Augmented Reality for Operator Training on Industrial Workplaces – Comparing the Microsoft HoloLens vs. Small and Big Screen Tactile Devices

Andreas Pusch and Frédéric Noël[(✉)]

Univ. Grenoble Alpes, CNRS, Grenoble INP, G-SCOP, 38000 Grenoble, France
frederic.noel@grenoble-inp.fr

Abstract. The digital revolution towards the industry standard 4.0 offers many ways to improve established methods and processes. In this paper, we report on the lessons learned about the pros and cons of Augmented-Reality-based operator training using the Microsoft HoloLens as compared to small and big screen tactile devices. Together with our industrial partner, we have chosen an encapsulation assembly task as use case. We have enriched the original training material with digital twins of the workplace, animations, videos, and contemporary forms of interaction, all of which made available in an optimised fashion on three different support technologies. Feedback from our testers, and those in charge of designing training courses, is suggesting that notably the HoloLens version of our prototype has the potential not only to replace current training methods, but to go beyond them up to the point where even novices can pass the training autonomously. It thus seems promising to integrate Augmented Reality into training programmes and so to complete the digital chain within the industry life management.

Keywords: Industry 4.0 · Digital twins · Augmented Reality · MS HoloLens · Tactile displays · Operator training · Assembly tasks · Ergonomics · Usability

1 Introduction

In the course of the industrial revolution towards the *Industry 4.0* standard (or *Smart Factories*, or *Cyber-Physical Systems*), systems and processes become entirely inter-operable, information fully transparent up to *digital twins*, and automated assistance available at both decision making and operational levels. Communication and cooperation between all involved entities takes place in real-time, internally as well as across organisations, allowing for highly efficient value chains and product life cycles.

This paper addresses issues related to the automated assistance during operator training in the automotive supplier industry, notably for the assembly of encapsulations. We consider this an important first step towards other through-life services, such as maintenance or everyday operation support. Having the factory and all its workplaces represented as virtual models helps transform traditional *work instructions*

(WI), often available only on paper, card board, or as static charts, into multi-platform, interactive training modules that benefit from all the recent advances also in fields of *Human-Computer-Interaction (HCI)* and *Augmented Reality (AR)* (see [1] for a survey).

Our overall objectives consisted of (**1**) integrating the digital workplace and product twins with the standard WI used at our industrial partner's factory (=> *Cooper Standard Vitré*, France), (**2**) providing interactive, automated training modules optimised for three platforms (=> *Microsoft HoloLens, NVIDIA SHIELD tablet, iiyama ProLite touch-screen monitor*), while (**3**) designing each version iteratively and compliant with their specific ergonomic principles (=> mainly adopted from the HCI domain), in order to (**4**) compare their utility experimentally by also looking at potentials and limitations.

However, novel AR applications and processes will have to prove their added value before they can become an integral part of the everyday manufacturing business. With this comparative work, we intent to take the current state of the art one step forward towards identifying the potentials and limitations of recent devices and methods.

2   Related Work

Production line workers usually are specialised on specific tasks or task sequences, such as the part assembly we have studied at Cooper Standard Vitré. But since products like car models, for instance, change rapidly, they also need to adapt to new workplaces within short periods of time. In addition, there is a relatively high fluctuation among workers, as workforce requirements get continuously adjusted to the factory's actual production needs. As a consequence, factories often employ interim staff, which makes efficient training a crucial element for an efficient operational functioning. We will here review "traditional" training methods as well as the current streams in HCI exploring modern AR technologies and interaction techniques for the use in industrial training contexts. We further discuss the ergonomic foundations for interactive AR applications.

2.1   "Traditional" Operator Training

In many cases, training is being done "on the job". An experienced worker takes the role of a trainer, unless there is dedicated training staff available, and the trainee discovers step-by-step what is to be done and how. Tasks are either being explained or demonstrated, and then have to be repeated in a training environment, or, sometimes, directly be applied productively.

A means to standardise the training process are static WI charts. "Static" means that they are printed on paper, card board, or any other type of support. The trainer uses these instructions, which typically include brief textual descriptions, pictures, and critical checks to be performed, for recalling and communicating all key elements, whereas the trainee uses them as basic learning tool. Advancement and success are being evaluated and approved by the trainer.

## 2.2 Operator Training in Augmented Reality

Overview of AR Technologies and Interaction Techniques. While training in *Virtual Reality (VR)* has been studied and applied for decades in various professional domains [2], AR had (and still has) to overcome major technological issues before reaching the maturity required for practical use. When talking about AR, we can differentiate three main approaches to display augmented information on top of the real world – or images of it – each having its own advantages and disadvantages: (1) *Video see-through AR* [3], (2) *optical see-through AR* [4], and (3) *spatial AR* [5]. For the sake of conciseness, we will here only focus on optical see-through AR, which uses prisms, semi-transparent micro displays, or retinal laser scanners in front of one or both eyes to let images "float" on top of the real environment. Tracking techniques are sensor- or camera-based. However, due to the optical complexity and other technological challenges, it was only recently, with the advent of devices such as Microsoft's HoloLens [6] (=> for further details, see Sect. 4.1), that we see optical see-through AR passing a fundamental frontier.

Operator Training in AR. AR has been studied for many years in the context of education and training (see e.g. [1] and [7] for extensive surveys). Hence, various applications, methods, and advanced strategies for information presentation exist, meant to facilitate assembly tasks in a number of domains, incl. the restricted medical sector [8], covering hierarchical structures [9], but also to assist in maintenance and repair [10]. However, it has been found that AR still does not meet the rigorous requirements of operational use in the industry [11], in part due to device constraints, but also due to the way how and when interactive augmented contents are being displayed. Radkowski et al. [12] reported that it is crucial to present appropriate visual features during assembly, corresponding to the difficulty of the task, whereas [13] introduced a rule-based expert system optimising information presentation depending on the trainee's progress. In [14], an (automated) process for the actual creation of augmented contents based on assembly instructions and 3D product models is being proposed.

Ergonomics Principles for Interactive AR. For AR to be successfully employed in industrial environments, it is vital that methods and applications be robust and error-prone. Beside best practices and domain experience acquired in the field over time, another central source for the design of reliable and easy to use systems are national and international norms and standards. Although AR is still rarely being mentioned explicitly (=> a number of related standards are currently under development), we have been able to identify recommendations for system and interaction design, as well as for human factors in a broader sense. As a basis for our research, we have used the comprehensive EN ISO 9241 (*The ergonomics of human system interaction*) [15], with Travis' guide as an initial orientation aid [16], knowing that there are other important standards to be taken into account in the future. These principles have been among the main elements used for the initial design of our below training prototypes, before we have iteratively refined them together with our industrial partner.

3   Assembly Use Case

Workplaces in our partner's factory are arranged along different production lines. The workplace chosen for further study is called the *Finishing table*, a final assembly workplace (see Fig. 1). As it is usually being operated standing, its height can be adjusted by a wired controller, while the lifting mechanism can be interrupted by an emergency stop. The tabletop can also be inclined. The encapsulation to be assembled is made up of 5 separate pieces. During assembly, various checks and other actions have to be carried out, as indicated in the WI (=> not shown here for reasons of confidentiality).

(a)   (b)
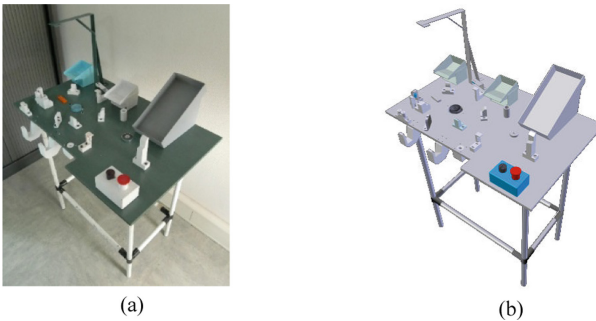
Fig. 1. The *Finishing table*: (a) the physical prototype and (b) its digital mock-up.

4   Prototypes

4.1   AR Prototype Using the Microsoft HoloLens

The MS HoloLens. The MS HoloLens is a powerful, hands-free, completely self-contained optical see-through head-mounted display (HMD) which allows to see virtual contents overlaid on the real environment. It comes with 6 degrees-of-freedom (DoF) tracking, environment understanding, spatial sound, as well as with basic gesture interaction and voice recognition. The resolution per eye is 1268×720 pixels, the field-of-projection (FoP) approx. $30°$ horizontally $\times$ $17.5°$ vertically. It is further equipped with a 2 megapixels forward-pointing camera. Various possibilities exist to fix the device to the user's head. There also is a small physical button device, called the *clicker*, which can be strapped to the middle finger, and then be held between the index and the thumb.

AR Training Application. In compliance with the protocols at Cooper Standard Vitré, we had to organise our training applications in phases and subphases. We further had to make sure that all key contents have been visited by the trainee, before allowing her or him to proceed. The application starts each training phase with an overview of all steps to be completes (see Fig. 2a). At first, the trainee would see only the main labels, later also the detailed instructions, pictures, and animations etc. (see Fig. 2b). To avoid or limit occlusions of the real environment, added contents can always be switch off. Videos for

each step are available at any time in the video pane, which also accommodates alternative navigation controls as well as the access to the overview. Trainees can navigate forth and back through the steps, pause videos and animations, replay audio instructions, and reposition the video pane, if needed.
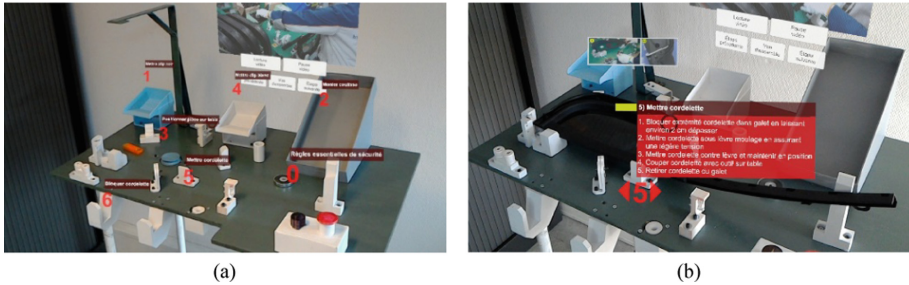
Fig. 2. Screenshots of the training application in AR mode: (a) Overview of the assembly task, (b) WI details of a specific step, plus added contents. (Color figure online)

Interactive user interface (UI) items and buttons can be selected using a 3D cursor and then performing either the *Air tap* gesture, saying "select", or using the clicker.

Limitations. Among the most important limitations of the current version of the HoloLens is its limited FoP. The AR training application has been designed to both show where the actual FoP ends (=> a blue border frame, not shown above) and guide the trainee's gaze, if relevant objects are being shown outside the current AR view. So, whenever any such object appears, its location with respect to the main camera's viewing frustum (=> the user's view) will be determined, and guidance cues (=> red arrows for step transitions, yellow arrows for additional contents of a given step) will be shown at the extremities of the projection space. These cues disappear as soon as the target object has been "found", so, when it has been seen or gazed at by the trainee.

Further limitations are the weight of the device, but also the modes of interaction. The Air tap gesture turned out to be difficult to perform, and the default speech recognition is restricted to US English. Rapid head movements may decompose the RGB images into their primary components, leading to chromatic aberrations. Head fixations, although rich and flexible, are complicated to adjust. As for the tracking, fluctuations of a few cms will have to be accepted. The device is generally rather fragile, making it risky to be used, notably in industrial settings.

Implementations Details. The training application has been developed on a standard workstation installed with *Unity 2017.3.0f3, Visual Studio Community* (needed for deployment), *Windows 10 SDK*, and all respective, freely available *HoloLens toolkits*. All 3D models have been generated with SolidWorks, and further been exported to the OBJ file format. Once running on the device, an external keyboard is needed to control certain aspects of the application, notably the manual calibration for aligning the virtual to the real workplace. A stopwatch module registers various times for further analyses.

4.2 Tablet Prototype

NVIDIA SHIELD Tablet. This (gaming) tablet comes with an $8''$ multi-touch display ($=>1920 \times 1200$ pixels). We have installed it with Android 7.0. It is equipped with two 5 megapixels cameras (=> forward- and backward-pointing), a microphone, as well as with position, orientation, and location tracking sensors. The tablet also allows for stylus input, direct stereo sound output, and for connecting up to Bluetooth 4.0 devices.

Tablet Training Application. As in the HoloLens prototype, we had to reflect the existing multi-phase training protocol. While multimodal contents are exactly the same, they are arranged differently to fit the tablet's small display (see Fig. 3a).

(a)   (b)
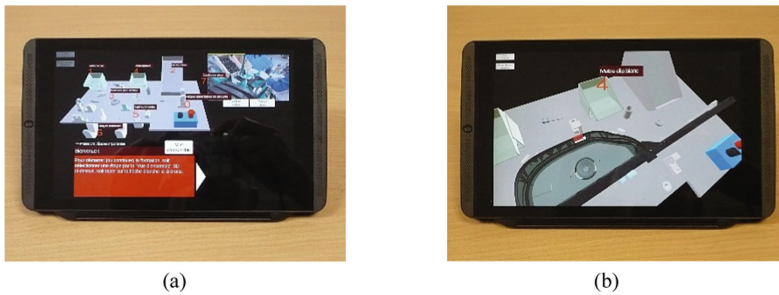
Fig. 3. Tablet prototype: (a) Initial training screen. Note the decomposition of the UI in dedicated zones for animations, videos, textual instructions, and additional images, (b) full-screen 3D view, which can be manipulated freely.

The dedicated zones allow to focus on the essential elements, without having to search for them. All non-textual contents can be zoomed by tapping on them. With one exception: The full-screen 3D view (see Fig. 3b). Once entered, the trainee can move the 3D scene around, turn it, and zoom further into it. Animations can be paused at any time.

Limitations. Despite its autonomy, maturity, and robustness, it will be difficult to use a tablet in a hands-free fashion with it being constantly in front of the eyes. In other words, even if a video see-through AR approach could theoretically be applied, practically, the tablet has to be mounted somewhere on or nearby the workplace. The trainee will thus have to force her- or himself to look at it, which means that information can be missed. Finally, the screen is small, limiting visibility and readability of the training material. For audio instructions, wireless headphones are highly recommended.

Implementation Details. The main differences as compared to the HoloLens prototype are that the *Android SDK* will be required instead of the Windows 10 SDK, and Visual Studio is purely optional. The deployment chain is otherwise completely integrated into Unity (=> for direct deployment, the tablet has to be switch into "developer mode"). All media should be transcoded for the Android target platform.

4.3 Big Screen Prototype

iiyama ProLite Touchscreen Monitor. The iiyama PL2735M has a 24″ full HD (=> 1920 × 1080 pixels) multi-touch display. It can be driven through a variety of graphics ports. Its USB 3.0 host connexion allows for an immediate integration with Windows 10 systems. So, touch events are being handled just as mouse clicks, or as touch events on a tablet. Microphone, camera, and stereo speakers are integrated as well.

Big Screen Training Application. The big screen version of the training application looks and behaves exactly the same as the tablet version. The major difference is in the display size. However, while visually more comfortable and easier to interact with due to its size, not to interfere with the physical workplace, but keeping the monitor accessible at the same time, can be challenge. A separate stand may be a good choice.

Limitations. The biggest advantage of a big display, its size, can quickly turn into its biggest disadvantage, if space matters. Also, manipulating bulkier pieces can be difficult, if a bigger screen is in the way somewhere. In addition, unless a complete system with a computer integrated into the display is being used, it will further be necessary to accommodate space for the central unit, cables etc.

The risk of missing important information is imminent here as well, although visual notifications are more likely to be detected (in the periphery), since they are bigger. For audio instructions, wireless headphones should be provided.

Implementation Details. Made with Unity, too, the development environment corresponds largely to that of the tablet training app (without the Android SDK). In fact, by changing the target platform during the build process, Unity allows to generate platform-specific executables from the same sources. It is also at that time, when all media will (usually automatically) be transcoded for the selected target platform.

5 Pilot Evaluation

5.1 Testing Conditions

Our current training prototypes have been preliminarily evaluated by 2 usability experts (=> heuristic evaluation and walkthroughs), 7 of our project partners from the industry, including 5 training specialists of different responsibility levels from Cooper Standard Vitré (=> free exploration), as well as 10+ invited researchers, engineers, and students (=> informal usability tests). The usability experts and 3 colleagues from Cooper Standard Vitré have also been involved in the iterative design process. Given the nature of the tests conducted so far, it should be noted that the results presented in Sect. 5.2 are mainly based on qualitative data and observations.

5.2   Results

The tables below (see Tables 1, 2 and 3) provide a condensed summary of the feedback obtained. Where appropriate, we explain our observations in more detail.

Table 1. Comparison of the device ergonomics.

	MS HoloLens	NVIDIA SHIELD	iiyama ProLite
Wearing comfort	Poor	Moderate	Good
Viewing comfort	Moderate	Moderate	Good
Field-of-view	Moderate	Poor	Moderate
Visual quality	Good	Poor	Moderate
Acoustic quality	Good	Moderate	Moderate
Hands-free use	Yes	Yes	Yes
Interaction	Multimodal	Tactile	Tactile
Device autonomy	Moderate	Good	n/a
Device robustness	Poor	Good	Moderate

Table 2. Comparison of the training applications as a function of the support technology.

	MS HoloLens	NVIDIA SHIELD	iiyama ProLite
Content visibility	Good	Poor	Moderate
Inform. awareness	Good	Very poor	Poor
Co-located AR	Yes	No	No
Real-time 2D/3D	Both	Both	Both
Comprehension	Good	Moderate	Moderate
Ease-of-use	Good	Poor	Moderate
Acceptance	Good	Moderate	Moderate

Table 3. Overall comparison from the manager's perspective.

	MS HoloLens	NVIDIA SHIELD	iiyama ProLite
Training time	Good	Poor	Moderate
Success rate	Good	Poor	Moderate
Autonom. training	Probably yes	No	Probably not
Auton. validation	At most partially	No	No
Eff. WI alternative	Yes	Probably not	Maybe
Work support tool	No	Yes	Maybe

Managers and training experts further asked about the cost of creating training modules. The good news is that, between systems, there is no big difference. And once created for one platform, having it on another will be considerably less expensive due to shared contents and codes, especially in the case of the two tactile versions which currently do not differ at all. However, it is clear that an authoring tool on top of Unity (or any similar engine) will be required in order to render the production process cost-effective and profitable.

6   Discussion

The benefits of the *MS HoloLens* make it a very interesting device for future industrial applications, incl. AR-based training or maintenance, where the operator immediately sees synthetic contents right on top of the real environment. It is self-contained, wireless, and fully equipped with all necessary tracking, multi-modal feedback, and interaction features. Development using Unity and Visual Studio is fairly straight-forward.

But despite its obvious advantages, it also (still) suffers from important limitations. Beside the poor wearing comfort and the relatively small FoP, its fragility makes it hard to imagine to see it being employed "in the industrial wild" any time soon. In controlled environments, however, it may already substitute traditional training methods.

The *NVIDIA SHIELD tablet*, on the other hand, is a well-proven and robust device. Its computational power is sufficient even for displaying more sophisticated real-time 3D graphics. Interacting with it via touch can nowadays be considered "natural". But its display size and the fact that the trainee has to actively look at it, make it a rather poor candidate for autonomous training programmes. In addition, it requires some space to be fixed to the workplace, so as to assure a hands-free use. It may prove useful, though, as a complement to the traditional training, or as an everyday support tool.

The iiyama ProLite touchscreen monitor is in most cases doing better than the tablet, of which it basically is a (much) bigger version. Its generous display size can become a handicap, if the workplace does not offer enough space to accommodate it. Moreover, manipulating bulky pieces can be a risky endeavour with such a big screen nearby.

To summarise, on most of the dimensions we have investigated, the MS HoloLens performs best, followed by the iiyama ProLite and the NVIDIA SHIELD tablet, resp. But the existing ergonomic constraints suggest to view even the "winner" with some caution! The tablet version of our training application may be a promising tool for recalling certain aspects of the work to be conducted, useful in job rotation contexts.

7   Conclusion and Future Work

We have demonstrated that it is timely and beneficial to integrate the development of novel training methods based on recent advances in HCI and AR into the transition process towards a Factory 4.0. Iterative design and first pilot tests helped us identify both positive and negative points of our three prototypes. The AR version based on the MS HoloLens, with co-located contents being projected right on top of the real workplace, appears to be the most efficient for actual training, despite its ergonomic constraints (=> wearing and viewing comfort). The tactile tablet version, however, may

prove useful as an everyday assistance or recall tool. Somewhere in the middle in terms of utility and acceptance landed the big screen version of our training application. For practical reasons (=> need for more space, risk of being "in the way" while manipulating more bulky pieces), we do not yet see it being employed operationally.

Among our next steps is to conduct formal experiments under lab and field conditions, i.e. with real operators at Cooper Standard Vitré, in order to confirm – or disprove – the various expert expectations. We also plan to study ways to render the validation phase more autonomous. In this context, we will exploit the MS HoloLens's hand tracking combined with the detection of operator actions and workplace state changes.

Finally, we plan to refine our development process, so that it can be applied to other use cases, incl. maintenance and control tasks. To this end, we will generalise code and content production, but also link with the evolving business knowledge. To support this process, and benefit all stakeholders, we intent to deepen and enlarge our partnerships.

References

1. van Krevelen, D.W.F., Poelman, R.: A survey of augmented reality technologies, applications, and limitations. Int. J. Virtual Real. **9**(2), 1–20 (2010)
2. Wang, P., Wu, P., Wang, J., Chi, H., Wang, X.: A critical review of the use of virtual reality in construction engineering, education, and training. Int. J. Environ. Res. Public Health **15**(6), 1204 (2018)
3. Edwards, E.K., Rolland, J.P., Keller, K.P.: Video see-through design for merging of real and virtual environments. In: Proceedings of VRAIS (1993)
4. Kiyokawa, K., Kurata, Y., Ohno, H.: An optical see-through display for mutual occlusion of real and virtual environments. In: Proceedings of IEEE & ACM ISAR (2000)
5. Bimber, O., Raskar, R.: Spatial Augmented Reality: Merging Real and Virtual Worlds. A K Peters/CRC Press, Boca Raton (2005)
6. Microsoft HoloLens (2008). https://www.microsoft.com/en-gb/hololens
7. Lee, K.: Augmented reality in education and training. TechTrends **56**(2), 13–21 (2012)
8. Nilsson, S., Johansson, B.: Fun and usable: augmented reality instructions in a hospital setting. In: Proceedings of 19th Australasian Conference on Computer-Human Interaction, pp. 123–130 (2007)
9. Zauner, J., Haller, M., Brandl, A., Hartman, W.: Authoring of a mixed reality assembly instructor for hierarchical structures. In: Proceedings of ISMAR, pp. 237–246 (2003)
10. Henderson, S., Feiner, S.: Exploring the benefits of augmented reality documentation for maintenance and repair. IEEE TVCG **17**(10), 1355–1368 (2011)
11. Tiefenbacher, P., Lehment, N.H., Rigoll, G.: Augmented reality evaluation: a concept utilizing virtual reality. In: Shumaker, R., Lackey, S. (eds.) VAMR 2014. LNCS, vol. 8525, pp. 226–236. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-07458-0_22

12. Radkowski, R., Herremaa, J., Olivera, J.: Augmented reality-based manual assembly support with visual features for different degrees of difficulty. Int. J. Hum. Comput. Interact. **31**, 337–349 (2015)
13. Syberfeldt, A., Danielsson, O., Holm, M., Wang, L.: Dynamic operator instructions based on augmented reality and expert systems. In: Proceedings CIRP Conference on Manufacturing Systems.(CMS) (2015)
14. Sääski, J., Salonen, T., Hakkarainen, M., Siltanen, S., Woodward, C., Lempiäinen, J.: Integration of design and assembly using augmented reality. In: Ratchev, S., Koelemeijer, S. (eds.) IPAS 2008. IIFIP, vol. 260, pp. 395–404. Springer, Boston (2008). https://doi.org/10.1007/978-0-387-77405-3_39
15. EN ISO 9241 - The ergonomics of human-system interaction (2018). https://www.iso.org/
16. Travis, D.: Bluffers' Guide to ISO 9241, 9th edn. Userfocus Ltd., London (2014)