HoloCollab: A Shared Virtual Platform for Physical Assembly Training using Spatially-Aware Head-Mounted Displays

Markus Funk
Siemens Corporate Technology
Berkeley, USA
funk.markus@siemens.com

Mareike Kritzler
Siemens Corporate Technology
Berkeley, USA
mareike.kritzler@siemens.com

Florian Michahelles
Siemens Corporate Technology
Berkeley, USA
florian.michahelles@siemens.com

ABSTRACT
Today’s industrial jobs require a skilled and trained workforce as tasks such as maintenance, service, and repair are becoming more complicated and more demanding. Therefore, both education and training for executing these tasks are becoming more important. Usually training is conducted on-site at designated training facilities with physical hardware. However, on-site hands-on training can be expensive as it requires designated training facilities that have to be maintained and need to be traveled to. With Augmented Reality (AR) becoming a substantial part of modern day manufacturing, using AR-based systems to train new workforces is becoming increasingly popular. In this paper, we investigate different training environments that use AR-based support during workforce training. We draw the design space for a shared collaborative AR-based learning space, and present the concept and implementation of HoloCollab which combines having a scalable virtual representation of an industrial scenario, such as assembly, with the benefit of having a trainer on-site with a trainee.

ACM Classification Keywords
H.5.2. Information Interfaces and Presentation (e.g. HCI): User Interfaces

Author Keywords
Remote Collaboration; Augmented Reality; Training; Assistive System; Workplaces

INTRODUCTION
A highly skilled and well trained industrial workforce is essential for a successful and productive factory and field service operation. In order to ensure a skilled and educated workforce, training of new workers is necessary. Young high school graduates, who apply for positions at factories or as field technicians, do not have the necessary practical qualifications and need to receive extensive on-the-job training. Often skilled veteran workers are a great source of knowledge and can mentor young workers during this training phase. We are experiencing an increase of these training programs as more and more companies now introduce an apprenticeship program. For example, Siemens created such a program for its gas turbine factory in Charlotte, North Carolina, USA in order to train and educate senior high school students. However, to facilitate such on-site hands-on training requires travel and a lot of hardware, which might not always be available and might be too expensive to buy in large quantities. In addition, procedures that are taught in classroom scenarios might be harder to memorize as it is difficult for trainees to take notes while learning practical hands-on procedures. Given these restrictions and the rise of new technologies, workforce training has the potential to change from traditional classroom type hands-on training to virtual training setups using Augmented Reality (AR).

A number of previous research projects have proposed to use AR to support persons in the domain of education and training, results show that using AR to improve learning...
environments has “vast potential implications and numerous benefits” [24]. For example, research prototypes used 3D web technology and AR in order to teach mechanical engineering to new workers [18]. Using AR technology, the authors conclude that this technology enables trainees to benefit from a new learning experience. There are also recent advancements in commercially available products, which support AR-based teaching experiences. A welding simulator, for example, utilizes AR to keep trainees safe during their training.

While 20 years ago, a matchmaking between requirements for a task and the skills of trainers had to be made to connect two workers that are on-site [20], today this matchmaking can be used to virtually connect to a trainer that is anywhere in the world. Since the invention of the telephone in 1876, researchers and engineers are making effort to create a bridge between two or more physically separated persons using technology. This started with the telephone providing a voice-only link between two remote persons and reaches from video calls [15] to virtual meetings in Virtual Reality using an avatar (e.g. AltspaceVR). Over the last 20 years, this technology has improved a lot [15]. Nowadays this telepresence technology can be used for various scenarios including teaching on-site trainees or supporting service technicians while trainers or experts are at a remote location. This is an important factor looking at business travel. We can see that the requirements for being at a remote location for business purposes is increasing. In fact, according to Statistica, in 2013 there were 444.9 million business trips alone inside the US. And with business becoming more distributed, this number is even expected to rise in the future. Using technology to reduce physical traveling and virtually conducting training sessions is not only beneficial for companies to reduce cost, it is also beneficial for the environment as it reduces the carbon dioxide pollution.

Especially a virtual learning environment [7] opens up new possibilities for both remote trainers and on-site trainees, as it can give two or more persons, who are at different locations, the illusion that they are at the same physical location. In addition to just connect two or more colleagues using a virtual environment, it can also be used to extend the user’s capabilities by adding tools or powers that a trainee would not have in the real environment. Therefore virtual learning environments can give trainers and trainees capabilities that are beyond the capabilities of physically being there [13].

In this paper, we describe the design space for such a shared collaborative AR-based learning space and we present HoloCollab, a novel system using head-mounted displays (HMDs) to create an augmented training experience for both trainer and trainee which implements two dimensions of the presented design space (see Figure 1). The contribution of this paper is two-fold: (1) We investigate the design space of an augmented training scenario, and (2) we present our prototypical implementation of HoloCollab.

While 20 years ago, a matchmaking between requirements for a task and the skills of trainers had to be made to connect two workers that are on-site [20], today this matchmaking can be used to virtually connect to a trainer that is anywhere in the world. Since the invention of the telephone in 1876, researchers and engineers are making effort to create a bridge between two or more physically separated persons using technology. This started with the telephone providing a voice-only link between two remote persons and reaches from video calls [15] to virtual meetings in Virtual Reality using an avatar (e.g. AltspaceVR). Over the last 20 years, this technology has improved a lot [15]. Nowadays this telepresence technology can be used for various scenarios including teaching on-site trainees or supporting service technicians while trainers or experts are at a remote location. This is an important factor looking at business travel. We can see that the requirements for being at a remote location for business purposes is increasing. In fact, according to Statistica, in 2013 there were 444.9 million business trips alone inside the US. And with business becoming more distributed, this number is even expected to rise in the future. Using technology to reduce physical traveling and virtually conducting training sessions is not only beneficial for companies to reduce cost, it is also beneficial for the environment as it reduces the carbon dioxide pollution.

Especially a virtual learning environment [7] opens up new possibilities for both remote trainers and on-site trainees, as it can give two or more persons, who are at different locations, the illusion that they are at the same physical location. In addition to just connect two or more colleagues using a virtual environment, it can also be used to extend the user’s capabilities by adding tools or powers that a trainee would not have in the real environment. Therefore virtual learning environments can give trainers and trainees capabilities that are beyond the capabilities of physically being there [13].

In this paper, we describe the design space for such a shared collaborative AR-based learning space and we present HoloCollab, a novel system using head-mounted displays (HMDs) to create an augmented training experience for both trainer and trainee which implements two dimensions of the presented design space (see Figure 1). The contribution of this paper is two-fold: (1) We investigate the design space of an augmented training scenario, and (2) we present our prototypical implementation of HoloCollab.


**Related Work**

Our research is mainly inspired by work from the areas of AR-supported assistance, and remote assistance or training systems. In the following, we introduce relevant related approaches from these two areas and discuss benefits and drawbacks.

### Augmented Reality Supported Assistance

In the field of providing AR assembly assistance, a number of new systems have been proposed. A comprehensive overview is presented by Büttner et al. [6]. Overall, currently there are three major technologies displaying assistance instructions: projection-based systems, hand-held devices, and HMDs.

In the area of projection-based AR systems, Bannat et al. [1] first introduced projecting pictorial assembly instructions directly on a trainee’s workplace. This idea has been extended by Funk et al. [9], who are using spatially-aware in-situ projection to teach assembly worksteps to workers with cognitive disabilities. Another projection-based remote assistance system is the TeleAdvisor system [12]. This system uses a camera-projector system to enable a remote expert to give instruction to an on-site worker.

When investigating hand-held assistance systems using AR, Gauglitz et al. [10] are using a hand-held device for providing remote assistance instructions for trainees. Their prototype is using image feature tracking to spatially annotate the live view of a hand-held device. The hand-held device can be as large as a tablet computer or as small as a mobile phone [2].

Considering HMD-based assistance systems, Büttner et al. [5] compared using HMDs and projected user interfaces to teach new assembly steps to workers. They concluded that projection-based interfaces lead to a faster assembly and less assembly errors. The TAC system [4] uses 3D elements that are displayed in a worker’s HMD to give instructions for a remote maintenance task. The research combines the TAC prototype with a voice and video communication system to improve the presence of the remote expert. Furthermore, Blattgerste et al. [3] are introducing a cross-hair visualization on the Microsoft HoloLens to visualize targets in assembly tasks. Interestingly, while using HMDs has been proven to be slower than projection-based assistance systems multiple times in related literature, the wearable character of HMDs is very attractive for mobile learning scenarios. Also, recent technological advances (e.g. Microsoft HoloLens) are making HMDs less bulky and are adding spatial awareness.

### Remote Assistance and Training

Considering the topic of remote assistance and training, AR is used in a number of different application areas using a number of different approaches. For example, in the RemoteBob project, Kritzler et al. [14] use a telepresence robot to give remote instructions to an on-site worker. Thereby, the remote expert can trigger actions and annotate the video stream using gestures that are detected using a leap motion controller. The annotated video stream is then presented to an on-site worker
on an iPad which is mounted on a mobile robot. Further, in the WACL project [23], a shoulder-mounted laser pointer is used to show a remote trainee where an action has to be performed. The laser pointer, which is mounted on the trainee’s shoulder, can be controlled by a remote expert. Thereby a mapping of the AR feedback to the physical environment is done automatically without the need for calibrating the device. This has been further evaluated in additional research [21] using a remote Lego-Duplo assembly task. Finally, AR also found its way into classroom learning situations [19], where models of solar systems and plants were taught to the students. Other research [17] used AR technology to let students manipulate and experience virtual objects using AR markers.

Summing up, related work has focused on providing assistance using HMDs, projection, and hand-held devices that augment either existing physical objects or purely virtual objects with instructions and remote help. In the HoloCollab project, we focus on extending the related approaches by providing a scalable platform that supports both remote and on-site training on virtual as well as real objects.

A SHARED COLLABORATIVE AR-BASED LEARNING SPACE

Informed by the related work, we investigate the design space of a shared collaborative AR-based learning space. We identified three design dimensions in the design space of a collaborative learning space: time, location, and materiality (virtual vs. real training objects).

- **Time**: Time means that teaching and learning can either happen synchronously at the same time or asynchronously at different times (e.g. pre-recorded video teaching vs. live tutoring).
- **Location**: Location means that trainer and trainee are either at different locations when the teaching happens or at the same location (e.g. live video teaching vs. on-site teaching).
- **Materiality**: Materiality means that the training object can either be virtual or real (e.g. virtual representation of an object vs. on-site training), which provides a different experience when receiving or giving training.

As we focus on a shared collaborative AR-based learning space, the scope of this paper is limited to a physical trainer that is able to react to problems and is able to communicate with the trainees using body language, voice, and gestures in combination with digital benefits like spatial annotation. Therefore, we constrain the design space of the collaborative learning environment in this paper to happen synchronously at the same time. In the following, we are analyzing each combination of the remaining design dimensions, space and materiality, and identify advantages and disadvantages of each combination. An overview of examples for each dimension of the design space is depicted in Figure 2.

**Same Physical Location - Real Object**
The first combination of the design dimensions mentioned above is a trainer and a trainee that are at the same physical location and that use a real object as their learning platform. This combination resembles a traditional teaching scenario.

**Same Physical Location - Virtual Object**
The second combination of the design dimensions is a trainer and a trainee that are at the same physical location and that use a virtual object as the learning platform.

**A SHARED COLLABORATIVE AR-BASED LEARNING SPACE**

Informed by the related work, we investigate the design space of a shared collaborative AR-based learning space. We identified three design dimensions in the design space of a collaborative learning space: time, location, and materiality (virtual vs. real training objects).

- **Time**: Time means that teaching and learning can either happen synchronously at the same time or asynchronously at different times (e.g. pre-recorded video teaching vs. live tutoring).
- **Location**: Location means that trainer and trainee are either at different locations when the teaching happens or at the same location (e.g. live video teaching vs. on-site teaching).
- **Materiality**: Materiality means that the training object can either be virtual or real (e.g. virtual representation of an object vs. on-site training), which provides a different experience when receiving or giving training.

As we focus on a shared collaborative AR-based learning space, the scope of this paper is limited to a physical trainer that is able to react to problems and is able to communicate with the trainees using body language, voice, and gestures in combination with digital benefits like spatial annotation. Therefore, we constrain the design space of the collaborative learning environment in this paper to happen synchronously at the same time. In the following, we are analyzing each combination of the remaining design dimensions, space and materiality, and identify advantages and disadvantages of each combination. An overview of examples for each dimension of the design space is depicted in Figure 2.

**Same Physical Location - Real Object**
The first combination of the design dimensions mentioned above is a trainer and a trainee that are at the same physical location and that use a real object as their learning platform. This combination resembles a traditional teaching scenario.

**Same Physical Location - Virtual Object**
The second combination of the design dimensions is a trainer and a trainee that are at the same physical location and that use a virtual object as the learning platform.
As previously described, an advantage of trainer and trainee being at the same physical location in a training scenario is to receive full gesture and body feedback as well as full insight. The following benefits occur additionally from using a virtual learning environment: First, switching scenarios is possible very easily as just a different model needs to be loaded in the virtual environment. Also especially when it comes to huge and expensive machines, using a virtual object to train trainees is less expensive than training on the real object. A trainer can scale up the number of trainees that are being supervised at the same time. Multiple trainees can share the same physical space with the trainer, and they can have their own virtual learning space. In case a direct supervision is needed, the trainer can just join the trainee in their private virtual learning scenario and provide help when a task is unclear or when the trainee is struggling. As a student might not need supervision the whole time, this scenario is ideal for optimizing the time of the trainer. It provides the opportunity to give a one-on-one training when necessary and provides the flexibility to switch to other students when required. On the other hand, a disadvantage of using a virtual learning environment is that it might not convey the real feeling of the machine that has to be interacted with. Only if the virtual representation of the machine is accurate and simulates every detail of the physical machine, the virtual training can prepare the trainee for interacting with a real physical machine.

**Different Physical Location - Virtual Object**

In this scenario of the design space, trainer and trainee do not have to be at the same location for conducting and participating in the training. This gives great flexibility to both parties involved in the learning experience. This opens up new possibilities for where learning takes place. For instance, the trainee can learn how to operate a machine while traveling just by connecting to a virtual learning environment and the virtual trainer. Further, this scenario allows to easily play back a previously recorded training as it needs to be recorded and transmitted to the virtual learning space in any case. The disadvantage of designing a learning system to be in a different physical location and using a virtual learning object, is that the actions of both trainer and trainee are transferred to a virtual scenario. Therefore, the trainer can only judge the trainees actions based on what is transmitted to the virtual scenario (this can happen on any granularity from translating movements and actions to an avatar to recording and transferring a full video of the users). On the other hand, the trainee can also only learn as much as it is being transmitted from the trainer, a stable transmission connection is necessary. Another negative aspect of this approach is that the connection between a trainee and a trainer might not be as personal as when meeting face-to-face. Previous research has shown that building up a personal connection to the trainer has a positive effect on the learning experience [11].

**Different Physical Location - Real Object**

In this scenario, the trainee and trainer are in different physical environments, however, one of them is with a real training object.

**Trainer is at real object**

First, we look at the case in which the trainer is at the location with the real object and the trainee is not on-site. One trainer could teach multiple trainees. For example, the trainer can talk about necessary steps to operate a machine while being directly at the machine. Further, the trainer is performing all necessary steps on-site, so it is less likely that the trainer will leave out necessary steps that are important as the work is performed directly at the physical machine. Depending on the technology that is used to transfer the learning scenario to a trainee, the trainee can interrupt the trainer when something is unclear and can point at the object in order to support a question and to receive a more detailed explanation.

**Trainee is at real object**

Secondly, we look at the other scenario of this design space dimension in which the trainee is at the location with the real object and the trainer is at a remote location (which is also the same situation as for remote expert support in running operations). In this scenario, the trainee has all freedom to interact with e.g. a real machine, which has benefits but also downsides. Most likely one trainer could teach only one trainee. The most convincing advantage of this scenario is that the trainee can experience the real object and does not have to interact with any form of simulation. The trainer can send comments and instructions to the learning site. However, in order to experience the full body feedback of the trainer, the trainer’s whole body has to be transferred virtually to the trainee’s site. In contrast, the downside of having a trainee on a real object is that especially when it comes to expensive machines, the machine cannot be used in production for the time the trainee is being taught. And also safety concerns might come up when an inexperienced trainee is operating a machine without a supervisor being physically present.

**SYSTEM**

We designed the HoloCollab system with two major goals in mind: First, providing scalable virtual assistance for trainees, and second, retaining a trainer component that can be either physically available or physically present with additional augmented capabilities (cf. beyond being there [13]). Considering the previously introduced design space, HoloCollab covers the dimension same location & real object and same location & virtual object.
A workflow in the HoloCollab system is a representation of work steps. Each workflow consists of \( n \) work steps, which have available tools, an end condition, and a visual feedback component. The available tools are virtual representations of tools that are spawned in front of the user (see Figure 4). E.g. a virtual screwdriver or a virtual wrench. These tools can be used for the current work step, or are distractor objects which have nothing to do with the current work step but are spawned to make the trainee put more effort into selecting the correct tool. An end condition is a logical condition that has to be fulfilled for the next work step to become active. An example of an end condition could be: using the virtual screwdriver at the correct position on the learning object. By having end conditions, the trainee can prove to the system that the current work step was understood and performed correctly. Finally, the visual feedback consists of a textual instruction (e.g. “use the screwdriver with the robot arm”) and a spatial annotation that can be directly placed on top of the object.

Feedback and Annotations

Once a workflow is loaded by a trainer, the trainer can decide if a virtual learning object or a physical learning object should be used for the training session. In case of a virtual representation, the learning object should be placed at any position in the room. By placing this virtual object, the system knows the exact position of the object and can place the feedback and annotations at the correct position. However, when a physical object is used as a learning object, the user has to additionally place a reference point at the bottom of the physical object in a setup phase. This is necessary to make the matching between the position of the feedback elements and the position of the physical object.

The previously mentioned feedback elements in HoloCollab are a textual feedback that explains the current work step, spatial target annotations, and 3D arrows indicating the direction. As depicted in Figure 4, the textual feedback is always presented in a textbox that moves with the user’s head movements. Thus, the trainee can always read the accompanying text first before asking a trainer for help. Considering the spatial annotations, in the HoloCollab system we are using a transparent green box that overlays either the physical or the virtual learning object to indicate an area at which an action has to be performed. Further, a 3D arrow is used to additionally highlight the position and direction of how the learning object should be approached by the trainee or the tool.

User Actions

The design of the system grants the trainer full control of a training session. Thus, the trainer can react to situations in which a trainee needs more or individual support. Actions, that the trainer can trigger, are activating and deactivating the feedback text, the spatial annotations, and the arrows, which point towards an area of action. As it is important that the user is always in full control of the system, we also made these actions available for the trainee. However, activating more feedback as a trainee without consulting a trainer first might result in a decrease of the learning experience as too much help makes solving a task very easy. Finally, in scenarios where the trainer is responsible for teaching multiple trainees, the trainer can easily jump between sessions and only see the feedback elements and virtual objects that belong to the selected student’s session. This setup allows for teaching of new skills and new workflows, but can also be used in an exam situation to check which tasks the trainee is already capable of performing without any help.

FIRST USER FEEDBACK AND LIMITATIONS

With our proof-of-concept implementation of the HoloCollab system, we covered the same location & virtual object and same location & physical object dimensions of the introduced design space. Initial tests showed that providing feedback that overlays the physical object provides a great alternative for both trainer and trainee to consume and present instructions. Also, having a virtual training object, which can be used to display feedback and annotations while still being able to talk about the feedback with a trainer in person, was perceived as very beneficial by both, trainers and trainees, during our initial tests. Having virtual objects enables each trainee to
have an own training object without being in somebody’s way and allows trainees to learn at their own pace. Also, the trainer enjoyed being able to switch between different training sessions to support multiple trainees at the same time.

We also discovered that the HoloCollab system comes with a few limitations. It felt unnatural for the trainees to not fully see the learning progress and learning environment of other trainees, as each learning environment is a separate virtual room. Further, the test users reported that using the Air Tap gesture of the Microsoft HoloLens is difficult to learn for them. Especially test users that were new to the device needed more training to handle the HMD correctly. We want to address this problem by also including support for more natural gestures in newer versions of HoloCollab. Finally, trainees reported that interacting with the virtual tools that are presented to them does not feel natural. Thus, the real task needs to be repeated with physical tools before finishing the learning of the task. However, for just learning the sequence of the task and which tools to use, HoloCollab has great potential.

In future work, we will extend our HoloCollab system to cover the other two dimensions of the design space: time and location. For the time design dimension, we want to be able to record full-body feedback of the trainer and postures and facial expressions that were made during the teaching. These could be played back by the trainee whenever the trainer feedback is needed and would not require a trainer to be present anymore. Considering the location design dimension, we can imagine to implement holographic representations of trainers and trainees (similar to the Holoporation project introduced by Orts-Escolano et al. [22]). We believe that this can further enhance the collaboration and the training outcome in a scenario where trainer and a trainee are physically separated, especially also for remote expert scenarios.

CONCLUSION

In this paper, we investigated the design space for a shared collaborative AR-based learning environment. We describe and explore the different design dimensions time, location, and materiality (real vs. virtual object) of the design space of creating AR-based training systems for industrial training scenarios. We elaborate on what systems following the different design dimensions can look like and explore their advantages and disadvantages. In our proof-of-concept implementation HoloCollab, we implement the design dimensions that utilize the same location for trainer and trainee using both virtual objects & real objects. In each instance we use a Microsoft HoloLens HMD for both trainer and trainee. The use of the HMDs for training scenarios extends the learning process with important feedback and empowers both the trainer and trainer with the opportunity to give annotations. Additionally, through network-connectivity, these training systems can access a huge database of instructions and can load the correct training scenario according to the current task that needs to be learned. We argue that the learning of new tasks, especially in industrial scenarios, can be improved using HoloCollab and that the presented design space will help to design and create new learner support systems in the future.

REFERENCES


http://dx.doi.org/10.1145/1395630.1395647

http://dx.doi.org/10.1145/3056540.3056547

http://dx.doi.org/10.1145/1785455.1785469

http://dx.doi.org/10.1145/2916767.2916779

http://dx.doi.org/10.1145/3056540.3076193


http://dx.doi.org/10.1145/3056540.3056548

9. Markus Funk, Sven Mayer, and Albrecht Schmidt. 2015. Using in-situ projection to support cognitively impaired


