assets, in an entirely virtual environment.

In the final alternative, the ADVR demo version was introduced. Initially, a support application was employed to display the CAD model of the compressor within a standard 3D CAD software environment on a computer desktop. As for the DVR solution, the user could pan, rotate, and zoom the 3D model of the entire product but use a keyboard and a mouse on the computer desktop instead of the touch screen on the handheld device. The Vuforia model target (360° option) was used to track the CAD model of the compressor (Fig. 47), as it was the real compressor. During the demonstration, the CAD model of the product is displayed on a large fixed screen, and the AR manual provides instructions attached to the 3D model.

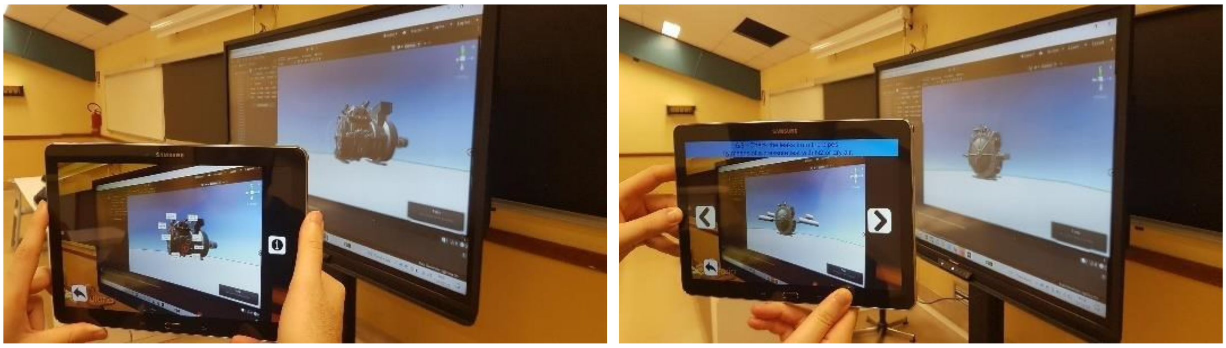


Fig. 47 User interaction example of the ADVR demo version: the handheld device shows the AR content, consisting of the proposed visual assets, referenced to the 3D CAD model of the compressor displayed on a secondary screen.

7.3.2. Material and Methods

The entire experiment carried out for each user was divided into two parts. The first one involved an initial session where the three prototypes of the AR manual were presented and explained to the users. The second part involved a subjective questionnaire.

In the first phase, experimenters provided participants with a detailed explanation of how each proposed approach works, with a particular emphasis on how users interact with the system. After the demonstration, users were requested to complete a subjective questionnaire that measures subjective satisfaction based on various factors related to user satisfaction, such as the content and how users interact with it. The proposed questionnaire is based on statements evaluating the user experience of mobile augmented reality services [210][16] and the Technology Acceptance Model (TAM) criteria [106]. Some items from the table “Examples of formative subjective statements with regard to the value and overall goodness of the service in terms of the UX category in question” and from the standard TAM questionnaire were picked and adapted to our scenario. The questionnaire comprises 10 items for each demonstration version, and participants rate their responses on a seven-point Likert scale, ranging from 1 (lowest) to 7 (highest). It encompasses eight evaluation categories: empowerment, efficiency, meaningfulness, intuitiveness, captivation, motivation, perceived ease of use, and intention to use (Table 18). At the end of the questionnaire, users were invited to share their comments regarding the strengths and weaknesses of each demo version.

Table 18 Questionnaire used for the user study.

Category	ID	Item
Empowerment	Q1	I think the application allows to pursue goals that are not supported by the other proposed technologies
Efficiency	Q2	I think the application can help me to choose visual assets for the final AR manual in an efficient way
Meaningfulness	Q3	I think with the application I can access information in the most appropriate place and time
	Q4	I think the content of the application makes sense in the context I use it
Intuitiveness	Q5	I think the application allows a natural way to interact with digital information
Captivation	Q6	I think I have a good conception of what is real and what is augmented in the AR application
Motivation	Q7	I think the application encourages me to produce and share information with other users in choosing visual assets
Perceived ease of use	Q8	I think learning how to use the application is easy for me
	Q9	I think the application simulates well the use of the final interface of the AR manual
Intention to use	Q10	Assuming I have access to the application, I intend to use it for choosing visual assets in an AR manual

15 individuals participated in the survey voluntarily and without compensation (2 females, 23 to 33 years old, mean = 25.2, SD = 2.71). All participants, except one Ph.D. student and a post-doc research fellow in Mechanical Engineering, were master's degree students in the same field. Before starting the experiment, participants were asked to rate their familiarity with AR and Virtual Reality (VR) on a scale of 1 to 7, where 1 represented "Not at all familiar" and 7 "Extremely familiar." On average, participants rated their familiarity with AR as 5.2 (SD = 1.33, Median = 5, Min = 2, Max = 7). Similarly, participants rated their familiarity with VR as 5.6 on average (SD = 0.88, Median = 6, Min = 4, Max = 7).

7.3.3. Results

A within-subject experiment was conducted where participants were exposed to three different experimental conditions: AR, AVR, and ADVR demo versions. The data obtained from all participants were analyzed to determine if they followed a normal distribution using the Shapiro-Wilk normality test (AS R94 algorithm). Since the original data did not meet the normal distribution criteria, non-parametric tests were chosen for further analysis. A statistical test called the Friedman 2-way ANOVA was employed to compare the three samples for each item. Additionally, a Wilcoxon ranks-sum test as a post-hoc test was used with the Bonferroni correction to examine the differences between the samples in more detail. The significance level was set at $p < 0.017$ to account for the adjusted threshold. All the results obtained from the statistical analysis were summarized in Fig. 48, showcasing the data collected from the experiment.

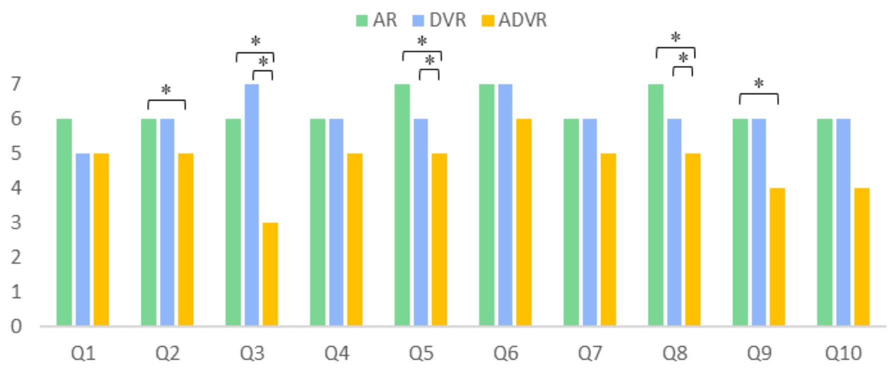


Fig. 48 Plot of the questionnaire median score for each item (: significant difference).*

In all three demo versions, no statistically significant differences were observed for the items Q1, Q4, Q6, Q7, and Q10.

However, when it came to perceived efficiency (item Q2), there was a significant difference found, $\chi2(2) = 18.250$, $p < 0.001$. Post hoc analysis with Wilcoxon signed-rank tests was carried out. The median perceived efficiency for each demo version was as follows: AR 6 (6 to 6), DVR 6 (5 to 6), ADVR 5 (3 to 6). There were no significant differences between the AR prototype and the DVR version ($Z = -1.265$, $p = 0.206$), nor between the DVR and ADVR versions ($Z = -2.254$, $p = 0.024$). However, there was a statistically significant reduction in perceived efficiency for the ADVR demo version compared to the AR one ($Z = -2.446$, $p = 0.014$).

A significant difference was observed in the perceived meaningfulness related to accessing information in the most appropriate place and time (item Q3), $\chi2(2) = 18.250$, $p < 0.001$. Post hoc analysis with Wilcoxon signed-rank tests was carried out. The median perceived meaningfulness for each demo version was as follows: AR 6 (3 to 6), DVR 7 (5 to 7), ADVR 3 (2 to 4). There was no significant difference between the AR prototype and the DVR version ($Z = -1.977$, $p = 0.048$). However, there were statistically significant reductions in perceived meaningfulness in the ADVR demo version compared to the AR version ($Z = -3.166$, $p = 0.002$) and the DVR one ($Z = -3.141$, $p = 0.002$).

A significant difference was observed in the perceived intuitiveness (item Q5) among the different demo versions, $\chi2(2) = 14.941$, $p < 0.001$. Post hoc analysis with Wilcoxon signed-rank tests was carried out. The median perceived intuitiveness for each demo version was as follows: AR 7 (6 to 7), DVR 6 (5 to 6), ADVR 5 (2 to 5). There was no significant difference between the AR prototype and the DVR version ($Z = -1.833$, $p = 0.067$). However, there were statistically significant reductions in perceived intuitiveness in the ADVR demo version compared to the AR version ($Z = -3.146$, $p = 0.002$) and the DVR one ($Z = -2.523$, $p = 0.012$).

A significant difference was found in the perceived ease of learning (item Q8), $\chi2(2) = 17.882$, $p < 0.001$. Post hoc analysis with Wilcoxon signed-rank tests was carried out. The median perceived ease of learning for each demo version was as follows: AR 7 (6 to 7), DVR 6 (6 to 7), ADVR 5 (4 to 6). There was no significant difference between the AR prototype and the DVR version ($Z = -1.000$, $p = 0.317$). However, there were statistically significant reductions in the perceived ease of learning in the ADVR demo version compared to the AR version ($Z = -2.820$, $p = 0.005$) and the DVR one ($Z = -2.831$, $p = 0.005$).

A significant difference was observed in the perceived ease of use regarding the similarity to the end-use with the real compressor (item Q9), $\chi2(2) = 8.851$, $p < 0.012$. Post hoc analysis with Wilcoxon signed-rank tests was carried out. The median perceived ease of use for each demo version was as follows: AR 6 (6 to 7), DVR 6 (5 to 7), ADVR 4 (4 to 6). There were no significant differences between the AR prototype and the DVR version ($Z = -1.026$, $p = 0.305$), nor between the DVR and ADVR versions ($Z = -1.588$, $p = 0.112$). However, there was a statistically significant reduction in the perceived ease of use in the ADVR demo version compared to the AR one ($Z = -2.914$, $p = 0.004$).

Regarding the feedback from participants on the AR demo version, 11 of them mentioned the naturalness and immediacy of interaction, and 6 the benefit of displaying virtual information aligned with a real scale CAD model. However, users also pointed out some drawbacks of using a natural feature, such as the tracking being significantly different from what the real compressor would look like when physically available. Specifically, 2 participants expressed difficulty distinguishing the visual assets from the virtual representation of the real object.

Regarding the feedback on the DVR prototype, 10 participants appreciated the ease of accessibility since not requiring any tracking. Additionally, 5 users found the interaction to be natural and intuitive. However, 8 participants expressed absolute detachment from the real context, making it challenging to distinguish between true AR and the virtual 3D model replacing the real product. Moreover, 2 participants pointed out that not using a 1:1 scale could be a disadvantage when evaluating the design quality of the visual assets.

Lastly, the ADVR solution presented some challenges, according to the feedback from participants. Specifically, 12 participants found it more demanding in terms of interaction because it required a secondary screen. Additionally, 2 users expressed concerns about the tracking system, as the model target is used on a 3D model displayed on a 2D screen, which could compromise the accuracy of the final result. On the positive side, 4 participants mentioned that they had no difficulty recognizing the visual assets from the virtual representation of the real object in the ADVR prototype. Furthermore, 5 users believed that this solution had the potential to facilitate information sharing among colleagues.

In the end, when asked to make their final choice (Fig. 49), most participants (67%) preferred the AR demo version over the DVR (20%) and ADVR (13%) prototypes.

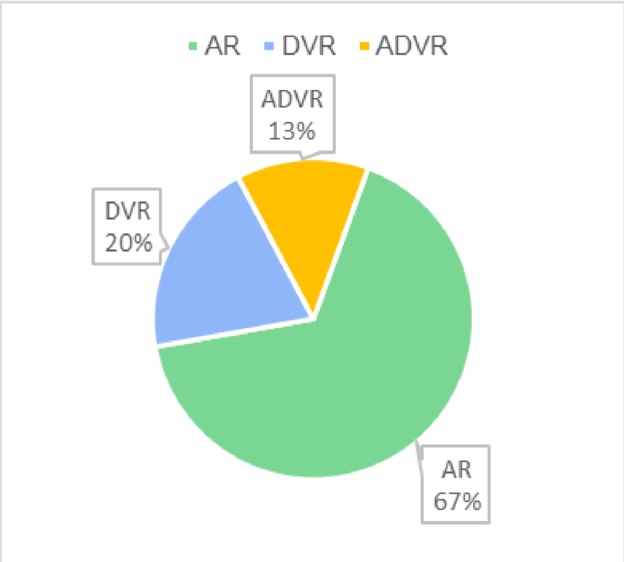


Fig. 49 Plot of the participants' final choices of the best demo version display in the design phase of a new product.

7.3.4. Lessons learned

The user evaluation of the AR, DVR, and ADVR demo versions provided valuable insights into addressing our research question: "*How to display a demo version of an AR manual in the design phase of a new product?*" The questionnaire results were further supported by the participants' technical comments, which shed light on the strengths and weaknesses of each prototype studied.

The primary drawback of the AR solution is its reliance on a physical printed image as a natural feature to attach AR instructions on the compressor. This requirement holds even if a CAD model is used instead of a real one. However, this limitation can be effectively addressed by implementing markerless tracking methods. Moreover, the AR approach offers the advantage of showcasing the product at a realistic scale, even if the real physical compressor is unavailable.

The DVR proposal uses a technique that is commonly called simulated AR [67], indirect AR [72], or immersive virtual AR [73]. Previous studies [72], [73] have shown that employing a simulated AR system yields results comparable to true AR in user evaluations but at lower costs and without the need to manage different devices and real-life locations, which is particularly advantageous for products that are not physically available. In this approach, a printed marker is not required during the demonstration; instead, only a handheld device is needed. The questionnaire responses regarding item Q3 also support the superiority of the DVR prototype over the other demo versions regarding perceived meaningfulness. However, the 3D model of the compressor is resized within a fully virtual environment, which results in less natural interaction. Users have to use the touch screen to change the view of the product, although item Q5 did not show a significant difference in the interaction between the DVR and AR demo versions. This outcome may be attributed to the comparison with the last ADVR solution, which found a significant reduction in both the naturalness of interaction (item Q5) and the perceived ease of learning (item Q8).

The ADVR prototype requires that the CAD model of the compressor is displayed on a separate device from the handheld one. Consequently, two users are involved in the interaction process: one runs the application with only AR instructions, while the other operates the 3D model in a virtual environment. Although this setup requires collaboration between the users during the demonstration and in the post-questionnaire feedback, it was found that this approach enables users to distinguish more easily between the virtual elements of true AR and simulated AR, which was not possible in the previous solutions. Additionally, an unexpected result emerged from the participants' comments, indicating that this ADVR solution stimulated user information sharing. The need to use a secondary large screen to interact with the 3D model of the compressor encourages users to use more handheld devices simultaneously, leading to shared discussions and opinions about the choice of visual assets.

After conducting the user study, the company evaluated all three solutions. Despite the user study participants showing a preference for the AR prototype, the company ultimately chose the ADVR solution. The main reason behind this decision could be the participants’ high familiarity with AR and VR technologies, resulting in no significant difference in their perception of what is real and what is augmented between the three prototypes (item Q6). However, in an industrial context, the company needs to also consider managers who may not be familiar with such innovative technologies. For these individuals, distinguishing between visual assets and the virtual representation of the real object might be more challenging. Moreover, they may be less capable of perceiving minor tracking inaccuracies that users identified in the ADVR solution. Considering these factors, the company found that the ADVR demo version provided a better representation of the final result that the true AR application would achieve. This is especially important when considering the type of tracking used in the physical product. Additionally, the ADVR solution encouraged sharing ideas among users for selecting the most appropriate visual assets for the AR manual, making it a preferred choice for the company.

Conclusion and future works

This dissertation aims to identify the most effective approach for conveying instructions in AR for manual documentation, addressing the existing gaps in the literature and the industrial domain. Based on the analysis of the current state of the art, the following observations can be made:

- Companies desire to use AR for their TD with a standardized framework that enables the development of an ARTD, even for technical writers who are not AR experts.
- Designing an ARTD requires decomposing work instructions into elemental information, for which appropriate visual assets must be suggested.
- Clear guidelines for decomposing work instructions into elemental information are absent.
- Clear guidelines on conveying information through visual assets in ARTD are lacking.
- Clear guidelines on selecting the most suitable visual assets based on the working context are missing.
- Product models are the most used visual assets for every information type; however, the selection process often lacks an objective criterion.

Therefore, a significant gap in the existing literature concerns the absence of clear guidelines on three key aspects: (i) the decomposition of work instructions into elemental information, (ii) the effective transmission of information through visual assets in ARTD, and (iii) the optimization of their properties according to the working context. This dissertation intended to fill these gaps. The findings provide guidelines to enhance operators' performance and simplify the ARTD authoring process, particularly for technical writers lacking expertise in AR.

The main outcomes can be summarized as follow:

- An information model based on the “Therbligs” method to decompose instructions into elemental information has been proposed. The six proposed information types (identity, location, order, way to, notification, and orientation) proved sufficient for analyzing all the instructions within the TTD that needed to be converted.
- The “minimal AR” authoring approach to choose visual assets based on the information to convey has been defined. There is no benefit in using AR signifiers that convey more information than the minimal signifier. This approach can streamline the authoring

process and prevent excessive efforts by selecting appropriate visual assets and avoiding information overload. In fact, in a real working context, workers positively accept the MRTD when applied to maintenance tasks using our “minimal AR” authoring approach.

- Guidelines for designing optimized AMs to locate in-view and not occluded components with regular shapes in complex machines have been provided:
  i. When dealing with a *regular* component, it is advised to use a 3D cylinder.
  ii. For *elongated* components, it is recommended to employ a 3D tube or cylinder.
  iii. In the case of a *plate* component, it is suggested to use a 2D AM that mimics the shape of the component boundary.
  iv. Regardless of the component shape, it is advised to employ yellow AMs with a blinking animation.
- "In-situ" and "3D side-by-side" presentation modes effectively displayed AR maintenance instructions in blind areas. Integrating in-situ instructions displayed using phantom rendering alongside a side-by-side 3D model proves beneficial for maintenance tasks.

In future works, we planned to deepen the "minimal AR" authoring approach by investigating the combined influence of object affordance and signifiers on user performance in true AR industrial scenarios. These studies will involve more complex tasks, such as assembling two parts or using specialized tools, heavily influenced by object affordance. Another aspect we will explore is the impact of operators' experience on their choice of visual assets. Depending on their skills, different combinations of visual assets may be necessary to provide the necessary information for effectively completing a task. As addressed for the LOCATION, examining the choice of visual assets when they convey the same amount of information for the other information types is interesting. Finally, by conducting these research projects in real industrial settings, we aim to develop guidelines for the ARTD to carry out concrete assessments of the economic impact that AR can offer for companies, compensating lacks present in the existing literature.

Acknowledgments

At the end of this journey, I would like to thank all the people who supported me:

- The Virginia Tech Polytechnic Institute, Blacksburg, Virginia, for hosting me during my PhD period abroad from 01/02/2023 to 02/06/2023. In particular, thanks to my supervisor Prof. Joseph L. Gabbard.
- The UPGRADING and CARLIEUKLIMA companies that provided us with the gas-fired radiant heating system for our user studies.
- All the VR3Lab people in these three years: Vito Modesto Manghisi, Antonio Boccaccio, Alessandro Evangelista, Lucilla Dammacco, Mine Dastan, Marina Ricci, Lorenzo Vaiani, Sara Romano, Vito De Giglio, Fabio Vangi, Francesco Laera, and Lorenzo Santoro. As friends, they supported me in facing the difficulties of my Ph.D. experience and for valuable suggestions. A special warm thanks to Michele Gattullo, a guide, colleague but, most of all, a friend who has been close to me in all the achievements during my Ph.D.
- All my coauthors.
- All the undergraduates supported in these years.

I thank all my family and friends for their moral support during this journey.

Finally, my supervisor Prof. Antonio E. Uva, and professors: Michele Fiorentino and Giuseppe Monno, for all the invaluable guidance given during these three years of my Ph.D. course.

Appendix 1: List of Figures

Appendix 2: List of Tables

Bibliography

[1] S. Henderson and S. Feiner, “Exploring the benefits of augmented reality documentation for maintenance and repair,” *IEEE Trans. Vis. Comput. Graph.*, vol. 17, no. 10, pp. 1355–1368, 2011, doi: 10.1109/TVCG.2010.245.

[2] A. E. Uva, M. Gattullo, V. M. Manghisi, D. Spagnulo, G. L. Cascella, and M. Fiorentino, “Evaluating the effectiveness of spatial augmented reality in smart manufacturing: a solution for manual working stations,” *Int. J. Adv. Manuf. Technol.*, vol. 94, no. 1–4, pp. 509–521, 2018, doi: 10.1007/s00170-017-0846-4.

[3] M. Lechner, “ARML 2.0 in the context of existing AR data formats,” *2013 6th Work. Softw. Eng. Archit. Realt. Interact. Syst. SEARIS 2013; Co-located with 2013 Virtual Real. Conf. - Proc.*, pp. 41–47, 2013, doi: 10.1109/SEARIS.2013.6798107.

[4] M. Lorenz, S. Knopp, and P. Klimant, “Industrial Augmented Reality: Requirements for an Augmented Reality Maintenance Worker Support System,” *Adjun. Proc. - 2018 IEEE Int. Symp. Mix. Augment. Reality, ISMAR-Adjunct 2018*, pp. 151–153, 2018, doi: 10.1109/ISMAR-Adjunct.2018.00055.

[5] J. M. Davila Delgado, L. Oyedele, T. Beach, and P. Demian, “Augmented and Virtual Reality in Construction: Drivers and Limitations for Industry Adoption,” *J. Constr. Eng. Manag.*, vol. 146, no. 7, p. 04020079, 2020, doi: 10.1061/(asce)co.1943-7862.0001844.

[6] M. Gattullo, G. W. Scurati, M. Fiorentino, A. E. Uva, F. Ferrise, and M. Bordegoni, “Towards augmented reality manuals for industry 4.0: A methodology,” *Robot. Comput. Integr. Manuf.*, vol. 56, 2019, doi: 10.1016/j.rcim.2018.10.001.

[7] K. Tainaka *et al.*, “Guideline and Tool for Designing an Assembly Task Support System Using Augmented Reality,” *Proc. - 2020 IEEE Int. Symp. Mix. Augment. Reality, ISMAR 2020*, pp. 486–497, 2020, doi: 10.1109/ISMAR50242.2020.00077.

[8] M. Gattullo, A. Evangelista, A. E. Uva, M. Fiorentino, and J. Gabbard, “What, How, and Why are Visual Assets used in Industrial Augmented Reality? A Systematic Review and Classification in Maintenance, Assembly, and Training (from 1997 to 2019),” *IEEE Trans. Vis. Comput. Graph.*, 2020, doi: 10.1109/TVCG.2020.3014614.

[9] M. Gattullo *et al.*, “Design preferences on Industrial Augmented Reality: A survey with

potential technical writers," *Adjun. Proc. 2020 IEEE Int. Symp. Mix. Augment. Reality, ISMAR-Adjunct 2020*, pp. 172–177, 2020, doi: 10.1109/ISMAR-Adjunct51615.2020.00054.

[10] S. Stork and A. Schubö, "Human cognition in manual assembly: Theories and applications," *Adv. Eng. Informatics*, vol. 24, no. 3, pp. 320–328, 2010, doi: 10.1016/j.aei.2010.05.010.

[11] R. Radkowski, J. Herrema, and J. Oliver, "Augmented Reality-Based Manual Assembly Support With Visual Features for Different Degrees of Difficulty," *Int. J. Hum. Comput. Interact.*, vol. 31, no. 5, pp. 337–349, 2015, doi: 10.1080/10447318.2014.994194.

[12] A. Syberfeldt, O. Danielsson, M. Holm, and L. Wang, "Dynamic Operator Instructions Based on Augmented Reality and Rule-based Expert Systems," *Procedia CIRP*, vol. 41, pp. 346–351, 2016, doi: 10.1016/j.procir.2015.12.113.

[13] C. Rolim, D. Schmalstieg, D. Kalkofen, and V. Teichrieb, "[POSTER] Design guidelines for generating augmented reality instructions," *Proc. 2015 IEEE Int. Symp. Mix. Augment. Reality, ISMAR 2015*, pp. 120–123, 2015, doi: 10.1109/ISMAR.2015.36.

[14] Anon, "No Title," *Market- sandmarkets, Market Leader -Augmented Reality and Virtual Reality Market*, 2022. https://www.marketsandmarkets.com/ResearchInsight/augmented-reality-virtual-reality-market.asp

[15] R. T. Azuma, "A Survey of Augmented Reality," *Presence Teleoperators Virtual Environ.*, vol. 6, no. 4, pp. 355–385, Aug. 1997, doi: 10.1162/pres.1997.6.4.355.

[16] R. W. Lindeman, H. Noma, and P. G. De Barros, "Hear-through and mic-through augmented reality: Using bone conduction to display spatialized audio," *2007 6th IEEE ACM Int. Symp. Mix. Augment. Reality, ISMAR*, pp. 173–176, 2007, doi: 10.1109/ISMAR.2007.4538843.

[17] O. Bau and I. Poupyrev, "REVEL: Tactile feedback technology for augmented reality," *ACM Trans. Graph.*, vol. 31, no. 4, pp. 1–11, 2012, doi: 10.1145/2185520.2185585.

[18] T. Narumi, S. Nishizaka, T. Kajinami, T. Tanikawa, and M. Hirose, "Augmented reality flavors: Gustatory display based on Edible Marker and cross-modal interaction," *Conf. Hum. Factors Comput. Syst. - Proc.*, pp. 93–102, 2011, doi: 10.1145/1978942.1978957.

[19] F. Zhou, H. B. L. Dun, and M. Billinghurst, "Trends in augmented reality tracking,

interaction and display: A review of ten years of ISMAR," *Proc. - 7th IEEE Int. Symp. Mix. Augment. Real. 2008, ISMAR 2008*, vol. 2, pp. 193–202, 2008, doi: 10.1109/ISMAR.2008.4637362.

[20] J. P. Rolland, Y. Baillot, and A. A. Goon, "A SURVEY OF TRACKING TECHNOLOGY FOR VIRTUAL ENVIRONMENTS Jannick P. Rolland, Yohan Baillot, and Alexei A. Goon Center for Research and Education in Optics and Lasers (CREOL) University of Central Florida, Orlando FL 32816," pp. 1–48, 2001.

[21] M. Bajura and U. Neumann, "Dynamic Registration Correction in Video-Based Augmented Reality Systems," *IEEE Comput. Graph. Appl.*, vol. 15, no. 5, pp. 52–60, 1995, doi: 10.1109/38.403828.

[22] M. Pressigout and É. Marchand, "Hybrid tracking algorithms for planar and non-planar structures subject to illumination changes," *Proc. - ISMAR 2006 Fifth IEEE ACM Int. Symp. Mix. Augment. Real.*, no. 1, pp. 52–55, 2006, doi: 10.1109/ISMAR.2006.297794.

[23] R. T. Azuma *et al.*, "Making Augmented Reality Work Outdoors Requires Hybrid Tracking," *Augment. Real.*, pp. 239–244, 1999, doi: 10.1201/9781439863992-28.

[24] H. Kato, M. Billinghurst, I. Poupyrev, K. Imamoto, and K. Tachibana, "Virtual object manipulation on a table-top AR environment," *Proc. - IEEE ACM Int. Symp. Augment. Reality, ISAR 2000*, pp. 111–119, 2000, doi: 10.1109/ISAR.2000.880934.

[25] J. Carmigniani and B. Furht, *Handbook of Augmented Reality*. 2011. doi: 10.1007/978-1-4614-0064-6.

[26] P. Fite-Georgel, "Is there a reality in Industrial Augmented Reality?," *2011 10th IEEE Int. Symp. Mix. Augment. Reality, ISMAR 2011*, pp. 201–210, 2011, doi: 10.1109/ISMAR.2011.6092387.

[27] T. Masood and J. Egger, "Augmented reality in support of Industry 4.0—Implementation challenges and success factors," *Robot. Comput. Integr. Manuf.*, vol. 58, no. August 2018, pp. 181–195, 2019, doi: 10.1016/j.rcim.2019.02.003.

[28] J. Zubizarreta, I. Aguinaga, and A. Amundarain, "A framework for augmented reality guidance in industry," *Int. J. Adv. Manuf. Technol.*, vol. 102, no. 9–12, pp. 4095–4108, 2019, doi: 10.1007/s00170-019-03527-2.

[29] C. S. Peirce, *Collected writings (8 vols.)*. Harvard Univ. Press. Cambridge, 1931.

[30] J. Wang, Y. Feng, C. Zeng, and S. Li, “An augmented reality based system for remote collaborative maintenance instruction of complex products,” *IEEE Int. Conf. Autom. Sci. Eng.*, vol. 2014-Janua, pp. 309–314, 2014, doi: 10.1109/CoASE.2014.6899343.

[31] ISO, *Technical drawings -- General principles of presentation -- Part 1: Introduction and index.* International Standards Organization Switzerland, 2003.

[32] *ASME Y14. 41-2003: Digital Product Definition Data Practices: An American National Standard*. 2003.

[33] J. L. Gabbard, G. M. Fitch, and H. Kim, “Behind the glass: Driver challenges and opportunities for AR automotive applications,” *Proc. IEEE*, vol. 102, no. 2, pp. 124–136, 2014, doi: 10.1109/JPROC.2013.2294642.

[34] A. Malta, M. Mendes, and T. Farinha, “Augmented reality maintenance assistant using yolov5,” *Appl. Sci.*, vol. 11, no. 11, pp. 1–14, 2021, doi: 10.3390/app11114758.

[35] J. Geng *et al.*, “A systematic design method of adaptive augmented reality work instruction for complex industrial operations,” *Comput. Ind.*, vol. 119, p. 103229, 2020, doi: 10.1016/j.compind.2020.103229.

[36] A. MacAllister, M. Hoover, S. Gilbert, J. Oliver, and R. Radkowski, “Comparing Visual Assembly Aids for Augmented Reality Work Instructions Recommended Citation CORE View metadata, citation and similar papers at core,” 2017, [Online]. Available: https://lib.dr.iastate.edu/me_conf/197

[37] C. Knöpfle, J. Weidenhausen, L. Chauvigné, and I. Stock, “Template based authoring for AR based service scenarios,” *Proc. - IEEE Virtual Real.*, pp. 237–240, 2005, doi: 10.1109/vr.2005.1492779.

[38] “Microsoft mixed reality guideline.” https://learn.microsoft.com/en-us/windows/mixed-reality/design/design

[39] “Google Augmented reality design guidelines.” https://developers.google.com/ar/design?hl=en

[40] “Apple Human interface guidelines—augmented reality.” https://developer.apple.com/design/human-interface-guidelines/augmented-reality

[41] C. H. Chu, C. J. Liao, and S. C. Lin, “Comparing augmented reality-assisted assembly functions-A case study on Dougong structure,” *Appl. Sci.*, vol. 10, no. 10, 2020, doi:

10.3390/APP10103383.

[42] S. Oh and W. Woo, "CAMAR: Context-aware Mobile Augmented Reality in Smart Space," *Proc. of IWUVR*, pp. 48–51, 2009, [Online]. Available: http://icserv.gist.ac.kr/mis/publications/data/2009/iwuvr-soh.pdf

[43] M. Gattullo *et al.*, "A context-aware technical information manager for presentation in augmented reality," *26th IEEE Conf. Virtual Real. 3D User Interfaces, VR 2019 - Proc.*, pp. 939–940, 2019, doi: 10.1109/VR.2019.8798262.

[44] M. Holm, O. Danielsson, A. Syberfeldt, P. Moore, and L. Wang, "Adaptive instructions to novice shop-floor operators using Augmented Reality," *J. Ind. Prod. Eng.*, vol. 34, no. 5, pp. 362–374, 2017, doi: 10.1080/21681015.2017.1320592.

[45] J. Wolfartsberger, M. Heiml, G. Schwarz, and S. Egger, "Multi-Modal Visualization of Working Instructions for Assembly Operations," *Int. J. Mech. Aerospace, Ind. Mechatron. Manuf. Eng.*, vol. 13, no. 2, pp. 107–112, 2019.

[46] S. Scheffer, A. Martinetti, R. Damgrave, S. Thiede, and L. van Dongen, "How to make augmented reality a tool for railway maintenance operations: Operator 4.0 perspective," *Appl. Sci.*, vol. 11, no. 6, 2021, doi: 10.3390/app11062656.

[47] Z. Yang *et al.*, "Influences of augmented reality assistance on performance and cognitive loads in different stages of assembly task," *Front. Psychol.*, vol. 10, no. JULY, pp. 1–17, 2019, doi: 10.3389/fpsyg.2019.01703.

[48] J. Gimeno, P. Morillo, J. M. Orduna, and M. Fernández, "An Advanced Authoring Tool for Augmented Reality Applications in Industry," *Actas las XXIII Jornadas Paralelismo (JP 2012). Elche Serv. Publicaciones la Univ. Miguel Hernández*, 2012.

[49] G. Reisinger, T. Komenda, P. Hold, and W. Sihn, "A Concept towards Automated Data-Driven Reconfiguration of Digital Assistance Systems," *Procedia Manuf.*, vol. 23, no. 2017, pp. 99–104, 2018, doi: 10.1016/j.promfg.2018.03.168.

[50] K. N. Kaipa, C. W. Morato, J. Liu, and S. K. Gupta, "Toward automated generation of multimodal assembly instructions for human operators," *Discip. Converg. Syst. Eng. Res.*, pp. 885–897, 2017, doi: 10.1007/978-3-319-62217-0_62.

[51] J. A. Erkoyuncu, I. F. del Amo, M. Dalle Mura, R. Roy, and G. Dini, "Improving efficiency of industrial maintenance with context aware adaptive authoring in augmented

reality," *CIRP Ann. - Manuf. Technol.*, vol. 66, no. 1, pp. 465–468, 2017, doi: 10.1016/j.cirp.2017.04.006.

[52] T. A. Pham and Y. Xiao, "Unsupervised workflow extraction from first-person video of mechanical assembly," *HotMobile 2018 - Proc. 19th Int. Work. Mob. Comput. Syst. Appl.*, vol. 2018-Febru, pp. 31–36, 2018, doi: 10.1145/3177102.3177112.

[53] A. Bocevska and Z. Kotevski, "Implementation of Interactive Augmented Reality in 3D Assembly Design Presentation," *Int. J. Comput. Sci. Inf. Technol.*, vol. 9, no. 2, pp. 141–149, 2017, doi: 10.5121/ijcsit.2017.9213.

[54] J. Blattgerste, P. Renner, and T. Pfeiffer, "Authorable augmented reality instructions for assistance and training in work environments," *ACM Int. Conf. Proceeding Ser.*, 2019, doi: 10.1145/3365610.3365646.

[55] D. Mourtzis, F. Xanthi, and V. Zogopoulos, "An adaptive framework for augmented reality instructions considering workforce skill," *Procedia CIRP*, vol. 81, no. March, pp. 363–368, 2019, doi: 10.1016/j.procir.2019.03.063.

[56] P. Renner, J. Blattgerste, and T. Pfeiffer, "A Path-Based Attention Guiding Technique for Assembly Environments with Target Occlusions," *25th IEEE Conf. Virtual Real. 3D User Interfaces, VR 2018 - Proc.*, pp. 671–672, 2018, doi: 10.1109/VR.2018.8446127.

[57] M. Chu, J. Matthews, and P. E. D. Love, "Integrating mobile Building Information Modelling and Augmented Reality systems: An experimental study," *Autom. Constr.*, vol. 85, no. October 2017, pp. 305–316, 2018, doi: 10.1016/j.autcon.2017.10.032.

[58] D. Norman, *The design of everyday things*. Psychol Everyday Things, 1988.

[59] D. Ferguson, "Therbligs: The Keys to Simplifying Work," *The Gilbreth Network (2000)*. https://gilbrethnetwork.tripod.com/therbligs.html

[60] R. Radkowski, "Investigation of visual features for augmented reality assembly assistance," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 9179, pp. 488–498, 2015, doi: 10.1007/978-3-319-21067-4_50.

[61] F. Jasche, S. Hofmann, and T. Ludwig, "Comparison of diferent types of augmented reality visualizations for instructions," *Conf. Hum. Factors Comput. Syst. - Proc.*, 2021, doi: 10.1145/3411764.3445724.

[62] J. Oyekan, W. Hutabarat, C. Turner, C. Arnoult, and A. Tiwari, “Using Therbligs to embed intelligence in workpieces for digital assistive assembly,” *J. Ambient Intell. Humaniz. Comput.*, vol. 11, no. 6, pp. 2489–2503, 2020, doi: 10.1007/s12652-019-01294-2.

[63] M. Gattullo, G. W. Scurati, A. Evangelista, F. Ferrise, M. Fiorentino, and A. E. Uva, “Informing the Use of Visual Assets in Industrial Augmented Reality,” *Lect. Notes Mech. Eng.*, pp. 106–117, 2020, doi: 10.1007/978-3-030-31154-4_10.

[64] E. Laviola, M. Gattullo, V. M. Manghisi, M. Fiorentino, and A. E. Uva, “Minimal AR: visual asset optimization for the authoring of augmented reality work instructions in manufacturing,” *Int. J. Adv. Manuf. Technol.*, 2021, doi: 10.1007/s00170-021-08449-6.

[65] A. Tang, C. Owen, F. Biocca, and W. Mou, “Comparative effectiveness of augmented reality in object assembly,” *Conf. Hum. Factors Comput. Syst. - Proc.*, no. 5, pp. 73–80, 2003, doi: 10.1145/642611.642626.

[66] J. Alves, B. Marques, M. Oliveira, T. Araújo, P. Dias, and B. S. Santos, “Comparing Spatial and Mobile Augmented Reality for Guiding Assembling Procedures with Task Validation,” in *2019 IEEE International Conference on Autonomous Robot Systems and Competitions (ICARSC)*, Apr. 2019, pp. 1–6. doi: 10.1109/ICARSC.2019.8733642.

[67] N. F. S. Jeffri and D. R. A. Rambli, “Problems with Physical Simulation in a Virtual Lego-based Assembly Task using Unity3D Engine,” *Proc. - 2020 IEEE Int. Conf. Artif. Intell. Virtual Reality, AIVR 2020*, pp. 298–300, 2020, doi: 10.1109/AIVR50618.2020.00060.

[68] L. Hou, X. Wang, L. Bernold, and P. E. D. Love, “Using Animated Augmented Reality to Cognitively Guide Assembly,” *J. Comput. Civ. Eng.*, vol. 27, no. 5, pp. 439–451, 2013, doi: 10.1061/(asce)cp.1943-5487.0000184.

[69] M. Richardson, T. E. Hunt, and C. Richardson, “Children’s construction task performance and spatial ability: Controlling task complexity and predicting mathematics performance,” *Percept. Mot. Skills*, vol. 119, no. 3, pp. 741–757, 2014, doi: 10.2466/22.24.PMS.119c28z8.

[70] R. B. Brünken, J. L. Plass, and D. L. Leutner, “Direct Measurement of Cognitive Load in Multimedia Leaming,” *Educ. Psychol. A Spec. Issue Educ. Psychol. Vol. 38*, vol. 38, pp. 53–61, 2003, doi: 10.4324/9780203764770-7.

[71] J. Cegarra and A. Chevalier Aline, “The use of Tholos software for combining measures

of mental workload: Toward theoretical and methodological improvements," *Behav. Res. Methods*, vol. 40, no. 4, pp. 988–1000, 2008, doi: 10.3758/BRM.40.4.988.

[72] J. Wither, Y. T. Tsai, and R. Azuma, "Indirect augmented reality," *Comput. Graph.*, vol. 35, no. 4, pp. 810–822, 2011, doi: 10.1016/j.cag.2011.04.010.

[73] G. Alce, K. Hermodsson, M. Wallergård, L. Thern, and T. Hadzovic, "A Prototyping Method to Simulate Wearable Augmented Reality Interaction in a Virtual Environment - A Pilot Study," *Int. J. Virtual World Hum. Comput. Interact.*, 2015, doi: 10.11159/vwhci.2015.003.

[74] C. Lee, G. A. Rincon, G. Meyer, T. Höllerer, and D. A. Bowman, "The effects of visual realism on search tasks in mixed reality simulation," *IEEE Trans. Vis. Comput. Graph.*, vol. 19, no. 4, pp. 547–556, 2013, doi: 10.1109/TVCG.2013.41.

[75] C. Lee, S. Bonebrake, T. Höllerer, and D. A. Bowman, "A replication study testing the validity of AR simulation in VR for controlled experiments," *Sci. Technol. Proc. - IEEE 2009 Int. Symp. Mix. Augment. Reality, ISMAR 2009*, pp. 203–204, 2009, doi: 10.1109/ISMAR.2009.5336464.

[76] E. Ragan, C. Wilkes, D. A. Bowman, and T. Höllerer, "Simulation of augmented reality systems in purely virtual environments," *Proc. - IEEE Virtual Real.*, pp. 287–288, 2009, doi: 10.1109/VR.2009.4811058.

[77] V. Tran and D. Kim, "An application of virtual reality in E-learning based LEGO-like brick assembling," 2016, pp. 783–786.

[78] C. M. Robertson, B. Maclntyre, and B. N. Walker, "An evaluation of graphical context when the graphics are outside of the task area," *Proc. - 7th IEEE Int. Symp. Mix. Augment. Real. 2008, ISMAR 2008*, pp. 73–76, 2008, doi: 10.1109/ISMAR.2008.4637328.

[79] A. Boccaccio *et al.*, "Exploiting augmented reality to display technical information on industry 4.0 P&ID," in *Lecture Notes in Mechanical Engineering*, Pleiades journals, 2019, pp. 282–291. doi: 10.1007/978-3-030-12346-8_28.

[80] R. Masoni *et al.*, "Supporting Remote Maintenance in Industry 4.0 through Augmented Reality," *Procedia Manuf.*, vol. 11, no. June, pp. 1296–1302, 2017, doi: 10.1016/j.promfg.2017.07.257.

[81] A. Doshi, R. T. Smith, B. H. Thomas, and C. Bouras, "Use of projector based augmented

reality to improve manual spot-welding precision and accuracy for automotive manufacturing," *Int. J. Adv. Manuf. Technol.*, vol. 89, no. 5–8, pp. 1279–1293, 2017, doi: 10.1007/s00170-016-9164-5.

[82] S. J. Henderson and S. Feiner, "Evaluating the benefits of augmented reality for task localization in maintenance of an armored personnel carrier turret," *Sci. Technol. Proc. - IEEE 2009 Int. Symp. Mix. Augment. Reality, ISMAR 2009*, pp. 135–144, 2009, doi: 10.1109/ISMAR.2009.5336486.

[83] R. Palmarini, J. A. Erkoyuncu, R. Roy, and H. Torabmostaedi, "A systematic review of augmented reality applications in maintenance," *Robot. Comput. Integr. Manuf.*, vol. 49, no. March 2017, pp. 215–228, 2018, doi: 10.1016/j.rcim.2017.06.002.

[84] L. Hou *et al.*, "Combining photogrammetry and augmented reality towards an integrated facility management system for the oil industry," *Proc. IEEE*, vol. 102, no. 2, pp. 204–220, 2014, doi: 10.1109/JPROC.2013.2295327.

[85] S. Werrlich, C. Lorber, P. A. Nguyen, C. E. F. Yanez, and G. Notni, "Assembly training: Comparing the effects of head-mounted displays and face-to-face training," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2018, vol. 10909 LNCS, pp. 462–476. doi: 10.1007/978-3-319-91581-4_35.

[86] J. Zauner, M. Haller, A. Brandl, and W. Hartmann, "Authoring of a mixed reality furniture assembly instructor," *ACM SIGGRAPH 2003 Sketches Appl. SIGGRAPH 2003*, no. UbiComp, p. 2002, 2003, doi: 10.1145/965400.965448.

[87] H. Eschen, T. Kötter, R. Rodeck, M. Harnisch, and T. Schüppstuhl, "Augmented and Virtual Reality for Inspection and Maintenance Processes in the Aviation Industry," *Procedia Manuf.*, vol. 19, no. 2017, pp. 156–163, 2018, doi: 10.1016/j.promfg.2018.01.022.

[88] D. B. Espíndola, C. E. Pereira, R. V. B. Henriques, and S. S. Botelho, "Using mixed reality in the visualization of maintenance processes," *IFAC Proc. Vol.*, vol. 1, no. PART 1, pp. 30–35, 2010, doi: 10.3182/20100701-2-pt-4012.00007.

[89] M. Speicher, B. D. Hall, and M. Nebeling, "What is mixed reality?," *Conf. Hum. Factors Comput. Syst. - Proc.*, pp. 1–15, 2019, doi: 10.1145/3290605.3300767.

[90] S. Lang, M. S. S. Dastagir Kota, D. Weigert, and F. Behrendt, "Mixed reality in production

and logistics: Discussing the application potentials of Microsoft HoloLensTM," *Procedia Comput. Sci.*, vol. 149, pp. 118–129, 2019, doi: 10.1016/j.procs.2019.01.115.

[91] M. Kerin and D. T. Pham, "A review of emerging industry 4.0 technologies in remanufacturing," *J. Clean. Prod.*, vol. 237, p. 117805, 2019, doi: 10.1016/j.jclepro.2019.117805.

[92] C. J. Dede, J. Jacobson, and J. Richards, *Virtual, Augmented, and Mixed Realities in Education*. 2017. doi: 10.1007/978-981-10-5490-7.

[93] Q. Guo, "Learning in a Mixed Reality System in the Context of ‚Industrie 4.0'," *J. Tech. Educ.*, vol. 3, no. 2, Dec. 2015, doi: 10.48513/JOTED.V3I2.60.

[94] G. Evans, J. Miller, M. Iglesias Pena, A. MacAllister, and E. Winer, "Evaluating the Microsoft HoloLens through an augmented reality assembly application," *Degrad. Environ. Sensing, Process. Disp. 2017*, vol. 10197, p. 101970V, 2017, doi: 10.1117/12.2262626.

[95] A. Siyaev and G. S. Jo, "Towards aircraft maintenance metaverse using speech interactions with virtual objects in mixed reality," *Sensors*, vol. 21, no. 6, pp. 1–21, 2021, doi: 10.3390/s21062066.

[96] S. Kim, M. A. Nussbaum, and J. L. Gabbard, "Influences of augmented reality head-worn display type and user interface design on performance and usability in simulated warehouse order picking," *Appl. Ergon.*, vol. 74, no. May 2018, pp. 186–193, 2019, doi: 10.1016/j.apergo.2018.08.026.

[97] D. Ariansyah, F. Rosa, and G. Colombo, "Smart maintenance: A wearable augmented reality application integrated with CMMS to minimize unscheduled downtime," *Comput. Aided. Des. Appl.*, vol. 17, no. 4, pp. 740–751, 2020, doi: 10.14733/cadaps.2020.740-751.

[98] R. Schlagowski, L. Merkel, and C. Meitinger, "Design of an assistant system for industrial maintenance tasks and implementation of a prototype using augmented reality," *IEEE Int. Conf. Ind. Eng. Eng. Manag.*, vol. 2017-Decem, pp. 294–298, 2018, doi: 10.1109/IEEM.2017.8289899.

[99] P. Piedimonte and S. L. Ullo, "Applicability of the mixed reality to maintenance and training processes of C4I systems in Italian Air Force," *5th IEEE Int. Work. Metrol. AeroSpace, Metroaerosp. 2018 - Proc.*, pp. 559–564, 2018, doi: 10.1109/MetroAeroSpace.2018.8453612.

[100] A. Pusch and F. Noël, “Augmented Reality for Operator Training on Industrial Workplaces – Comparing the Microsoft HoloLens vs. Small and Big Screen Tactile Devices,” in *IFIP Advances in Information and Communication Technology*, 2019, vol. 565 IFIP, pp. 3–13. doi: 10.1007/978-3-030-42250-9_1.

[101] F. Alves *et al.*, “Deployment of a Smart and Predictive Maintenance System in an Industrial Case Study,” *IEEE Int. Symp. Ind. Electron.*, vol. 2020-June, pp. 493–498, 2020, doi: 10.1109/ISIE45063.2020.9152441.

[102] R. C. M. Yusoff, R. Ibrahim, H. B. Zaman, and A. Ahmad, “Evaluation of user acceptance of mixed reality technology,” *Australas. J. Educ. Technol.*, vol. 27, no. 8, pp. 1369–1387, 2011, doi: 10.14742/ajet.899.

[103] B. C. Müller *et al.*, “Motion Tracking Applied in Assembly for Worker Training in different Locations,” *Procedia CIRP*, vol. 48, pp. 460–465, 2016, doi: 10.1016/j.procir.2016.04.117.

[104] E. Laviola, S. Romano, M. Gattullo, and A. E. Uva, “Evaluating the Worker Technology Acceptance of a Mixed Reality Technical Documentation,” in *Virtual Reality and Mixed Reality*, 2023.

[105] “Microsoft HoloLens, 2022. Mixed Reality Technology for Business.” https://www.microsoft.com/en-us/hololens

[106] F. D. Davis, R. P. Bagozzi, and P. R. Warshaw, “User Acceptance of Computer Technology: A Comparison of Two Theoretical Models,” *Manage. Sci.*, vol. 35, no. 8, pp. 982–1003, 1989, doi: 10.1287/mnsc.35.8.982.

[107] M. Gattullo, A. E. Uva, M. Fiorentino, and J. L. Gabbard, “Legibility in Industrial AR: Text Style, Color Coding, and Illuminance,” *IEEE Comput. Graph. Appl.*, vol. 35, no. 2, pp. 52–61, Mar. 2015, doi: 10.1109/MCG.2015.36.

[108] “Unity Real-Time Development Platform, 2022. | 3D, 2D VR & AR Engine.” https://unity.com/

[109] “Releases·microsoft/MixedRealityToolkit-Unity.” https://github.com/microsoft/MixedRealityToolkit-Unity/releases/

[110] “Vuforia engine developer portal.” https://developer.vuforia.com/

[111] C. Papakostas, C. Troussas, A. Krouska, and C. Sgouropoulou, “User acceptance of

augmented reality welding simulator in engineering training," *Educ. Inf. Technol.*, vol. 27, no. 1, pp. 791–817, 2022, doi: 10.1007/s10639-020-10418-7.

[112] J. Iqbal and M. S. Sidhu, "Acceptance of dance training system based on augmented reality and technology acceptance model (TAM)," *Virtual Real.*, vol. 26, no. 1, pp. 33–54, 2022, doi: 10.1007/s10055-021-00529-y.

[113] A. Sanna, F. Manuri, F. Lamberti, G. Paravati, and P. Pezzolla, "Using handheld devices to sup port augmented reality-based maintenance and assembly tasks," in *2015 IEEE International Conference on Consumer Electronics, ICCE 2015*, Mar. 2015, pp. 178–179. doi: 10.1109/ICCE.2015.7066370.

[114] L. M. Daling, A. Abdelrazeq, and I. Isenhardt, "A Comparison of Augmented and Virtual Reality Features in Industrial Trainings," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2020, vol. 12191 LNCS, pp. 47–65. doi: 10.1007/978-3-030-49698-2_4.

[115] T. Chesney, "ACCEPTANCE MODEL FOR USEFUL AND FUN INFORMATION SYSTEMS," *Hum. Technol. An Interdiscip. J. Humans ICT Environ.*, 2006, Accessed: Aug. 04, 2023. [Online]. Available: www.humantechnology.jyu.fi

[116] J. R. A. Santos, "Cronbach's alpha: A tool for assessing the reliability of scales," *J. Ext.*, vol. 37, pp. 1–5, 1999.

[117] S. Romano, E. Laviola, M. Gattullo, M. Fiorentino, and A. E. Uva, "More Arrows in the Quiver : Investigating the Use of Auxiliary Models to Localize In-view Components with Augmented Reality," *2023 IEEE Int. Symp. Mix. Augment. Real.*, 2023.

[118] M. Gattullo, E. Laviola, and A. E. A. E. Uva, "From Therbligs to Visual Assets : a Technique to Convey Work Instructions in Augmented Reality Technical Documentation," in *Advances on Mechanics, Design Engineering and Manufacturing IV*, Springer, Cham, 2022, pp. 1327–1339. doi: 10.1007/978-3-031-15928-2_116.

[119] T. Lavric, E. Bricard, M. Preda, and T. Zaharia, "Exploring low-cost visual assets for conveying assembly instructions in AR," *2021 Int. Conf. Innov. Intell. Syst. Appl. INISTA 2021 - Proc.*, pp. 1–6, 2021, doi: 10.1109/INISTA52262.2021.9548570.

[120] D. Markov-Vetter, M. Luboschik, A. T. Islam, P. Gauger, and O. Staadt, "The Effect of Spatial Reference on Visual Attention and Workload during Viewpoint Guidance in Augmented Reality," *Proc. - SUI 2020 ACM Symp. Spat. User Interact.*, 2020, doi:

10.1145/3385959.3418449.

[121] W. Li, J. Wang, S. Jiao, M. Wang, and S. Li, “Research on the visual elements of augmented reality assembly processes,” *Virtual Real. Intell. Hardw.*, vol. 1, no. 6, pp. 622–634, 2019, doi: 10.1016/j.vrih.2019.09.006.

[122] F. Obermair *et al.*, “Maintenance with Augmented Reality Remote Support in Comparison to Paper-Based Instructions: Experiment and Analysis,” *2020 IEEE 7th Int. Conf. Ind. Eng. Appl. ICIEA 2020*, pp. 942–947, 2020, doi: 10.1109/ICIEA49774.2020.9102078.

[123] J. Blattgerste, P. Renner, B. Strenge, and T. Pfeiffer, “In-situ instructions exceed side-by-side instructions in augmented reality assisted assembly,” *ACM Int. Conf. Proceeding Ser.*, pp. 133–140, 2018, doi: 10.1145/3197768.3197778.

[124] J. Wieland, R. C. H. Garcia, H. Reiterer, and T. Feuchtner, “Arrow, Bézier Curve, or Halos?-Comparing 3D Out-of-View Object Visualization Techniques for Handheld Augmented Reality,” *Proc. - 2022 IEEE Int. Symp. Mix. Augment. Reality, ISMAR 2022*, pp. 797–806, 2022, doi: 10.1109/ISMAR55827.2022.00098.

[125] Curtis P. Arledge, “Filled-in vs. Outline Icons: The Impact of Icon Style on Usability,” *Arledge, Curtis*, p. 68, 2014, [Online]. Available: https://cdr.lib.unc.edu/concern/masters_papers/6w924g35w

[126] M. Funk, T. Kosch, and A. Schmidt, “Interactive worker assistance: Comparing the effects of in-situ projection, head-mounted displays, tablet, and paper instructions,” *UbiComp 2016 - Proc. 2016 ACM Int. Jt. Conf. Pervasive Ubiquitous Comput.*, pp. 934–939, 2016, doi: 10.1145/2971648.2971706.

[127] J. Hahn, B. Ludwig, and C. Wolff, “Augmented reality-based training of the PCB assembly process,” *ACM Int. Conf. Proceeding Ser.*, vol. 30-Novembe, no. Mum, pp. 395–399, 2015, doi: 10.1145/2836041.2841215.

[128] R. Radkowski, T. Garrett, J. Ingebrand, and D. Wehr, “Trackingexpert - A versatile tracking toolbox for augmented reality,” *Proc. ASME Des. Eng. Tech. Conf.*, vol. 1B-2016, pp. 1–10, 2016, doi: 10.1115/DETC2016-60401.

[129] P. Renner and T. Pfeiffer, “Attention Guiding Using Augmented Reality in Complex Environments,” *25th IEEE Conf. Virtual Real. 3D User Interfaces, VR 2018 - Proc.*, pp. 771–772, 2018, doi: 10.1109/VR.2018.8446396.

[130] V. Paelke, "Augmented reality in the smart factory: Supporting workers in an industry 4.0. environment," *19th IEEE Int. Conf. Emerg. Technol. Fact. Autom. ETFA 2014*, pp. 1–4, 2014, doi: 10.1109/ETFA.2014.7005252.

[131] J. Blattgerste, B. Strenge, P. Renner, T. Pfeiffer, and K. Essig, "Comparing conventional and augmented reality instructions for manual assembly tasks," *ACM Int. Conf. Proceeding Ser.*, vol. Part F1285, pp. 75–82, 2017, doi: 10.1145/3056540.3056547.

[132] B. H. Limbu, H. Jarodzka, R. Klemke, F. Wild, and M. Specht, "From AR to Expertise: A User Study of an Augmented Reality Training to Support Expertise Development," *J. Univers. Comput. Sci.*, vol. 24, no. 2, pp. 108–128, 2018.

[133] G. W. Scurati, M. Gattullo, M. Fiorentino, F. Ferrise, M. Bordegoni, and A. E. Uva, "Converting maintenance actions into standard symbols for Augmented Reality applications in Industry 4.0," *Comput. Ind.*, vol. 98, pp. 68–79, 2018, doi: 10.1016/j.compind.2018.02.001.

[134] S. Chen *et al.*, "SEMarbeta: Mobile sketch-gesture-video remote support for car drivers," *ACM Int. Conf. Proceeding Ser.*, pp. 69–76, 2013, doi: 10.1145/2459236.2459249.

[135] C. Koch, M. Neges, M. König, and M. Abramovici, "Natural markers for augmented reality-based indoor navigation and facility maintenance," *Autom. Constr.*, vol. 48, pp. 18–30, 2014, doi: 10.1016/j.autcon.2014.08.009.

[136] S. J. Henderson and S. K. Feiner, "Augmented reality in the psychomotor phase of a procedural task," *2011 10th IEEE Int. Symp. Mix. Augment. Reality, ISMAR 2011*, pp. 191–200, 2011, doi: 10.1109/ISMAR.2011.6092386.

[137] U. Gruenefeld, L. Prädel, and W. Heuten, "Locating nearby physical objects in augmented reality," *ACM Int. Conf. Proceeding Ser.*, 2019, doi: 10.1145/3365610.3365620.

[138] S. Singh, "Impact of color on marketing," *Manag. Decis.*, vol. 44, no. 6, pp. 783–789, 2006, doi: 10.1108/00251740610673332.

[139] B. Schwerdtfeger and G. Klinker, "Supporting order picking with augmented reality," *Proc. - 7th IEEE Int. Symp. Mix. Augment. Real. 2008, ISMAR 2008*, pp. 91–94, 2008, doi: 10.1109/ISMAR.2008.4637331.

[140] S. Webel, U. Bockholt, T. Engelke, N. Gavish, and F. Tecchia, "Design Recommendations for Augmented Reality based Training of Maintenance Skills," in

Recent Trends of Mobile Collaborative Augmented Reality Systems, Springer New York, 2011, pp. 69–82. doi: 10.1007/978-1-4419-9845-3_5.

[141] M. Funk, J. Heusler, E. Akcay, K. Weiland, and A. Schmidt, "Haptic, auditory, or visual? Towards optimal error feedback at manual assembly workplaces," *ACM Int. Conf. Proceeding Ser.*, vol. 29-June-20, pp. 1–6, 2016, doi: 10.1145/2910674.2910683.

[142] U. Gruenefeld, D. Lange, L. Hammer, S. Boll, and W. Heuten, "FlyingARrow: Pointing towards out-of-view objects on augmented reality devices," *PerDis 2018 - Proc. 7th ACM Int. Symp. Pervasive Displays*, 2018, doi: 10.1145/3205873.3205881.

[143] B. Thomas, B. Close, J. Donoghue, J. Squires, P. De Bondi, and W. Piekarski, "First person indoor/outdoor augmented reality application: ARQuake," *Pers. Ubiquitous Comput.*, vol. 6, no. 1, pp. 75–86, 2002, doi: 10.1007/s007790200007.

[144] M. Di Donato, M. Fiorentino, A. E. Uva, M. Gattullo, and G. Monno, "Text legibility for projected Augmented Reality on industrial workbenches," *Comput. Ind.*, vol. 70, pp. 70–78, 2015, doi: 10.1016/j.compind.2015.02.008.

[145] M. A. Livingston, "Quantification of visual capabilities using augmented reality displays," *Proc. - ISMAR 2006 Fifth IEEE ACM Int. Symp. Mix. Augment. Real.*, pp. 3–12, 2006, doi: 10.1109/ISMAR.2006.297788.

[146] E. Kruijff, J. E. Swan, and S. Feiner, "Perceptual issues in augmented reality revisited," *9th IEEE Int. Symp. Mix. Augment. Real. 2010 Sci. Technol. ISMAR 2010 - Proc.*, pp. 3–12, 2010, doi: 10.1109/ISMAR.2010.5643530.

[147] C. Merenda, M. Smith, J. Gabbard, G. Burnett, and D. Large, "Effects of real-world backgrounds on user interface color naming and matching in automotive AR HUDs," *2016 IEEE VR 2016 Work. Percept. Cogn. Issues AR, PERCAR 2016*, pp. 1–6, 2016, doi: 10.1109/PERCAR.2016.7562419.

[148] J. Ping, B. H. Thomas, J. Baumeister, J. Guo, D. Weng, and Y. Liu, "Effects of shading model and opacity on depth perception in optical see-through augmented reality," *J. Soc. Inf. Disp.*, vol. 28, no. 11, pp. 892–904, 2020, doi: 10.1002/jsid.947.

[149] S. Wiedenmaier, O. Oehme, L. Schmidt, and H. Luczak, "Augmented Reality (AR) for Assembly Processes Design and Experimental Evaluation," *Int. J. Hum. Comput. Interact.*, vol. 16, no. 3, pp. 497–514, 2003, doi: 10.1207/S15327590IJHC1603_7.

[150] B. Volmer *et al.*, “A comparison of predictive spatial augmented reality cues for procedural tasks,” *IEEE Trans. Vis. Comput. Graph.*, vol. 24, no. 11, pp. 2846–2856, 2018, doi: 10.1109/TVCG.2018.2868587.

[151] S. Büttner, O. Sand, and C. Röcker, “Extending the design space in industrial manufacturing through mobile projection,” *MobileHCI 2015 - Proc. 17th Int. Conf. Human-Computer Interact. with Mob. Devices Serv. Adjun.*, pp. 1130–1133, 2015, doi: 10.1145/2786567.2794342.

[152] M. Fiorentino, A. E. Uva, M. Gattullo, S. Debernardis, and G. Monno, “Augmented reality on large screen for interactive maintenance instructions,” *Comput. Ind.*, vol. 65, no. 2, pp. 270–278, 2014, doi: 10.1016/j.compind.2013.11.004.

[153] Y. Wang, Y. Gao, R. Huang, W. Cui, H. Zhang, and D. Zhang, “Animated Presentation of Static Infographics with InfoMotion,” *Comput. Graph. Forum*, vol. 40, no. 3, pp. 507–518, 2021, doi: 10.1111/cgf.14325.

[154] L. Bonanni, C. H. Lee, and T. Selker, “Attention-based design of augmented reality interfaces,” *Conf. Hum. Factors Comput. Syst. - Proc.*, pp. 1228–1231, 2005, doi: 10.1145/1056808.1056883.

[155] S. Faroult, “Getting the message across,” vol. 78, no. 1, 1978.

[156] Bauckhage, “BOUNDING BOX SPLITTING FOR ROBUST SHAPE CLASSIFICATION Christian Bauckhage and John K . Tsotsos Centre for Vision Research,” pp. 9–12, 2005.

[157] D. Dimitrov, M. Holst, C. Knauer, and K. Kriegel, “Experimental study of bounding box algorithms,” *GRAPP 2008 - Proc. 3rd Int. Conf. Comput. Graph. Theory Appl.*, pp. 15–22, 2008, doi: 10.5220/0001096600150022.

[158] P. J. SCHNEIDER and D. H. EBERLY, “INTERSECTION IN 3D,” in *Geom. Tools Comput. Graph.*, 2003, pp. 481–662. doi: 10.1016/B978-155860594-7/50014-X.

[159] B. D’Amico and F. Pomponi, “A compactness measure of sustainable building forms,” *R. Soc. Open Sci.*, vol. 6, no. 6, 2019, doi: 10.1098/rsos.181265.

[160] J. Sturges and T. W. A. Whitfield, “Locating basic colours in the munsell space,” *Color Res. Appl.*, vol. 20, no. 6, pp. 364–376, 1995, doi: 10.1002/col.5080200605.

[161] A. Mulloni, H. Seichter, and D. Schmalstieg, “Handheld augmented reality indoor

navigation with activity-based instructions," *Mob. HCI 2011 - 13th Int. Conf. Human-Computer Interact. with Mob. Devices Serv.*, pp. 211–220, 2011, doi: 10.1145/2037373.2037406.

[162] J. M. Wolfe and T. S. Horowitz, "Wolfe2004What," vol. 5, no. June, pp. 1–7, 2004.

[163] P. Fraga-Lamas, J. Varela-Barbeito, and T. M. Fernandez-Carames, "Next Generation Auto-Identification and Traceability Technologies for Industry 5.0: A Methodology and Practical Use Case for the Shipbuilding Industry," *IEEE Access*, vol. 9, pp. 140700–140730, 2021, doi: 10.1109/ACCESS.2021.3119775.

[164] A. Evangelista, V. M. Manghisi, F. Laera, M. Gattullo, A. E. Uva, and M. Fiorentino, "CompassbAR: A Technique for Visualizing Out-of-View Objects in a Mixed Reality Environment," in *Lecture Notes in Mechanical Engineering*, 2022, pp. 141–148. doi: 10.1007/978-3-030-91234-5_14.

[165] T. Fukiage, T. Oishi, and K. Ikeuchi, "Reduction of contradictory partial occlusion in mixed reality by using characteristics of transparency perception," *ISMAR 2012 - 11th IEEE Int. Symp. Mix. Augment. Real. 2012, Sci. Technol. Pap.*, pp. 129–139, 2012, doi: 10.1109/ISMAR.2012.6402549.

[166] M. A. Livingston *et al.*, "Resolving multiple occluded layers in augmented reality," *Proc. - 2nd IEEE ACM Int. Symp. Mix. Augment. Reality, ISMAR 2003*, pp. 56–65, 2003, doi: 10.1109/ISMAR.2003.1240688.

[167] E. Laviola, M. Gattullo, A. Evangelista, M. Fiorentino, and A. E. Uva, "In-situ or side-by-side? A user study on augmented reality maintenance instructions in blind areas," *Comput. Ind.*, vol. 144, no. June 2022, p. 103795, 2023, doi: 10.1016/j.compind.2022.103795.

[168] S. Feng, W. He, S. Zhang, and M. Billinghurst, "Seeing is believing: AR-assisted blind area assembly to support hand–eye coordination," *Int. J. Adv. Manuf. Technol.*, vol. 119, no. 11–12, pp. 8149–8158, 2022, doi: 10.1007/s00170-021-08546-6.

[169] S. Makris, P. Karagiannis, S. Koukas, and A. S. Matthaiakis, "Augmented reality system for operator support in human–robot collaborative assembly," *CIRP Ann. - Manuf. Technol.*, vol. 65, no. 1, pp. 61–64, 2016, doi: 10.1016/j.cirp.2016.04.038.

[170] B. M. Khuong, K. Kiyokawa, A. Miller, J. J. La Viola, T. Mashita, and H. Takemura, "The effectiveness of an AR-based context-aware assembly support system in object

assembly," *Proc. - IEEE Virtual Real.*, pp. 57–62, 2014, doi: 10.1109/VR.2014.6802051.

[171] J. R. Li, L. P. Khoo, and S. B. Tor, "An object-oriented intelligent disassembly sequence planner for maintenance," *Comput. Ind.*, vol. 56, no. 7, pp. 699–718, 2005, doi: 10.1016/j.compind.2005.03.005.

[172] J. G. Kang and P. Xirouchakis, "Disassembly sequencing for maintenance: A survey," *Proc. Inst. Mech. Eng. Part B J. Eng. Manuf.*, vol. 220, no. 10, pp. 1697–1716, 2006, doi: 10.1243/09544054JEM596.

[173] M. C. F. MacEdo and A. L. Apolinario, "Occlusion Handling in Augmented Reality: Past, Present and Future," *IEEE Trans. Vis. Comput. Graph.*, vol. 29, no. 2, pp. 1590–1609, 2023, doi: 10.1109/TVCG.2021.3117866.

[174] C. H. Chu and C. H. Ko, "An experimental study on augmented reality assisted manual assembly with occluded components," *J. Manuf. Syst.*, vol. 61, no. March, pp. 685–695, 2021, doi: 10.1016/j.jmsy.2021.04.003.

[175] D. E. Breen, R. T. Whitaker, E. Rose, and M. Tuceryan, "Interactive occlusion and automatic object placement for augmented reality," *Comput. Graph. Forum*, vol. 15, no. 3, pp. 11–22, 1996, doi: 10.1111/1467-8659.1530011.

[176] M. O. Berger, "Resolving occlusion in augmented reality: A contour based approach without 3D reconstruction," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, pp. 91–96, 1997, doi: 10.1109/cvpr.1997.609304.

[177] Y. Tian, Y. Long, D. Xia, H. Yao, and J. Zhang, "Handling occlusions in augmented reality based on 3D reconstruction method," *Neurocomputing*, vol. 156, pp. 96–104, 2015, doi: 10.1016/j.neucom.2014.12.081.

[178] A. Holynski and J. Kopf, "Fast depth densification for occlusion-aware augmented reality," *SIGGRAPH Asia 2018 Tech. Pap. SIGGRAPH Asia 2018*, vol. 37, no. 6, pp. 1–11, 2018, doi: 10.1145/3272127.3275083.

[179] G. Klein and T. Drummond, "Sensor fusion and occlusion refinement for tablet-based AR," *ISMAR 2004 Proc. Third IEEE ACM Int. Symp. Mix. Augment. Real.*, no. Ismar, pp. 38–47, 2004, doi: 10.1109/ISMAR.2004.54.

[180] D. Kalkofen, E. Mendez, and D. Schmalstieg, "Interactive focus and context visualization for augmented reality," *2007 6th IEEE ACM Int. Symp. Mix. Augment. Reality, ISMAR*,

pp. 191–201, 2007, doi: 10.1109/ISMAR.2007.4538846.

[181] M. Bajura, H. Fuchs, and R. Ohbuchi, "Merging virtual objects with the real world: seeing ultrasound imagery within the patient," *Comput. Graph.*, vol. 26, no. 2, pp. 203–210, 1992, doi: 10.1145/142920.134061.

[182] C. Sandor, A. Cunningham, A. Dey, and V. V. Mattila, "An augmented reality X-ray system based on visual saliency," *9th IEEE Int. Symp. Mix. Augment. Real. 2010 Sci. Technol. ISMAR 2010 - Proc.*, pp. 27–36, 2010, doi: 10.1109/ISMAR.2010.5643547.

[183] N. Navab, M. Feuerstein, and C. Bichlmeier, "Laparoscopic virtual mirror new interaction paradigm for monitor based augmented reality," *Proc. - IEEE Virtual Real.*, pp. 43–50, 2007, doi: 10.1109/VR.2007.352462.

[184] C. Bichlmeier, N. Navab, S. M. Heining, and M. Feuerstein, "The Virtual Mirror: A New Interaction Paradigm for Augmented Reality Environments," *IEEE Trans. Med. Imaging*, vol. 28, no. 9, pp. 1498–1510, 2009, doi: 10.1109/TMI.2009.2018622.

[185] A. Martin-Gomez, A. Winkler, K. Yu, D. Roth, U. Eck, and N. Navab, "Augmented Mirrors," *Proc. - 2020 IEEE Int. Symp. Mix. Augment. Reality, ISMAR 2020*, pp. 217–226, 2020, doi: 10.1109/ISMAR50242.2020.00045.

[186] G. Westerfield, A. Mitrovic, and M. Billinghurst, "Intelligent augmented reality training for motherboard assembly," *Int. J. Artif. Intell. Educ.*, vol. 25, no. 1, pp. 157–172, 2015, doi: 10.1007/s40593-014-0032-x.

[187] B. Besbes, S. N. Collette, M. Tamaazousti, S. Bourgeois, and V. Gay-Bellile, "An interactive Augmented Reality system: A prototype for industrial maintenance training applications," *ISMAR 2012 - 11th IEEE Int. Symp. Mix. Augment. Real. 2012, Sci. Technol. Pap.*, pp. 269–270, 2012, doi: 10.1109/ISMAR.2012.6402568.

[188] J. Servan, F. Mas, J. L. Menéndez, and J. Ríos, "Assembly work instruction deployment using augmented reality," *Key Eng. Mater.*, vol. 502, pp. 25–30, 2012, doi: 10.4028/www.scientific.net/KEM.502.25.

[189] E. Laviola, S. Romano, and M. Gattullo, "Virtual Mirror for Maintenance: an Augmented Reality Support Tool," in *Design Tools and Methods in Industrial Engineering*, 2023.

[190] S. Zhang, W. He, S. Wang, S. Feng, Z. Hou, and Y. Hu, "An AR-Enabled See-Through System for Vision Blind Areas," *Commun. Comput. Inf. Sci.*, vol. 1498 CCIS, pp. 206–

213, 2021, doi: 10.1007/978-3-030-90176-9_28.

[191] S. G. Hart, “NASA-task load index (NASA-TLX); 20 years later,” *Proc. Hum. Factors Ergon. Soc.*, pp. 904–908, 2006, doi: 10.1177/154193120605000909.

[192] J. B. Du Prel, B. Röhrig, G. Hommel, and M. Blettner, “Auswahl statistischer testverfahren - Teil 12 der serie zur bewertung wissenschaftlicher publikationen,” *Dtsch. Arztebl.*, vol. 107, no. 19, pp. 343–348, 2010, doi: 10.3238/arztebl.2010.0343.

[193] N. Meshkati and P. Hancock, *Human mental workload*. 2011. [Online]. Available: https://books.google.it/books/about/Human_Mental_Workload.html?id=ItG1YGvRJ9oC &redir_esc=y

[194] M. Roxas, T. Hori, T. Fukiage, Y. Okamoto, and T. Oishi, “Occlusion handling using semantic segmentation and visibility-based rendering for mixed reality,” *Proc. ACM Symp. Virtual Real. Softw. Technol. VRST*, 2018, doi: 10.1145/3281505.3281546.

[195] Y. Kameda, T. Takemasa, and Y. Ohta, “Outdoor see-through vision utilizing surveillance cameras,” *ISMAR 2004 Proc. Third IEEE ACM Int. Symp. Mix. Augment. Real.*, no. Ismar, pp. 151–160, 2004, doi: 10.1109/ISMAR.2004.45.

[196] D. Kalkofen, E. Veas, S. Zollmann, M. Steinberger, and D. Schmalstieg, “Adaptive ghosted views for Augmented Reality,” *2013 IEEE Int. Symp. Mix. Augment. Reality, ISMAR 2013*, vol. 1, no. c, pp. 1–9, 2013, doi: 10.1109/ISMAR.2013.6671758.

[197] G. Ballestin, F. Solari, and M. Chessa, “Perception and Action in Peripersonal Space: A Comparison between Video and Optical See-Through Augmented Reality Devices,” *Adjun. Proc. - 2018 IEEE Int. Symp. Mix. Augment. Reality, ISMAR-Adjunct 2018*, pp. 184–189, 2018, doi: 10.1109/ISMAR-Adjunct.2018.00063.

[198] S. Romano, E. Laviola, and M. Gattullo, “ADAM : Automatic Development of Auxiliary Models . An Authoring Tool for Augmented Reality Technical Documentation,” in *Design Tools and Methods in Industrial Engineering*, 2023.

[199] J. Brooke, “SUS: A ‘Quick and Dirty’ Usability Scale,” *Usability Eval. Ind.*, no. July, pp. 207–212, 1996, doi: 10.1201/9781498710411-35.

[200] A. Bangor, P. Kortum, and J. Miller, “Determining what individual SUS scores mean; adding an adjective rating,” *J. usability Stud.*, vol. 4, no. 3, pp. 114–23, 2009.

[201] “Vuforia Work Instructions.” https://www.ptc.com/en/products/vuforia/work-

instructions

[202] "Microsoft Dynamics 365 Guides." https://dynamics.microsoft.com/it-it/mixed-reality/guides/

[203] V. Zogopoulos, E. Geurts, D. Gors, and S. Kauffmann, "Authoring Tool for Automatic Generation of Augmented Reality Instruction Sequence for Manual Operations," *Procedia CIRP*, vol. 106, no. March, pp. 84–89, 2022, doi: 10.1016/j.procir.2022.02.159.

[204] D. Mourtzis, A. Vlachou, and V. Zogopoulos, "Cloud-based augmented reality remote maintenance through shop-floor monitoring: A product-service system approach," *J. Manuf. Sci. Eng. Trans. ASME*, vol. 139, no. 6, pp. 1–11, 2017, doi: 10.1115/1.4035721.

[205] A. Ceruti, P. Marzocca, A. Liverani, and C. Bil, "Maintenance in aeronautics in an Industry 4.0 context: The role of Augmented Reality and Additive Manufacturing," *J. Comput. Des. Eng.*, vol. 6, no. 4, pp. 516–526, 2019, doi: 10.1016/j.jcde.2019.02.001.

[206] K. Komonen, "A cost model of industrial maintenance for profitability analysis and benchmarking," *Int. J. Prod. Econ.*, vol. 79, no. 1, pp. 15–31, 2002, doi: 10.1016/S0925-5273(00)00187-0.

[207] R. De Amicis, A. Ceruti, D. Francia, L. Frizziero, and B. Simões, "Augmented Reality for virtual user manual," *Int. J. Interact. Des. Manuf.*, vol. 12, no. 2, pp. 689–697, 2018, doi: 10.1007/s12008-017-0451-7.

[208] E. Lampen, J. Teuber, F. Gaisbauer, T. Bär, T. Pfeiffer, and S. Wachsmuth, "Combining simulation and augmented reality methods for enhanced worker assistance in manual assembly," *Procedia CIRP*, vol. 81, no. March, pp. 588–659, 2019, doi: 10.1016/j.procir.2019.03.160.

[209] E. A. L. Lee and K. W. Wong, "Learning with desktop virtual reality: Low spatial ability learners are more positively affected," *Comput. Educ.*, vol. 79, pp. 49–58, Oct. 2014, doi: 10.1016/j.compedu.2014.07.010.

[210] T. Olsson, "Human Factors in Augmented Reality Environments," in *Hum. Factors Augment. Real. Environ.*, 2013, pp. 203–232. doi: https://doi.org/10.1007/978-1-4614-4205-9.