Computer Science and Information Systems 19(2):1047–1073 https://doi.org/10.2298/CSIS211123013L

A Low-Cost AR Training System for Manual Assembly Operations*

Traian Lavric[1,2], Emmanuel Bricard[1], Marius Preda[2], and Titus Zaharia[2]

[1] IP Paris - Telecom SudParis, 91011, Évry, France
{traian.lavric, marius.preda, titus.zaharia}@telecom-sudparis.eu
[2] ELM LEBLANC SAS, 937400, Drancy, France
{traian.lavric, emmanuel.bricard}@fr.bosch.com

Abstract. This research work proposes an AR training system adapted to industry, designed by considering key challenges identified during a long-term case study conducted in a boiler-manufacturing factory. The proposed system relies on low-cost visual assets (i.e., text, image, video, and predefined auxiliary content) and requires solely a head-mounted display (HMD) device (i.e., Hololens 2) for both authoring and training. We evaluate our proposal in a real-world use case by conducting a field study and two field experiments, involving 5 assembly workstations and 30 participants divided into 2 groups: (i) low-cost group (G-LA) and (ii) computer-aided design (CAD)-based group (G-CAD). The most significant findings are as follows. The error rate of 2.2% reported by G-LA during the first assembly cycle (WEC) suggests that low-cost visual assets are sufficient for effectively delivering manual assembly expertise via AR to novice workers. Our comparative evaluation shows that CAD-based AR instructions lead to faster assembly (-7%, -18% and -24% over 3 assembly cycles) but persuade lower user attentiveness, eventually leading to higher error rates (+38% during the WEC). The overall decrease of the instructions reading time by 47% and by 35% in the 2nd and 3rd assembly cycles, respectively, suggest that participants become less dependent on the AR work instructions rapidly. By considering these findings, we question the worthiness of authoring CAD-based AR work instructions in similar industrial use cases.

Keywords: augmented reality, training, content authoring, work instructions, assembly, user study, industry 4.0.

1. Introduction

The industrial revolution also known under the label of Industry 4.0 provides a set of enabling technologies that support the development of individualized products in a cost-effective manner [41]. Augmented Reality (AR) is one of the key technologies that have demonstrated its benefits as a knowledge-sharing tool, among other applications in a variety of domains including education, medicine, tourism and entertainment [1][5]. Studies show that AR training systems can be more efficient in terms of task completion time and error rates when compared to classical training procedures (i.e., paper instructions) [44][2][40][10][7]. Although AR has been investigated as a guidance tool for manufacturing process since more than two decades [4], only recently, technological advancements

* The present paper is an extended and revised version of our preliminary conference report that was presented in INISTA 2021 [16]. This paper significantly expands the evaluation of the proposed AR training method.

enabled a resurgence of the AR use. However, despite the exponential progress that AR has experienced in recent years, no significant breakthrough can be noted in the industrial environment, due to various challenges [25][26]. AR systems have been mostly designed and evaluated in controlled environments, under laboratory settings, as recent surveys show [29][6]. Palmarini *et al.* [37] claimed that AR technology is not sufficiently mature for complying with strong industrial requirements such as robustness and reliability. Another recent study, conducted by Masood and Egger [26], identified and classified AR challenges into three main categories: technology, organization and environment and uncovered a gap between academic and industrial challenges. The authors suggested that field studies must be conducted in order to ensure the successful implementation of AR systems in industrial sectors.

We aimed to address these recommendations by elaborating an AR training solution for a concrete use case, a boiler-manufacturing factory. To this purpose, we conducted a long-term case study for obtaining a comprehensive picture of the needs and requirements, from both technical and organizational perspectives, that an AR training system should address, to be adopted in such context. The key success factors identified during our case study are effectiveness and viability. A summary of the most significant challenges that an AR training system should address, to be considered for adoption in the considered use case are safety, user acceptance, viability, technical setup, existing digital resources, assembly environment and process. A detailed description of our case study and its findings are presented in [16].

This work has an industrial focus; however, it explores relevant AR-related research topics identified by Kim *et al.* [14], including interaction techniques, user interfaces (UI), AR applications, evaluation, AR authoring, visualization and multimodal AR. Additionally, it addresses AR assembly concerns identified by Wang *et al.* [47], including time-consuming authoring procedures and appropriate guidance for complex, multi-step assembly tasks. Finally, it tries to answer a research question inquiring optimal ways for conveying instructions in Industrial Augmented Reality (IAR) [9]. We adopted therefore a human-centered design (HCD) approach to provide an intuitive, hands-free AR training system adapted to the shop floor environment, by addressing some of the most relevant industrial concerns identified during our case study and in the literature as well [6][21][27][39][35]. We evaluated the proposed AR training method, particularly the conveyance of the AR step-by-step instructions by conducting two field experiments. The first [16], a preliminary one that involved 12 participants, aimed at assessing the effectiveness and usability of our proposed low-cost AR training method. The second [17], an extension of the first, involved 20 additional participants and aimed at assessing the worthiness of using CAD data for conveying manual assembly expertise via AR. This paper presents a comprehensive overview of the two field experiments and discusses unpublished preliminary data in respect to the proposed AR authoring method, collected during a field study. The overall reported evaluation results collected from the considered experiments suggest that capturing and conveying expert knowledge via AR by using uniquely low-cost spatially registered visual assets is potentially the most efficient and viable option until the creation, manipulation and storage of CAD data and animations become more convenient, especially in industrial sectors.

The rest of this paper is organized as follows. Section 2 describes our proposed AR training method. Section 3 presents the technical implementation of the system. Section 4 describes the field experiments. We discuss the most significant findings in Section 5. A summary of the conclusions is presented in Section 6. Finally, limitations and suggestions for future work are discussed in Section 7.

2. Proposed Method

In this section, we justify the most relevant choices on which our proposal relies (see Section 2.1), we discuss the main design principles of our approach (see Section 2.2) and finally, we elaborate the proposed methodology, for both training and authoring (see Section 2.3). A summary of the main concerns that our proposed methodology aims to address, are further listed:

- **Content**: the AR system should not rely on existing digital data.
- **User**: the AR system should be adapted to shop floor personnel, particularly to assembly line experts and novice workers.
- **Environment**: the AR system should be hands-free, usable and effective, independently on the assembly environment.

2.1. AR Device, Visual Assets and Spatial Registration

To address the aforementioned concerns from a hardware perspective, our findings indicate that the best compromise is using cable-less HMD AR devices. Handheld devices (i.e., smartphones) do not answer the hands-free requirement while spatial augmented reality (SAR) systems [28][46] are not considered viable for the considered manufacturing context, shows our study. The methodology and implementation of the proposed AR training system relies therefore on the state-of-the-art AR device, Microsoft® HoloLens 2 [30], further referred to as Hololens 2.

The second most important aspect is represented by the way the assembly information is conveyed via AR. Literature shows that digital assets used to convey information in AR include text, audio, static 2D/3D and dynamic 2D/3D [20]. The visual ones are classified as text, sign/symbol, image/picture, video, drawing, 3D model and animations [9][20]. However, as identified in a recent study [9], there is no agreement in the literature regarding optimal ways of conveying instructions via AR. Tainaka *et al.* [43] empirically observed however that low-cost visual assets provide satisfactory results in conveying most manual assembly operations. Lee *et al.* [19] demonstrated the potential of first-person view (FPV) videos for conveying task instructions. In addition, we remark a potentially significant advantage of low-cost assets: unlike CAD models, these can be captured by state-of-the-art AR devices (i.e., Hololens 2), in-situ, as part of the AR authoring procedure itself. The authoring of the AR instructions is therefore not limited by existing digital content, preparation or post-processing steps, as proposed by commercial AR tools like Vuforia Expert Capture [38] and Microsoft Dynamics 365 Guides [33], further referred to as Guides. A summary of the most relevant concerns related to the usage of CAD models in AR, identified during our informal experiments are availability [21] and preparation, positioning during the authoring, occlusion, and real time spatial registration particularly for objects in motion. We expect that, by not depending on spatially registered CAD data, we remove the risk of rendering poor AR training experiences and even potential safety issues, which might arise due to imprecise world registration. We rely therefore our AR training proposal on low-cost visual assets, including text, image, video, and predefined auxiliary content.

The last aspect that we considered was the content registration, a core function of most AR systems, still an open issue of research. We identified three main types of information registration methods for HMD-based AR: object, head and environment-based [43]. Marker-based represents the most utilized (57%) registration technique among industrial applications [42]. Other techniques - i.e., 2D/3D recognition, sensor-based, location-based and marker less - do not comply with industrial requirements and are generally limited to test environments [42]. To address robustness and precision requirements, our training proposal relies on head (head-gaze technique) and environment (marker-based technique) registration methods.

2.2. UX Design Principles

User acceptance is identified as one of the most important success factors in the literature [26][27] and during our case study as well. Our informal experiments performed with shop floor workers suggest that a simplistic user experience (UX) is likely the best, considering the profile of the end users and the organization of the manufacturing environment. To ensure the usability of the proposed training method, we adopted a HCD approach: from the usage perspective, the proposed authoring tool should allow shop floor experts easily capture their assembly expertise, independently on the assembly environment. A standalone application, which does not require additional steps (i.e., desktop preparation, fine-tuning or offline content capture) and can be operated in-situ, in a "What You See Is What You Get" (WYSIWYG) manner, is potentially the most adapted. Lee *et al.* [18] demonstrated the advantages of immersive AR authoring in one of the first AR studies of this kind. Recently, Lorenz *et al.* [22] suggested that workstation experts are the most suited to create the AR instructions while better visualization techniques are needed during the authoring process, claims supported by our informal experiments as well.

Further, we analyzed and adopted information-presentation methods (i.e. registration, media types, semi-transparent effect and rotation) proposed as guidelines for AR assembly task support [43] and explored information access and peripheral awareness methods discussed in a study related to information access methods for HMD AR [23]. We followed and adopted guidelines to ensure the usability and effectiveness of the proposed solution, on the shop floor, independently on the assembly environment. We finally designed a hybrid solution, by combining and adjusting these guidelines [43] and techniques [19], to provide a contextualized information conveyance method adapted to the considered manual assembly scenario. We used implicit interaction techniques, including eye tracking and head position, along with common interaction techniques [36] like speech and touch, information outlined in Table 1.

A summary of the main HCD principles around which our proposed AR training system was elaborated, is listed further:

- **Familiarity**: use familiar UI patterns (buttons, arrows) and assets (text, images, and video) to increase user confidence and trust during the usage of the application.
- **Guidance**: use visual cues and implicit interaction techniques to guide the user during the training procedure, in the least intrusive manner.

– **Simplicity**: use a standard information delivery method regardless the variety of the assembly operations. Require deliberate input from the user only when necessary.
– **Comfort and safety**: do not clutter the UI and render the AR content at key locations of the assembly environment, as indicated during the authoring.

2.3. Methodology

The literature already shows that conveying instructions via AR produces better results when compared to classical training procedures [44][7][2]. However, previous research work does not yet provide optimal, standardized ways of delivering step-by-step instructions via AR, even less regarding the authoring of these AR instructions. It is not clear thus which AR visual modalities are optimal for conveying manual assembly information, especially under industrial requirements and challenges. Further, we describe in detail our proposed methodology, which aims to address this research question for the considered boiler-manufacturing use case.

The 2W1H (What, Where and How) Principle In the absence of a standardized method for digitally capturing and conveying manual assembly instructions in AR, we propose a technique that aims to address this concern. We note that each assembly operation, independently of its type and complexity, can be described by three variables: *what*, *where* and *how*. By using this technique, we try to replicate the oral human-to-human explanation of manual operations, as noted during our assembly training experiment and observations. What briefly describes the assembly operation, where indicates the physical location of the assembly operation and finally, how describes how the assembly is performed. This approach is based on the principle proposed by the Greek philosopher Aristotle, known as the "Five Ws (*Who*, *What*, *When*, *Where* and *Why*) and *How*", which represent the six basic questions in problem solving. In the considered use case, *who* – the trainee, *when* – now and *why* – training/authoring procedure, are known, therefore not considered as variables. Our hypothesis is that by following the 2W1H principle, the authors of the AR instructions will be able to describe any manual operation effectively and in a formalized manner, independently on the assembly environment and process. We aim as well to ensure a simple and consistent assembly information conveyance via AR, potentially easy to follow by novice shop floor workers, generally people without technical or AR expertise.

Assembly Instructions Chunking For the 2W1H principle to be applicable, each AR instruction should describe a single assembly operation. As an example, the assembly instruction "*Grab an upright and place it on the structure*" as defined in one of the existing paper instructions analyzed during our case study, becomes two separate "2W1H-friendly" instructions: (1) "*Grab an upright*" and (2) "*Place the upright on the structure*". By using this technique, we expect multiple benefits, as follows. First, the authoring and the training procedures are formalized and consistent, independently on the assembly environment or process. Secondly, by asking the author (during the authoring of the AR instructions) and the trainee (during the training procedure) to perform a single task at a time potentially decreases the assembly complexity, the mental workload, and the error rate. Finally, by limiting the number of virtual elements we avoid the UI clutter. Benefits of a similar chunking technique were recently discussed by Tainaka *et al.* in [43].

Visual Representation of an Assembly Task Regarding the visual representation of assembly tasks, we apply the 2W1H principle for describing them by using the considered low-cost visual assets. Each assembly operation is therefore visually composed of three elements:

- A text instruction, briefly describing the assembly operation (**what**).
- An arrow pointing to the physical location of the assembly operation (**where**).
- A FPV image or video illustrating complex assembly operations (optional) (**how**).

3. System Implementation

To evaluate our proposal, we developed two applications: (i) one for capturing the expert knowledge in AR (see Section 3.1) and (ii) one for conveying the authored AR instructions for training purposes (see Section 3.2). Both applications were developed for Hololens 2 by using Unity 3D (v. 2019.4.10f) [45] and MRTK v. 2.4.0) [32].

3.1. AR Instructions Authoring (Authoring Tool, On-the-Fly, In-Situ - ATOFIS)

The AR device on which we rely the implementation of our proposed authoring method is Hololens 2, which supports text insertion, photo, and video capture, as well as spatial registration of the virtual elements. Such functionalities supported the development of a standalone AR authoring tool that allows shop floor experts to capture their expertise *in-situ*, directly and only inside the AR device. The authoring is a procedure that does not require any prerequisites except an AR device connected to the internet and a unique (per workstation) QR code. The authored AR instructions are ready to be used for training immediately, as soon as the authoring procedure is finalized. Further, let us describe how the content authoring is performed for a single instruction (see Fig. 1), by following the proposed 2W1H principle, as discussed in Section 2.3. The same process applies for creating any AR instruction.

Fig. 1. AR authoring example. a) Step 1. Insert text instruction by using the virtual keyboard or dictation; b) Step 1 validated, step 2 active; c) Step 2. Positioning of the location arrow by using far interaction technique; d) Step 2 validated, step 3 active; e) Step 3. Photo-capture view; f) Step 3. Photo taken, the author validates or removes it.

At any time during the authoring procedure, the author uses hand gestures and voice commands (see Table 1) in order to interact with a 2D panel (Fig. 1.a), b), d)), displayed in front of him by using head registration technique (see Section 2.1). The authoring panel has multiple functions, as further detailed. Firstly, it displays the current assembly instruction number and the authoring step within the current AR instruction. Secondly, it allows the author to access the AR functions for text insertion and FPV photo and video capture. Finally, it allows the author to validate the captured data, advance to the next authoring step and create a new AR instruction. The application implements functions like visualization, selection, and editing of existing AR instructions; however, these are not discussed in the present paper, as they are not essential for the authoring procedure itself. Further, let us describe how the author creates a step-by-step AR instruction by following the 2W1H principle.

What: At this stage (1/3), along with the 2D authoring panel, a virtual keyboard is displayed in front of the user (Fig. 1.a). The author uses the keyboard to insert a text for briefly describing the current assembly task, by using one of the two modalities: (i) natural hand gesture technique, which require touching the virtual keystrokes or (ii) dictation by using voice, a function that is activated by the user by clicking the microphone button, part of the virtual keyboard. The user goes to the next authoring step by clicking a validation button, part of the 2D panel.

Where: At this stage (2/3), the author is required to place a virtual arrow for indicating the physical location of the assembly task (Fig. 1.c)). The arrow is displayed in front of the user, as a static object. The author uses his hand [31] to grab, place, scale and rotate it. Finally, the author validates its and implicitly the authoring step by clicking a "validate" button displayed under the arrow.

How: At this stage (3/3), the author captures a FPV image (Fig. 1.d)) or demonstration video to describe the assembly operation. It is up to the author to decide whether an image, a video or none of the two is required along with the text description and the indication arrow to effectively describe the assembly operation. The author captures one of the two by using the corresponding buttons of the authoring panel or the voice commands: "*photo*", "*video*" and "*stop video*". The author spatially registers the captured media at a convenient location in the real world by using hand interaction techniques [31]. The author is supposed to position the media preferably in the same field of view with the assembly location, to limit the head movement during the training and potentially decrease the assembly time and the effort required by the trainee while following the AR instructions. We note that during training, the images and the videos automatically rotate towards the user; therefore, during authoring, the author should not spend time rotating these elements to face specific real-world locations.

The contextualization of the media elements is one of the main differences between our approach and state-of-the-art AR authoring tools like Guides. Our previous work [15] suggests that the author should decide where the augmentation media is presented during training, instead of teaching and allowing trainees to interact and change its position. Another significant difference compared to existing authoring tools is represented by the way that the AR instructions are created, in a *WYSIWYG* manner, allowing the user to author the AR instructions during the assembly process, to visualize and validate his creation right away, during the authoring itself.

3.2. AR Instructions Conveyance (Training)

Further, let us detail how the assembly information is conveyed and how the user interacts with the visual elements during training. An example of the proposed training interface is presented in Fig. 2. Note that Fig. 2.d illustrates the usage of a CAD model, which replaces a location arrow (see Section 4.2). We note that all instructions are conveyed in the same manner and that the same AR device used for authoring the AR instructions, Hololens 2, is used for conveying these instructions, as follows.

What: Each instruction starts by displaying a text panel (Fig. 2.a)) in front of the user, between 0.6 to 0.7 meters away. The text panel follows user's head for 1 second (head registration) then it stops (environment registration). We ensure that the text is not overlooked by the user and, at the same time, that the panel does not visually interfere for more than necessary. The "sticking time" of 1 second is adjusted for our use case, based on the required movements of the user during the assembly procedure. The user hides the panel by clicking a "hide" button or by using the voice command "hide". Complementing the "hide" button with a voice command was required for cases when the text panel is rendered behind the physical environment, unreachable to hand touch. Our use case validates thus the requirement of multimodal interfaces discussed in [13].

Where: The next step consists in identifying the assembly location, pointed at by a spatially registered arrow (Fig. 2.b)). If the location is not in the field of view (FOV) of the user, a fixed-screen registered arrow (Fig. 2.a) and c)) guides the user towards it. Other techniques for localizing out-of-view objects in AR, like EyeSee360 and audio-tactile stimuli [24] and the "virtual tunnel" [11] are proposed in the literature. However, we rely on the arrow guidance-based technique for several reasons: (i) arrows are familiar visual cues, potentially easy to follow in unfamiliar environments like AR; (ii) visually, arrows are less intrusive and easier to integrate with other AR graphical elements; (iii) the technical implementation of such technique does not represent a challenge. A similar spatial cue technique was recently proposed in [19].

How: Optionally, a FPV image or demonstration video (Fig. 2.d)) describing the assembly operation is displayed in the proximity of the assembly location. Its position is spatially registered by the author so that the visual element and the assembly location are in the FOV of the user, minimizing therefore user's head movement while switching the attention between the two. The eye gaze controls the video playback, meaning that the video plays as long as the user looks at it. The implicit video playback interaction technique allows the trainee to follow video instructions without requiring deliberate input. We address thus the hands-free requirement while avoiding the UI clutter with the classical visual playback controls. We respect the principles discussed in Section 2.2, by allowing the trainee to focus on the assembly process, not on the application usage.

The user visualizes the next/previous instruction by clicking the "next"/"previous" button or by using the corresponding voice command. The "help me" voice command brings the text panel in front of the user. We note that unlike [19], our proposal uses text and images, in addition to video and indication arrows. In our approach, the FPV video is presented during the training experience exactly as captured in the authoring procedure.

Fig. 2 presents an example of the training workflow for performing two assembly operations. The first assembly task requires the worker to grab two uprights from the storage area (Fig. 2.a), b)). The second one requires the worker to place one of the uprights on the mobile assembly structure of the workstation (Fig. 2.b) and c)). We note that the directional arrow is orange and horizontal while the location one is blue and vertical.

Fig. 2. AR training example: a) Instruction 1a: text description ("*grab 2 uprights*") & directional arrow; b) Instruction 1b: Location arrow & FPV image illustrating the operation; c) Instruction 2a: text description ("*place the 1st upright*") & directional arrow; d) Instruction 2b: Location arrow & FPV video demonstrating the assembly operation.

3.3. Visualization and interaction techniques

Further, let us present the set of visualization and interaction techniques that aims to make it easy for shop floor workers to understand the training interface and be able to follow the AR instructions in the least intrusive manner.

Speech and touch: the user hides the text panel by clicking a "hide" button or by saying the voice command "hide". Similarly, we use voice commands to complement the instruction navigation buttons "next" and "previous". Touch and voice are interaction modalities that complement each other. Our UI requires multimodal interactions for cases when the virtual elements are unreachable by hand. Finally, the "help me" voice command brings the text panel in front of the user if this was hidden or left out of sight.

Head gaze: an implicit interaction technique, used to place dynamically the virtual elements, based on user's position and orientation. We used this technique for placing the text panel in front of the user and for rotating the virtual elements to always face the user. This way, the user is not required to move to certain physical locations for reading the text instruction, watching the associated FPV video demonstration, or inspecting the image.

Eye gaze: another implicit interaction technique that we used to control the video playback involuntarily: the video plays as long as the user looks at it. This technique addresses the hands-free UI requirement identified during our case study. At the same time, this allowed us to remove the "classical" video playback elements, avoiding therefore cluttering the UI. By using implicit interaction techniques, we aim to minimize intentional input from the user and provide guidance in the least intrusive manner. Table 1 presents a summary of the interaction techniques used during both the authoring (A) and training (T) procedures. Table 2 summarizes the information conveyance workflow. We use the frame of reference (FoR) notation for referring to the registration methods: screen-fixed (SF) and world-fixed (WF) [8].

Table 1. Interaction techniques.

Interaction technique	User input	Ouput
Speech	(A) "*photo*", "*video*", "*stop video*"	FPV image and video
	(T) "*next*", "*previous*", "*hide*", "*help me*"	Instruction navigation, show/hide virtual elements
Touch	Virtual elements (e.g., buttons)	(A) Add/update text instruction, add/replace location arrow, take a photo, record a video (T) Hide visual elements, step navigation
Hands object manipulation	(A) Instinctual interaction [41]	Scale, rotate and move virtual elements
Head gaze	(A+T) Implicit interaction	Dynamic positioning of virtual elements
Eye gaze	(A+T) Implicit interaction	Video playback

Table 2. UI and information conveyance

2W1H	Media type	FoR	Information	User action
What	Text instruction	SF / WF	Briefly describes the assembly operation	Reads text, then hides or ignores the panel
Where	Indication arrow	SF	Guides the user toward the assembly location	Turns the head towards the indicated direction
	Location arrow	WF	Indicates the assembly location	Identifies the location
How	Image / video	WF	Illustrates the assembly	Performs the assembly

4. Field Experiments

We conducted two field experiments and one field study in the boiler-manufacturing factory where we conducted our long-term case study to (i) measure the effectiveness and usability of the proposed AR training method and to (ii) validate our hypothesis that low-cost visual assets are sufficient for describing and conveying complex assembly operations via AR. The first field experiment was a preliminary one, with the main objective of measuring the usability and the effectiveness of the low-cost visual assets for conveying manual assembly instructions. The second experiment, an extension of the first, had as main objectives to (i) assess the potential benefits of authoring CAD-based AR instructions and to (ii) validate the hypothesis that our low-cost-based approach is potentially the most adapted technique for conveying assembly information in similar industrial use cases, until significant AR technical concerns are addressed.

We note that the AR instructions used in all three experiments were created by using ATOFIS, an implementation of the proposed authoring method, discussed in Section 3.1.

However, we discuss authoring statistics exclusively in the study presented in Section 4.3, specifically conducted for this purpose. As the first two field experiments, FE1 (see Section 4.1) and FE2 (see Section 4.2), were conducted under the same assembly setup, we present relevant information regarding the two of them jointly, as follows. Both experiments concerned the assembly of a boiler frame. The procedure consisted of 38 assembly tasks performed on the mobile structure of the first workstation of a manual assembly line. We grouped the assembly tasks (ATx) into four types:

- 14 x AT1 – picking (assembly components and tools)
- 8 x AT2 – installing / placement (assembly components)
- 12 x AT3 – screwing & riveting (screws and rivets)
- 4 x AT4 – manipulating (assembly structure and tools)

We used ATOFIS to author two sets of AR instructions. The first set, evaluated in FE1, was based solely on low-cost visual assets. The second instruction set was identical with the first, except that CAD models replaced the location arrows in assembly instructions of type AT2. The field experiment FE2 evaluated and compared both instruction sets. We note that every AT2 instruction had a FPV demonstration video associated to it in the first instruction set, complemented with a CAD model in the second. There was no potential benefit in complementing with CAD models the other assembly types (AT1, AT3 and AT4), reason for which these instructions were the same between the two instruction sets.

4.1. Field Experiment 1 (FE1)

The main objective of the field experiment presented in this section was to evaluate the effectiveness of the proposed low-cost-based AR training approach. A comprehensive description of the experiment set-up, participants, evaluation procedure, results and conclusions is presented in our previous work [16]. In this section, we present a summary of the most relevant findings reported in field experiment.

In the absence of an agreed upon framework to assess AR training systems for manual assembly process, we adopted the two evaluation methods identified by Wang *et al.* [47] in their AR assembly research survey: effectiveness and usability. We assessed the effectiveness of the proposed training method by measuring the error rate, the assembly completion time (ACT) and the instruction reading time (IRT). ACT represents the time spent for completing an assembly task; IRT represents the time spent on reading the low-cost visual assets. We evaluated the usability of the proposed training method by using the System Usability Scale (SUS) questionnaire [3]. We note that in FE2 (see Section 4.2) the extension of FE1, we present the evaluation results that include those reported in FE1, reason for which this section does not present in detail the findings further discussed.

Our measurement reported that 75% (9 out of 12) of the participants committed one or two errors during the first assembly cycle; however, we observe that the average error rate of 2.63% is very low, a value which potentially suggests a lack of user attentiveness rather than an issue regarding the AR information conveyance method. The convergence of the error rate to zero and the progress of the ACT over the course of the three assembly cycles potentially demonstrate the usability and effectiveness of our proposed training method. We note that all the reported errors except one consisted in a wrong orientation of the assembly component, supporting thus the effectiveness of the proposed AR training method, particularly for tasks of type AT1, AT2 and AT4.

A relevant finding revealed during the experiment indicates that participants get familiarized and recall the assembly instructions at a very fast pace: the time spent on reading the AR instructions decreases by 60% in the third assembly cycle, indicating a rapid diminishing utility of the AR instructions. The worthiness of authoring AR instructions by using CAD data, animations and other "expensive" media that would make the authoring more laborious is therefore questioned, a claim that is partially demonstrated in the second field experiment, FE2, discussed in the next section.

Fig. 3 illustrates the time required by each participant for completing each assembly task over three assembly cycles. We note the average time spent by the participants to read the instructions, *AvgRead* (blue line), matching closely the video length, *Video* (black line), during the first assembly cycle. We observe as well that *AvgRead* is flattening over the course of the three cycles. The peaks of *Video* over *AvgRead* in Fig. 3.c) indicate that participants stop watching the video entirely at this point, potentially suggesting that after only two cycles, videos can be replaced by images, even removed. The flattening of *AvgTotal* (red line), indicates the learning progress and the familiarization with the assembly.

Finally, we used the SUS questionnaire (see Table 7) to evaluate the overall usability of the proposed training method, by using a five-item Likert scale ranging from "*strongly agree*" to "*strongly disagree*". The reported usability score was 88.33 (SD = 9.02). All the participants that ranked Q4 with a score <= 60, claimed that they might need human support during the first assembly cycle, further referred to as the workstation exploration cycle (WEC). This claim is supported by the reported error rate during the WEC, where 75% of the participants committed one or two errors

The findings of the field experiment presented in this section suggest that spatially registered 2D visual assets together with a specific, human-centered set of interaction and visualization techniques could provide an effective AR-based conveyance method for describing manual assembly operations in industrial context. We note however that the reported error rate during the WEC suggests that a better technique for describing error-prone assembly operations is required, particularly for novice operators and during WEC.

4.2. Field Experiment 2 (FE2)

The field experiment presented in this section had multiple objectives, further listed:

- (O1) Extend and demonstrate the findings of FE2.
- (O2) Identify potential benefits of using CAD models for manual conveying assembly information by comparing the usability and effectiveness of two sets of AR instructions: low-cost vs. CAD-complemented.
- (O3) Answer a research question [9] suggesting that studies are needed to identify optimal ways to convey instructions in industrial sectors via AR.
- (O4) Validate the HCD principles discussed in Section 2.2, by measuring the perceived usability of the system and the mental workload of the participants.

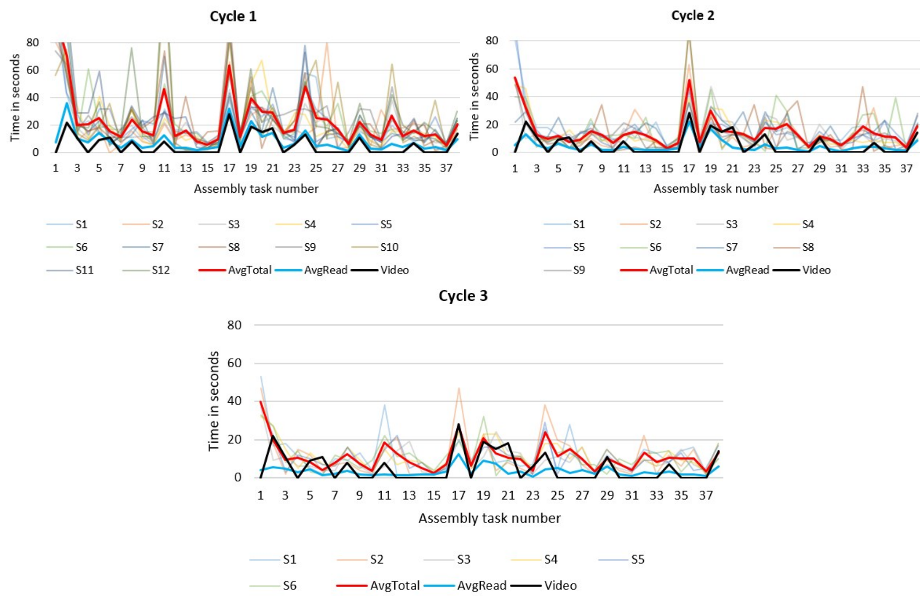

Fig. 3. Average assembly (*AvgTotal*) and reading (*AvgRead*) times per instruction, per a) cycle 1, b) cycle 2 and c) cycle 3

Similarly, as for the FE1, in this paper we present the most significant aspects of the field experiment FE2. A comprehensive description of this study is presented in our previous work [17]. We note that FE2 extends FE1 from 12 participants to 30 and from one instruction set (low-cost-based) to two instruction sets (low-cost and CAD-based). We created two groups, **G-LA** and **G-CAD**, each composed of 15 participants, for evaluating the two instruction sets: **LA** = Low-cost Assistive-based instruction set and **CAD** = CAD-based instruction set. Five participants have assembly experience in each group. We created two subgroups for each group: **G-LA-N** = novice participants from G-LA and **G-LA-E** = experienced participants from G-LA. Similarly, for G-CAD: **G-CAD-N** and **G-CAD-E**. We grouped the participants as such, to identify if assembly experience has a notable influence on the training performance.

Table 3 outlines this information.

Table 3. Evaluation groups

Group	**G-LA**		**G-CAD**	
Subgroup	G-LA-N	G-LA-E	G-CAD-N	G-CAD-E
Number of participants	10	5	10	5
Assembly experience	No	Yes	No	Yes
Instruction set number	1		2	

We assess the effectiveness and the usability of the proposed training system in the same manner as described in FE1, per instruction set and per subgroup, to address the main four objectives aforementioned. Further, we present a summary of the most relevant findings of this field experiment.

Table 4 presents the number of participants, per group, performing the nth assembly cycle. For each assembly cycle, we present the percentage of participants committing errors, the average error rate per instruction set, the total ACT and IRT (% of the ACT), and finally the average ACT of assembly operations of type AT2. We measure the IRT to identify differences and to estimate the utility of low-cost visual assets over multiple assembly cycles, within each instruction sets. A comparison between the two instruction subsets of type AT2 (CAD-complemented) is performed separately. Table 4 however presents the reported data collected on all instructions, to identify the impact of the CAD-based instructions over the whole instruction set, a practical evaluation approach for the considered use case.

The reported error rate and type shows that except one, all assembly errors were committed on operations of type AT2, during the first two cycles. We observe that subtle assembly details are prone to be overlooked, especially by participants without assembly experience, which commit more errors, as shown in Table 5, reason for which we believe that a better visual modality is needed for highlighting key assembly details, particularly for novice workers, during the WEC.

Table 4. Evaluation measurements

Group	**G-LA**			**G-CAD**		
Cycle no.	1	2	3	1	2	3
Participants no.	15	12	8	15	15	9
Participant error rate (%)	66%	25%	0%	66%	20%	0%
Total number of errors	13	3	0	18	3	0
Error rate per set (%)	2.2%	0.6%	0%	3.1%	0.5%	0%
Avg. ACT (s)	884s	538s	367s	838s	475s	336s
ACT progress (nth-1)		39%	31%		43%	29%
Avg. IRT (%)	37%	29%	25%	31%	27%	19%
Avg. ACT of AT2 (s)	290s	165s	98s	268s	130s	74s

Table 5 presents the average error rate committed per participant in each subgroup during the WEC. The error rates of the following cycles are not significant, therefore not discussed.

Table 5. Error rates per subgroup during the WEC

Group	**G-LA**		**G-CAD**	
Subgroup	G-LA-N	G-LA-E	G-CAD-N	G-CAD-E
Avg. errors per participant	1.1	0.4	1.4	0.8
Novice vs. experienced	-63%		-42%	
G-LA vs G-CAD	+38%			

We note that participants with assembly experience commit fewer errors in both groups (-63% and -42%) and surprisingly that G-CAD commits more errors than G-LA (38%).

The IRT measurement reveals that G-CAD participants watch the instructional videos less (during AT2 instructions) that G-LA, relying therefore on the CAD information more, potentially explaining their higher error rate during the WEC. The reported measurements indicate that G-CAD participants use 29% less time for watching videos, leading to a 7% decrease in the ACT, but to an increase in the error rate by 38% (see Table 5). The error rate convergence to zero after three assembly cycles support the hypothesis that both instruction sets are reliable for conveying manual assembly information in the considered use case. However, for assembly tasks of type AT2, the evaluation results suggest that human supervision might be necessary during the WEC.

The mean (M) and the standard deviation (SD) between the ACT over the three assembly cycles presented in Table 6, support the WEC paradigm and indicate that the participants start familiarizing with the assembly operations at a rapid pace. These findings underpin the conclusion of the FE1 and support the hypothesis that questions the worthiness of authoring of CAD-based AR instructions in similar industrial use cases.

Table 6. Mean and standard deviation of the ACT and IRT over three assembly cycles

Cycle number	**1**		**2**		**3**	
Global ACT/IRT	ACT	IRT	ACT	IRT	ACT	IRT
M	22.67	7.85	13.34	3.80	9.26	2.07
SD	13.04	6.89	6.73	3.07	4.04	1.32

Fig. 4 illustrates the mean ACT of all participants, per cycle. The ACT "flattening" over the three assembly cycles supports our claim, indicating the learning progress.

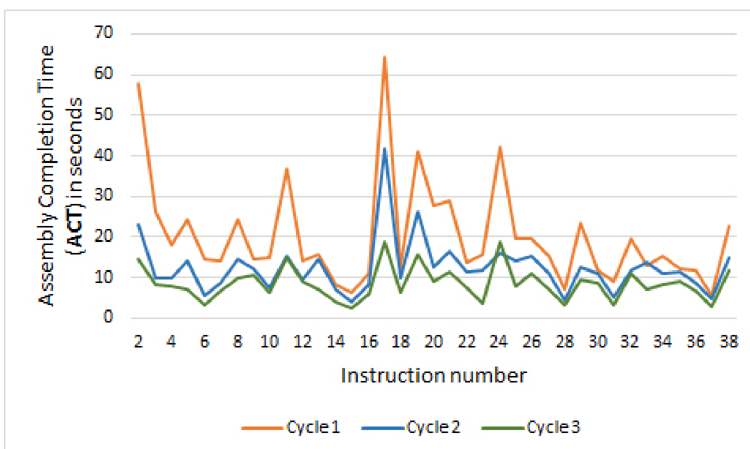

Fig. 4. ACT per instruction over 3 assembly cycles

The overall value of the IRT decrease by 47.5% and by 35.7% in the 2nd and 3rd cycles indicates that participants become less dependent on the AR instructions rapidly (see Table 6). By considering only the AT2 subset, we observe that CAD-based instructions lead to a faster assembly progress: G-CAD requires less time to perform the assembly operations of type AT2 compared to G-LA: -7%, -20% and -22% over the three assembly cycles (see Table 4).

In addition, we used the Pearson's correlation test to analyze the correlation between the ACT performances between the two groups, over the assembly operations of type AT2. The test shows a high correlation between the mean ACT of the two groups, during all assembly cycles: [r = 0.94, p = 0.142] for the first cycle, [r = 0.99, p = 0.003] for the second one and finally [r = 0.96, p = 0.007] for the third assembly cycle. We observe a high correlation (r = 0.94) without statistical significance (p = 0.142) during the first cycle; however, very strong correlations with statistical significance are reported for the second and third assembly cycles, respectively.

Finally, a subjective evaluation of the training method, including both instruction sets was performed. We used the SUS questionnaire (see Table 7) to evaluate the overall usability of the proposed training method (see Fig. 5).

Table 7. SUS questionnaire used to evaluate our proposed AR training system

No.	Question
1	I think that I would like to use this system frequently.
2	I found the system unnecessarily complex.
3	I thought the system was easy to use.
4	I think that I would need the support of a technical person to be able to use this system.
5	I found the various functions in this system were well integrated.
6	I thought there was too much inconsistency in this system.
7	I would imagine that most people would learn to use this system very quickly.
8	I found the system very cumbersome to use.
9	I felt very confident using the system.
10	I needed to learn a lot of things before I could get going with this system.

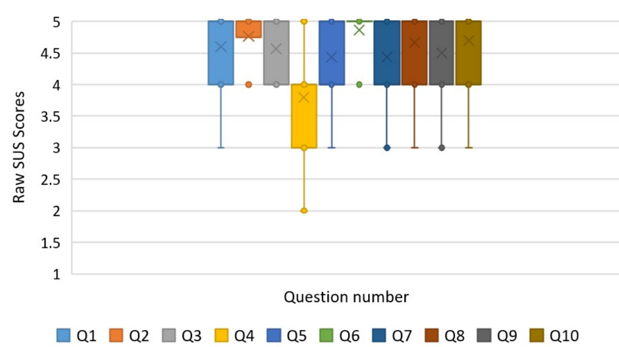

Fig. 5. Reported overall raw SUS scores

A one-way analysis of variance reveals no significant differences between G-LA and G-CAD [F(1,28)=0.01, p=0.89)] or between G-Experienced and G-Novice [F(1,28)=0.71, p=0.40)]. The overall reported perceived usability for all the participants is 4.53 (SD=0.25), indicating that the proposed method validates the HCD principles presented in Section 2.2. Similarly, as in FE1, Q4 reports the lowest rating: S=3.80, underlining the claim that human supervision is required during the WEC. However, Q4 reports a significant difference between G-LA vs. G-CAD [F(1,28)=5.34, p=0.02)] potentially indicating that CAD-based AR instructions lead to higher user confidence, evidence supported as well by the IRT difference of assembly operations of type AT2.

Further, we used the NASA-TLX questionnaire [12] to measure the mental workload of the participants. The raw NASA-TLX scores reported the following values: S=24.42, SD=4.75 for G-LA; S=24.22, SD=5.00 for G-CAD; S=25.25, SD=6.13 for G-Experienced and S=23.85, SD=5.49 for G-Novice (see Fig. 6). A one-way analysis of variance (ANOVA) finds no statistically significant differences ($p>.05$) between all groups and on all dimensions. However, our post-experiment evaluation reveals that participants with assembly experience have higher expectations from a temporal perspective, affecting their perceived performance level.

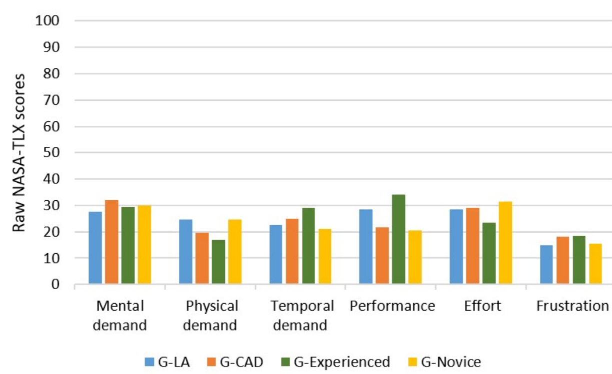

Fig. 6. Raw NASA-TLX scores per dimension and per group

4.3. Field Study (FS)

The study presented in this section was conducted by one of the authors and aimed at evaluating the proposed AR authoring system, ATOFIS (see Section 3.1), particularly from a time perspective. We conducted this work to provide an estimate of the time required for authoring the AR instructions of an assembly workstation and, concurrently, to compare our proposal with one of the most representative industrial AR authoring systems. We analyzed the creation of step-by-step low-cost-based AR instructions by measuring their authoring time and their media type composition. We note that the current study does not deal with CAD data, nor does it evaluate the authoring difficulty. Unlike FE1 and FE2, this section discusses unpublished data; therefore, a complete description of the considered study is presented, as further described.

Study set-up and evaluation procedure

The study was conducted in the same industrial environment as FE2. Five workstations have been uniformly chosen for authoring the corresponding step-by-step assembly instructions in AR. The numbers of assembly operations per each workstation are 33, 11, 33, 20 and 35. The author of the instructions is expert in performing the assembly tasks of the selected workstations and in manipulating the AR authoring system. We expect therefore the collected data of the field study to provide a reliable assumption regarding the authoring procedure when performed by experienced shop floor workers (e.g., line manager), at ease with the proposed system. In addition, this study aimed to validate authoring expectations from an industrial perspective, particularly regarding time constraints, adaptability, and robustness, before conducting a large-scale field experiment with shop floor experts, a costly and difficult study.

To evaluate objectively our proposed authoring method from a time perspective, we decided to compare it against Guides, the most popular state-of-the-art AR authoring tool. We do not detail the authoring workflow of Guides, available at [34]; however, we note the authoring workflow of Guides being comprised of three stages: media capture, PC authoring and HMD authoring. To avoid bias, for each workstation we wrote down the assembly instructions, together with the media type that shall be captured and the number of location arrows. This way we aimed to ensure the creation of the same AR instructions with both authoring systems: text description, number of location arrows and image or video (see the 2W1H principle, Section 2.3). Any assembly instruction was therefore written down by respecting the following template: "*DESCRIPTION, M, N*", where "M" can be either I (image) or V (video) and represents the captured media type while N is an integer number and represents the number of the location arrows of the assembly task in question. An example of such an assembly instruction is "*Screw the 4 screws at the indicated locations, I, 4*". To avoid bias, for every other workstation, the authoring started with the last system used for authoring the previous one.

Results and Interpretation

Table 8 presents the number of AR instructions and their corresponding media assets and characters, per workstation. Together with the information displayed in Table 9, we estimate what an AR assembly instruction is composed of, on average. Further, by considering the information presented in Table 11, we estimate the time required for authoring the AR instructions for a new workstation.

Table 8. Authoring data captured per workstation

Workstation	W1	W2	W3	W4	W5	Total
No. of instructions	33	11	33	20	35	132
No. of videos	12	6	13	9	9	49
No. of images	17	5	16	11	19	68
No. of arrows	43	22	41	24	45	175
No. of characters	1214	358	794	347	1106	3819

Table 9 presents the average number of characters, location arrows, images and videos used to author an AR instruction, per workstation.

Table 9. Average number of media used for authoring an AR instruction, per workstation

Workstation	W1	W2	W3	W4	W5	Total avg.
Avg. per instruction	33	11	33	20	35	132
No. of characters	37	33	24	18	32	28.8
No. of location arrows	1.3	2	1.2	1.3	1.3	1.4
No. of images	0.51	0.45	0.48	0.55	0.54	0.5
No. of videos	0.36	0.54	0.39	0.45	0.26	0.4

We used the reported preliminary data to estimate the composition, on average, of an AR instruction and the overall required authoring time of a new assembly workstation. The average number of videos per workstation indicates that the author of the AR instructions considers that 40% of the total assembly tasks are difficult or error-prone and require a video demonstration. The average number of images indicates that 50% of the total assembly tasks are relatively easy and that 10% of them are obvious and do not require a visual representation. Finally, the number of characters used to describe an assembly task is 28.8, on average, as indicated in Table 9.

Table 10 presents the authoring times in seconds, per workstation, for both AR authoring systems, Guides and ATOFIS. Additionally, the time differences, in percentage, in the favor of ATOFIS are presented.

Table 10. Authoring times comparison between ATOFIS and Guides

Workstation	W1	W2	W3	W4	W5	Total
Guides (seconds)	2822s	991s	2459s	1804s	2706s	10782s
ATOFIS (seconds)	1805s	606s	1663s	1058s	1837s	6969s
Difference (%)	-36%	-39%	-32%	-41%	-32%	-35%
No. of images	0.51	0.45	0.48	0.55	0.54	0.5
No. of videos	0.36	0.54	0.39	0.45	0.26	0.4

We observe an authoring time improvement of 35% on average, in the favor of ATOFIS. We speculate therefore that, from a time perspective, our proposed authoring system will outperform Guides by 35% on average in a similar context. By considering the low number of data points (5), a definitive claim cannot be made, however.

Further, a remarkable finding observed from the time measurements (see Table 11), where the average authoring times required to create a single AR instruction, per workstation, by using ATOFIS and Guides, respectively, are presented. It is interesting to note the extremely low standard deviation (SD =1.32) and variance (VAR=1.76) between the average authoring times of an AR instruction, for all workstations, reported by ATOFIS.

By considering the fact that the evaluated workstations are representative, they were uniformly selected, and their cycle times is very similar (guaranteed by the assembly line balancing procedure), we could expect that the authoring time of a new workstation, Wx, composed of N assembly instructions, will approximately be N*53.2 seconds. We highlight however, the fact that our estimation is based on a limited number of data points, five. Unlike ATOFIS, Guides reported higher values for both the standard deviation (SD=6.61)

Table 11. Average AR instruction authoring time, per workstation

Workstation	W1	W2	W3	W4	W5	Total avg.
ATOFIS (seconds)	54s	55s	51s	53s	53s	53.2s
Guides (seconds)	85s	90s	74s	90s	77s	83.2s

and the variance (VAR =43.76). This data suggests that a better authoring time approximation could be estimated for our proposed authoring system, information that plays an important role in the planning and potentially adoption of the proposed AR system in a similar industrial use case.

5. Discussion

5.1. Training Field Experiments (FE1 & FE2)

The percentage of participants committing errors during the WEC, 66% in each of the two groups (L-GA and L-CAD), invalidates the hypothesis that novice workers can perform the training completely unsupervised. CAD-complemented AR instructions lead to faster ACT (-7% in the first assembly cycle, -20% in the second and -22% in the third), but to a higher error rate during the first assembly cycle (+38%). These results suggest that (i) FPV video demonstrations are more reliable for conveying error-prone assembly operations to novice workers and secondly, that (ii) CAD-based instructions lead to faster assembly for operations of type AT2, especially after the WEC. It seems therefore that FPV video demonstrations are more effective for conveying assembly information during the WEC, however leading to higher assembly times in the following cycles, when compared to CAD models. The reported IRT shows that G-CAD uses 29% less time for reading the instructions during the WEC, indicating that participants prefer CAD-based guidance to video demonstrations. By considering the higher error rate and the lower IRT of G-CAD during the WEC, we speculate that CAD models persuade higher user confidence and lower user attentiveness. The ACT decrease of only 7% in the favor of G-CAD and the error rate increase of 38% reported during the first assembly cycle support the hypothesis than video demonstrations are more effective than CAD models for conveying assembly instructions of type AT2 to novice workers. The overall decrease of the IRT by 47% and 35% in the second and third assembly cycles, respectively, suggest that participants become less dependent on the AR instructions rapidly, questioning therefore the worthiness of authoring CAD-based AR instructions in similar industrial use cases.

A comparative evaluation between the authoring of the two instruction sets used in the field experiments FE1 and FE2 was not conducted. However, we strongly argue, based on the work carried out for authoring the two instruction sets, that the overall authoring effort (technical expertise and time) for creating CAD-based AR instructions is significantly higher compared to creating low-cost-based AR instructions. By considering that only a single error of type non-AT2 was reported during all assembly cycles, we anticipate that low-cost visual assets can effectively convey assembly information of type AT1, AT3 and AT4 in similar assembly use cases, even during the WEC. Additionally, all participants, independently on their assembly performance reported during the experiment, believe that spatially registered low-cost visual assets are sufficient for conveying assembly instructions via AR. Our picking technique differs from the one proposed in [11], however, the effectiveness of pick-by-AR technique is supported by the results of our field experiments.

Assembly experience leads to a better training performance. We observed during the experiment that participants with assembly experience perform significantly faster in AT3 tasks (e.g. screwing or riveting) and commit fewer errors (-52%) during the WEC. They do not commit errors during the following cycles and their ACT is better, independently on the group: -9%, -19% and -16% on average over the three assembly cycles. AR experience however does not affect the performance. We do not observe a lower mental workload nor usability advantages for participants with AR experience, potentially demonstrating the usability of the proposed AR training method.

The ACT of G-CAD suggests that registered CAD models allow a faster identification of the position and orientation of the assembly components, as long as a precise spatial registration is guaranteed. However, the reliability of spatially registered CAD-based AR instructions is questioned until an accurate continuous object registration technique will be provided. We note that the CAD-based AR training experience is highly dependent on the quality of the spatial registration, a concern that was partially addressed in the experiment by the assembly environment itself, as most workstation components had a fixed position, unchanged between the authoring and training procedures. We note as well that CAD models seem to interfere visually with the assembly location, making it difficult to perform operations in non-obvious locations, as reported by few participants. A similar concern was observed in video-based instructions as well, where some of the participants spent more time than expected to identify the corresponding real world assembly location indicated in the demonstration video.

Finally, it seems that successive and repetitive assembly operations like screwing and riveting can be grouped and conveyed as a single AR instruction. Participants performing at least three assembly cycles either suggested or agreed on this affirmation, indicating that the instruction chunking technique was not adapted for certain repetitive assembly operations, particularity after the WEC.

5.2. Authoring Field Study (FS) Experiment

The preliminary data collected during the authoring field study indicates that an expert in manipulating the proposed authoring system, would require approximately 30 minutes for capturing his assembly expertise of a workstation with a nominal cycle time of approximately 3 minutes and 30 seconds (30-35 assembly tasks), in a similar assembly context. The media types captured during the authoring of the five selected workstations for the evaluation indicate that 40% of the assembly operations require video demonstrations, 50% require an image while 10% only require a text description and an indication to the physical location of the assembly. Together with the average reported authoring time per instruction (see Table 11), this information potentially allows one to estimate the authoring time of a workstation that is not representative for the considered use case (e.g., automotive). To do so, the corresponding assembly tasks should be grouped by difficulty into: "*require video demonstration*" (complex assembly task), "*require image*" (rather easy task), "*no visual information required*" (elementary/routine task). We are not aware of a standard method for objectively classifying the assembly operations by their difficulty. Consequently, we believe that the author of the AR instructions, a shop floor expert, potentially an assembly instructor, should decide what type of description each assembly requires for a novice worker to be able to perform it correctly and efficiently.

Secondly, we demonstrated that our proposed authoring method performs, on average, 35% faster than Guides. The preliminary collected data suggests that a precise authoring time estimation of a manual assembly workstation, similar to our use case, can be made by using our proposed method. In addition to the authoring time gain, we expect that our proposal has other advantages including faster learning curve, lower mental and physical effort, less expertise required and ultimately better adapted to industrial usage. In future work we will conduct a participant-based evaluation of the proposed AR authoring system, to validate the aforementioned claims and to measure the usability and the user preference of the system as well.

6. Conclusions

In this research work, we presented an AR training system for manual assembly, adapted to industrial context. We discussed the design and the implementation of the proposed AR authoring tool, dedicated to shop floor experts for capturing assembly knowledge in a one-step authoring process, entirely performed in an HMD AR device (i.e., Hololens 2). Further, we presented how the captured information, represented by a set of step-by-step instructions, is conveyed, and consumed by novice workers via AR, for training purposes. During our long-term case study, we found that, to address industrial challenges and requirements, the best compromise was to rely the proposed AR training system on low-cost visual assets like text, image, video, and predefined auxiliary content, instead of CAD data and animations. To validate our hypotheses, we conducted two field experiments in a real-world industrial use case.

The findings of the first field experiment (FE1) suggested that spatially registered 2D low-cost visual assets are sufficient and effective for conveying manufacturing expertise to novice workers via AR. In the second field experiment, FE2 (an extension of the first), we comparatively evaluated a CAD-complemented instruction set with the initial one (low-cost-based) to identify potential benefits of conveying assembly information by using non-animated, registered CAD models. We found that CAD data persuades lower user attentiveness, eventually leading to a higher error rate for components with a high degree of symmetry (error-prone), but to faster overall assembly completion times, particularly after the WEC. By considering the progress of the time spent by the participants in reading the AR instructions over three assembly cycles, we concluded that the worthiness of authoring CAD-based instructions in similar industrial use cases is questionable, until significant technical and organizational AR challenges are addressed. The overall reported effectiveness and usability scores are favorable, indicating that the proposed AR training method can potentially be used in concrete real world industrial use cases, with a remark that a better technique for underlining subtle assembly details is required for ensuring error-free, completely unsupervised AR training procedures. We expect that our approach can be generalized and adopted in other manufacturing use cases where the 2W1H principle can be applied.

In the third experiment, a preliminary field study (FS), we aimed at evaluating the proposed authoring method from a time perspective, to estimate the required authoring time of a lambda assembly workstation. In addition, we compared our proposal with the most representative industrial AR authoring tool, Guides, and found that our system is 35% faster. Secondly, the evaluation results suggest that our authoring system provides a better time estimation for creating AR instructions, presumably for workstations that are not representative to the considered use case.

Finally, we believe that the industry does not need to wait for better registration techniques, 3D content authoring processes or interfaces. We demonstrated that easy to author, low-cost visual assets together with specific interaction and visualization techniques available in state-of-the-art AR devices could provide effective AR training experiences in complex, real-world industrial environments. At the same time, we demonstrated that organizational and technical AR challenges could be overcome, as long as the conception of the solution is elaborated and tested in the right context, with the direct involvement of the potential end users.

7. Future Work

The main limitation of our work is that the proposed training method was evaluated on a single assembly workstation. To obtain unquestioning statistical data regarding its effectiveness and usability, full training procedures involving novice assembly workers and multiple workstations might be required. We anticipate that future evaluations considering detailed user profiles including cognitive skills, will reveal important findings regarding optimal ways of conveying profile-adapted instructions in AR. From a training perspective, we plan to conduct a large-scale evaluation in other manual assembly use cases and to extend the current evaluation (FE2) to multiple workstations, ideally performing complete training procedures on multiple novice workers.

Regarding the authoring of the AR instructions, in our future work we will aim to evaluate objectively the proposed AR authoring method by conducting a field experiment, ideally by involving line managers and other shop floor experts. Similarity with the first two experiments described in this paper (FE1 and FE2), we will aim to measure the effectiveness and the usability of the overall training system as follows: prior to authoring the AR instructions, participant N will be trained by using the AR instructions created by participant N-1. We will objectively evaluate our proposed AR system against Guides, by alternating the two systems, for both the training and the authoring procedures, for every next participant.

Finally, a comparative evaluation between authoring low-cost versus CAD-based AR instructions might be conducted to objectively evaluate the authoring effort of the two instruction sets and demonstrate our hypothesis, which questions the worthiness of creating CAD-based AR instructions for conveying assembly expertise in similar industrial context.

References

1. Bellalouna, F.: Industrial use cases for augmented reality application. In: 2020 11th IEEE International Conference on Cognitive Infocommunications (CogInfoCom). pp. 10–18 (2020), https://doi.org/10.1109/CogInfoCom50765.2020.9237882

2. Bosch, T., Könemann, R., Cock, H., Rhijn, G.: The effects of projected versus display instructions on productivity, quality and workload in a simulated assembly task. In: Proceedings of the 10th International Conference on PErvasive Technologies Related to Assistive Environments (PETRA '17. p. 412–415. Association for Computing Machinery, New York, NY, USA (2017), https://doi.org/10.1145/3056540.3076189, dOI:
3. Brooke, J.: System Usability Scale (SUS). https://www.usability.gov/how-to-and-tools/methods/system-usability-scale.html (1986), online; accessed 2021-07-06
4. Caudell, T., Mizell, D.: Augmented reality: an application of heads-up display technology to manual manufacturing processes (2003), https://doi.org/10.1109/hicss.1992.183317
5. Dey, A., Billinghurst, M., Lindeman, R., Swan, J.: A systematic review of 10 years of augmented reality usability studies: 2005 to 2014. Front. Robot. AI 5 (2018), https://doi.org/10.3389/frobt.2018.00037
6. Egger, J., Masood, T.: Augmented reality in support of intelligent manufacturing – a systematic literature review. Comput. Ind. Eng 140, 106195, (2020-02), https://doi.org/10.1016/j.cie.2019.106195
7. Funk, M., Kosch, T., Schmidt, A.: 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, (2018-07), https://doi.org/10.1145/2971648.2971706
8. Gabbard, J., Fitch, G., Kim, H.: Behind the glass: Driver challenges and opportunities for ar automotive applications. Proc. IEEE 102(2), 124–136, (2014), https://doi.org/10.1109/JPROC.2013.2294642
9. Gattullo, M., Evangelista, A., Uva, A., Fiorentino, M., Gabbard, J.: 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 2626(c), 1–1, (2020), https://doi.org/10.1109/tvcg.2020.3014614
10. Hahn, J., Ludwig, B., Wolff, C.: Augmented reality-based training of the pcb assembly process. In: Proceedings of the 14th International Conference on Mobile and Ubiquitous Multimedia (MUM '15. p. 395–399. Association for Computing Machinery, New York, NY, USA (2015), https://doi.org/10.1145/2836041.2841215, dOI:
11. Hanson, R., Falkenström, W., Miettinen, M.: Augmented reality as a means of conveying picking information in kit preparation for mixed-model assembly. Comput. Ind. Eng 113(August), 570–575, (2017-11), https://doi.org/10.1016/j.cie.2017.09.048
12. Hart, S.G.: Nasa-Task Load Index (NASA-TLX); 20 Years Later. In: Proceedings of the Human Factors and Ergonomics Society Annual Meeting. p. 904–908 (2016), https://doi.org/10.1177/154193120605000909.K
13. Irawati, S., Green, S., Billinghurst, M., Duenser, A., Ko, H.: An evaluation of an augmented reality multimodal interface using speech and paddle gestures. Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics (January), 272–283, (2006), https://doi.org/10.1007/11941354_28
14. Kim, K., Billinghurst, M., Bruder, G., Duh, H., Welch, G.: Revisiting trends in augmented reality research: A review of the 2nd decade of ismar (2008-2017. IEEE Trans. Vis. Comput. Graph 24(11), 2947–2962, (2018), https://doi.org/10.1109/TVCG.2018.2868591
15. Lavric, T., Bricard, E., Preda, M., Zaharia, T.: An AR Work Instructions Authoring Tool for Human-Operated Industrial Assembly Lines. In: 2020 IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR. pp. 174–183, (2020), https://doi.org/10.1109/AIVR50618.2020.00037
16. Lavric, T., Bricard, E., Preda, M., Zaharia, T.: Exploring low-cost visual assets for conveying assembly instructions in ar. In: 2021 International Conference on INnovations in Intelli-

gent SysTems and Applications (INISTA. pp. 1–6, (2021), https://doi.org/10.1109/INISTA52262.2021.9548570
17. Lavric, T., Bricard, E., Preda, M., Zaharia, T.: An industry-adapted ar training method for manual assembly operations. In: HCI International 2021 - Late Breaking Papers: Multimodality, eXtended Reality, and Artificial Intelligence. pp. 282–304. Springer International Publishing, Cham (2021), https://doi.org/10.1007/978-3-030-90963-5_22
18. Lee, G., Nelles, C., Billinghurst, M., Kim, G.: Immersive Authoring of Tangible Augmented Reality Applications Introduction Application Domain Analysis Immersive Authoring Design. Ismar (2004), https://doi.org/10.1109/ISMAR.2004.34
19. Lee, G.A., Ahn, S., Hoff, W., Billinghurst, M.: Enhancing first-person view task instruction videos with augmented reality cues. In: 2020 IEEE International Symposium on Mixed and Augmented Reality (ISMAR). pp. 498–508 (2020), https://doi.org/10.1109/ISMAR50242.2020.00078
20. Li, W., Wang, J., Jiao, S., Wang, M., Li, S.: Research on the visual elements of augmented reality assembly processes. Virtual Real. Intell. Hardw 1(6), 622–634, (2019-12), https://doi.org/10.1016/j.vrih.2019.09.006
21. Lopik, K., Sinclair, M., Sharpe, R., Conway, P., West, A.: Developing augmented reality capabilities for industry 4.0 small enterprises: Lessons learnt from a content authoring case study. Comput. Ind 117, 103208, (2020-05), https://doi.org/10.1016/j.compind.2020.103208
22. Lorenz, M., Knopp, S., Klimant, P.: Industrial augmented reality: Requirements for an augmented reality maintenance worker support system. In: Adjun. Proc. - 2018 IEEE Int. Symp. Mix. Augment. Reality, ISMAR-Adjunct 2018. pp. 151–153, (2018), https://doi.org/10.1109/ISMAR-Adjunct.2018.00055
23. Lu, F., Davari, S., Lisle, L., Li, Y., Bowman, D.: Glanceable ar: Evaluating information access methods for head-worn augmented reality. In: 2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR), Mar. pp. 930–939, (2020), https://doi.org/10.1109/VR46266.2020.00113
24. Marquardt, A., Trepkowski, C., Eibich, T., Maiero, J., Kruijff, E., Schoning, J.: Comparing non-visual and visual guidance methods for narrow field of view augmented reality displays. IEEE Trans. Vis. Comput. Graph pp. 1–1, (2020), https://doi.org/10.1109/tvcg.2020.3023605
25. Martinetti, A., Marques, H., Singh, S., Dongen, L.: Reflections on the limited pervasiveness of augmented reality in industrial sectors. Applied Sciences 9. 3382 (2019), https://doi.org/10.3390/app9163382
26. Masood, T., Egger, J.: Augmented reality in support of industry 4.0—implementation challenges and success factors. Robotics and Computer-Integrated Manufacturing 58, 181–195 (2019), https://www.sciencedirect.com/science/article/pii/S0736584518304101
27. Masood, T., Egger, J.: Adopting augmented reality in the age of industrial digitalisation. Comput. Ind. 115(C) (feb 2020), https://doi.org/10.1016/j.compind.2019.07.002
28. Mengoni, M., Ceccacci, S., Generosi, A., Leopardi, A.: Spatial augmented reality: An application for human work in smart manufacturing environment. Procedia Manuf 17, 476–483, (2018), https://doi.org/j.promfg.2018.10.072
29. Merino, L., Schwarzl, M., Kraus, M., Sedlmair, M., Schmalstieg, D., Weiskopf, D.: Evaluating mixed and augmented reality: A systematic literature review (2009-2019 (2020-10), https://doi.org/10.1109/ISMAR50242.2020.00069.
30. Microsoft: HoloLens 2 — AR Headset. https://www.microsoft.com/en-us/hololens/hardware (2021), online; accessed 2021-07-06

31. Microsoft: Introducing instinctual interactions. https://docs.microsoft.com/en-us/windows/mixed-reality/design/interaction-fundamentals (2021), online; accessed 2021-07-06
32. Microsoft: Microsoft Mixed Reality Toolkit v2.4.0. https://github.com/microsoft/MixedRealityToolkit-Unity/releases/tag/v2.4.0 (2021), online; accessed 2021-07-06
33. Microsoft: Mixed Reality Dynamics 365 Guides (2021), https://dynamics.microsoft.com/en-us/mixed-reality/guides/, , accessed: 27 April 2021.
34. Microsoft: Overview of authoring a guide in Dynamics 365 Guides. https://docs.microsoft.com/en-us/dynamics365/mixed-reality/guides/authoring-overview (2021), online; accessed 2021-07-06
35. Nicolai, T., Sindt, T., Kenn, H., Witt, H.: Case study of wearable computing for aircraft maintenance. 3rd Int. Forum Appl. Wearable Comput (June), 1–12, (2006)
36. Nizam, S., Abidin, R., Hashim, N., Lam, M., Arshad, H., Majid, N.: A review of multimodal interaction technique in augmented reality environment. Int. J. Adv. Sci. Eng. Inf. Technol 8(4–2), 1460, (2018-09), https://doi.org/10.18517/ijaseit.8.4-2.6824
37. Palmarini, R., Erkoyuncu, J.A., Roy, R., Torabmostaedi, H.: A systematic review of augmented reality applications in maintenance. In: Robotics and Computer-Integrated Manufacturing, vol. 49 (2018), https://doi.org/10.1016/j.rcim.2017.06.002
38. P.T.C.: Vuforia expert capture (2021), https://www.ptc.com/en/products/augmented-reality/vuforia-expert-capture, accessed:
39. Quandt, M., Knoke, B., Gorldt, C., Freitag, M., Thoben, K.: General requirements for industrial augmented reality applications. Procedia CIRP 72, 1130–1135, (2018), https://doi.org/10.1016/j.procir.2018.03.061
40. Sanna, A., Manuri, F., Lamberti, F., Paravati, G., Pezzolla, P.: Using handheld devices to support augmented reality-based maintenance and assembly tasks. In: 2015 IEEE International Conference on Consumer Electronics (ICCE), Las Vegas. pp. 178–179,. NV (2015), https://doi.org/10.1109/ICCE.2015.7066370
41. Sethi, A., Sethi, S.: Flexibility in manufacturing: A survey. Int. J. Flex. Manuf. Syst 2(4), 289–328, (1990), https://doi.org/10.1007/BF00186471
42. Souza Cardoso, L., Mariano, F., Zorzal, E.: A survey of industrial augmented reality. Comput. Ind. Eng 139, 106159, (2019-11), https://doi.org/10.1016/j.cie.2019.106159
43. Tainaka, K., Fujimoto, Y., Kanbara, M., Kato, H., Moteki, A., Kuraki, K., Osamura, K., Yoshitake, T., Fukuoka, T.: Guideline and tool for designing an assembly task support system using augmented reality. In: 2020 IEEE International Symposium on Mixed and Augmented Reality (ISMAR). pp. 486–497. IEEE Computer Society, Los Alamitos, CA, USA (nov 2020), https://doi.ieeecomputersociety.org/10.1109/ISMAR50242.2020.00077
44. Tang, A., Owen, C., Biocca, F., Mou, W.: Comparative effectiveness of augmented reality in object assembly. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. p. 73–80. CHI ’03, Association for Computing Machinery, New York, NY, USA (2003), https://doi.org/10.1145/642611.642626
45. UnityTechnologies: Unity3d - 2019.4. https://unity3d.com/get-unity/download/archive (2021), online; accessed 2021-07-06
46. Uva, A., Gattullo, M., Manghisi, V., Spagnulo, D., Cascella, G., Fiorentino, M.: Evaluating the effectiveness of spatial augmented reality in smart manufacturing: a solution for manual working stations. Int. J. Adv. Manuf. Technol 94(1–4), 509–521, (2018), https://doi.org/10.1007/s00170-017-0846-4
47. Wang, X., Ong, S.K., Nee, A.: A comprehensive survey of augmented reality assembly research. Adv. Manuf 4(1), 1–22, (2016), https://doi.org/10.1007/s40436-015-0131-4

Traian Lavric is an industrial PhD student at Bosch France and Télécom SudParis. He received an Engineering Degree in Computer Science from Transylvania University Brașov and worked as R&D engineer at Télécom SudParis. His research interests include augmented and virtual reality, content authoring, human-computer interaction and interactive multimedia content.

Emmanuel Bricard is IT director of elm.leblanc, French leader in thermal comfort and part of Bosch group. Engineer and holder of an Executive MBA, he heads the R&D department ”Innovation Center for Operations” that focuses on Industry 4.0 technologies, from IoT to augmented reality.

Marius Preda is associate professor at “Institut MINES-Télécom” and Chairman of the 3D Graphics group of ISO MPEG. He contributed to various ISO standards with technologies in the fields of 3D graphics, virtual worlds and augmented reality. He received a Degree in Engineering from POLITEHNICA Bucharest, a PhD in Mathematics and Informatics from University Paris V and an eMBA from IMT Business School, Paris.

Titus Zaharia received the Engineering degree in electronics and telecommunications and the M.Sc. degree from University POLITEHNICA, Bucharest, in 1995 and 1996, respectively, and the PhD in mathematics and computer science from University Paris V. He is a full professor in Télécom SudParis. His research interests include visual content representation methods, with 2D/3D compression, reconstruction, recognition, and indexing applications.

Received: October 15, 2021; Accepted: December 01, 2021.