Blending AR and VR to Increase Situational Awareness during Training

Austin Garcia, Eliot Winer
Iowa State University
Ames, Iowa
acgarcia@iastate.edu, ewiner@iastate.edu

ABSTRACT

Virtual training is estimated to make up $14 billion of the US military budget annually, with most going towards virtual reality (VR) and augmented reality (AR) applications. Advancements in commodity VR and AR head mounted displays (HMDs) have driven much of this recent growth by making virtual training solutions more practical and cost-effective. However, these training applications create a disparity in virtual environment perception between a trainer and a trainee. A trainee is in the HMD while a trainer is usually not, creating a barrier for collaborative communication. One solution is to display a trainee's viewpoint on a 2D display, but this makes it difficult for a trainer to have adequate situational awareness of a trainee. Another option is to place a trainer inside the virtual environment. This offers improved perception of a trainee's progress but removes a trainer's real-world situational awareness, which can be vital when facilitating training on complex processes or equipment. Neither of these options provide an adequate solution to bridge this visual communication gap.

User studies have shown that asymmetric collaboration between VR and AR users results in improved performance and engagement during collaborative tasks. These findings show the potential of applying asymmetric collaboration for commercial and military VR training. This paper introduces a prototype AR mobile application that establishes a new method for asymmetric collaboration within a training scenario. The approach uses body tracking to identify a trainee and align the VR scene around them in real-time. Using AR, a trainer can view a trainee in the real-world with augmented content around them replicating the virtual environment. Occlusion of the trainee's body accurately portrays their depth within the scene. Evaluation of this application found that the implementation of real-time body tracking and networking on a mobile device can maintain an operable framerate.

ABOUT THE AUTHORS

Austin Garcia is a master's student in Human Computer Interaction at Iowa State University. He is currently researching asymmetric collaboration between users in virtual and augmented reality. His previous research focused on virtual reality training for manufacturing, emergency scenario simulation, human-agent interaction, and online cooperation.

Eliot Winer, Ph.D., is the director of the Virtual Reality Applications Center and professor of Mechanical Engineering, Electrical and Computer Engineering, and Aerospace Engineering at Iowa State University. Dr. Winer has over 20 years of experience working in virtual reality and 3D computer graphics technologies on sponsored projects for the Department of Defense, Air Force Office of Scientific Research, Department of the Army, National Science Foundation, Department of Agriculture, Boeing, and John Deere.

Blending AR and VR to Increase Situational Awareness during Training

Austin Garcia, Eliot Winer
Iowa State University
Ames, Iowa
acgarcia@iastate.edu, ewiner@iastate.edu

INTRODUCTION

The U.S. training market was $92.3 billion in 2020-2021 (2021 Training Industry Report, 2021) with the Military Simulation and Training market predicted to grow from $9.2 billion in 2022 to $12.2 billion by 2027 (Military Simulation and Training Market, 2022). Recent advancements in augmented reality (AR) and virtual reality (VR) technology have opened new avenues for how training can be implemented. VR training with commodity head mounted displays (HMD) has seen widespread adoption in both military and commercial settings within the past decade. As a result, the amount of immersive training applications and academic work have been expanding in a variety of industries such as chemical processing (Garcia Fracaro et al., 2022), assembly and maintenance (Huang & Roscoe, 2021; Renganayagalu et al., 2021), aerospace (Renganayagalu et al., 2021), health care (Renganayagalu et al., 2021), and defense (Renganayagalu et al., 2021; Synthetic Training Environment, 2022). VR allows a user to immerse themselves into a virtual environment by obscuring the real world. Alternatively, AR augments the real world by placing information over top of a user's view either using see through displays or by capturing the real-world through a camera and displaying to a user in real-time (e.g., mobile device). Traditionally, VR required large and expensive systems such as cave automatic virtual environments (CAVE™). Newly developed HMDs targeted at consumers have lowered the entry cost for VR as well as proven more practical to use than CAVE™ systems. Additionally, AR and VR development tools have advanced alongside the hardware, streamlining the process of making immersive applications. Now, a simple VR setup only consists of an HMD and an open area for the VR user to move around. Expensive equipment, as well as dangerous scenarios, can be represented by 3D models and animations while training environments can be reset with a push of a button. These hardware and software advances have led militaries to adopt immersive training to prepare individuals for highly consequential tasks. Furthermore, the ability of VR and AR to immerse a user is extremely valuable when training for dynamic systems such as aircraft (Endsley, 1995). The U.S. military has seen the benefits immersive training offers and has already begun integrating it into their training programs. The U.S. Army's Synthetic Training Environment (STE) utilizes VR with plans to incorporate AR to help warfighters with general training, mission rehearsal, and more (Stone, 2021; Synthetic Training Environment (STE)). High demand means that military usage of VR and AR as training tools and supplements shows no sign of slowing down anytime soon (Liu et al., 2018).

One current issue, in both military and commercial settings, is that with VR HMD training is a very individual experience developed to benefit the trainee's experience at the expense of the trainer's experience. Training typically consists of not only the trainee, but a trainer who facilitates tasks such as hardware setup, assisting the trainee, providing feedback, or assessing the trainee's capabilities. Additionally, a trainer may also oversee multiple trainees at a time. Typically, a trainer conducting VR training can only see the virtual environment from the trainee's point of view through a 2D display. A 2D display provides some situational awareness (i.e., 3D environment projected onto a 2D display) but does not provide the trainer with a qualitative depiction of the virtual world. To increase situational awareness, a trainer can be placed into VR alongside the trainee (Grandi et al., 2019), but this nullifies the trainer's real-world situational awareness as well as adds additional complexity to the overall training process. A trainer may be required to exit VR to complete trainee evaluations, assist with specific training content, or solve technical issues. If the trainee is new to VR a trainer may be required to attend to issues like twisted cords, tracking space errors, or other novice VR user issues that require the trainer's real-world presence (Ouverson et al., 2020). Whether the trainer is watching a 2D display or in VR, the aforementioned issues can create a barrier for collaborative communication.

One promising approach to help bridge the existing communication gap in VR training is asymmetric collaboration. Asymmetric collaboration occurs when users employ devices with varying levels of immersiveness to access a shared virtual space and cooperate within that environment. Higher levels of engagement (Gugenheimer et al., 2017; Renganayagalu et al. 2021), presence (Chan & Minamizawa, 2017; Gugenheimer et al., 2017), enjoyment

(Gugenheimer et al., 2017), and quickened task durations (Cavallo et al., 2019; Grandi et al., 2019) have been reported when applying asymmetric collaboration to environments implementing VR and AR devices. Since training often involves multiple users with varying objectives (a trainer and trainee), asymmetric collaboration naturally fits into the training structure to better suit the objective required by the respective user. An immersive learning environment with clear guidance is desired for the trainee. By placing the trainee in VR, these goals can be accomplished. In contrast, simultaneous situational awareness of the real and virtual world as well as indication of the trainee's progress is desired for the trainer. The current options lack the ability to efficiently deliver these objectives to a trainer. Recent advances in mobile devices and their widespread usage have made the prospect of using AR to implement asymmetric collaboration with a VR user an attainable target.

This paper introduces a prototype system that implements a new method of asymmetric collaboration for virtual training between a VR HMD and an AR mobile device. Utilizing 3D mobile body tracking to identify the location of a VR trainee, this prototype provides an immersive AR view of the current state of the virtual training environment to the trainer. The system provides a method of bridging the existing communication gap between a trainer and trainee. In developing this new asymmetric collaboration method, it was found that mobile AR has the potential to serve as a useful supplement for VR training applications, as well as identified the limitations that need to be addressed before widespread implementation can happen.

BACKGROUND

Usage of VR Training in Military Environments

The ability to place a trainee in nearly any environment to replicate expensive or dangerous tasks has drawn the military to VR training. Tasks such as pilot training benefit immensely from increased presence throughout the training process (Endsley, 1995). However, increased presence isn't the only reason for military adoption of VR training. Lele (Lele, 2011) initially found the following five reasons for the increasing usage of VR by militaries: 1) leaps in the fidelity of immersive technology can allow military uses for more than just training, 2) modern military challenges require innovative solutions capable of adapting to the scenario, 3) virtual environments provide safe settings that allow mistakes without real world consequences, 4) VR's affordability, and 5) VR can focus content so a user only sees what is vital to the current situation. Although these reasons were postulated in 2011 they still hold true today. Liu (Liu et al., 2018) updated Lele's work and found VR usage aligns with existing demand with the result that military VR usage will become more widespread.

The U.S. Army has already committed resources to enable the use of VR and AR HMDs in multiple roles. Work with Microsoft has yielded the creation of the Integrated Visual Augmentation Systems Platform (Goldstein, 2021) that is vital to their STE program and is planned to incorporate AR training alongside the existing VR implementation (Stone, 2021; Synthetic Training Environment (STE)). The Army has already committed large amounts of resources after seeing the benefits offered by VR and AR training. Training applications such as the one this paper discusses are already in use in military environments.

Effectiveness of VR Training

VR training has become more widespread not only due to the influx of VR HMD's availability commercially, but because of the flexibility and effectiveness of VR training. As a result, the amount of VR/AR training applications and academic work has been expanding in a variety of industries. Fracaro (Garcia Fracaro et al., 2022) found that the chemical processing industry has seen a major increase in the last decade of publications that developed an immersive training application or conducted research with one. Renganayagalu (Renganayagalu et al., 2021) conducted a systematic review that found VR training was currently beneficial to at least five different industries. VR's training efficiency combined with reasonable costs have helped boost this recent growth. VR training has been found to be at least as effective as traditional training, while proper use cases and higher fidelity devices could lead to performance improvements (Kaplan et al., 2021). For example, Lohre (Lohre et al., 2020) ran a study that compared traditionally trained surgeons with immersive VR trained surgeons and found that the latter group was able to complete complicated tasks both faster and with a higher knowledge retention rate. A similar study resulting in users demonstrating fewer errors and quicker completion times than a traditionally trained group can be seen in Seymour (Seymour et al., 2002). These are just a few examples that demonstrate the training effectiveness of VR. In contrast, Renganayagalu's

(Renganayagalu et al., 2021) review of VR HMD professional training research found that due to the differing levels of quality of the reviewed literature, no fundamental inferences could be made about the overall effectiveness of VR training. However, Renganayagalu (Renganayagalu et al., 2021) did find that given the proper use case, high fidelity hardware, and thorough analysis, VR training could benefit trainee confidence, skill retention, performance, engagement, motivation, and increase the appeal of the subject. Compared to traditional training, VR training has demonstrated it is capable of quick, efficient training given the appropriate scenario and fidelity.

Effectiveness of Asymmetric Collaboration between VR and AR

Previous research into asymmetric collaboration with immersive technology has found promising trends. Ouverson's (Ouverson et al., 2020) scoping literature review found that research on asymmetric collaboration produced results that aligned well with the structure of virtual training. Previous research has shown that asymmetric collaboration for virtual training offers more than just theoretical benefits. Gugenheimer (Gugenheimer et al., 2017) created an application that employed asymmetric collaboration that allowed a non-HMD user to interact with an HMD user. They found that it improved presence, enjoyment, and engagement while collaborating when compared to a base condition using a 2D screen. Additionally, Gugenheimer (Gugenheimer et al., 2018) found that a non-HMD user was able to interact with an HMD user, but the HMD user held more responsibility and dominance over the non-HMD user. This is not an issue when considered in the training dynamic. The trainer's dominant role aligns with the dominance exhibited by the non-HMD user. Another possible benefit of asymmetric collaboration is that varying levels of immersion can shorten task durations when compared to symmetric VR, AR, or traditional 2D cases (Cavallo et al., 2019). In contrast, Grandi (Grandi et al., 2019) found that a symmetric VR-VR condition performed better and faster than an asymmetric VR-AR condition, which in turn performed better and faster than a symmetric AR-AR condition. However, the author's application is not attempting to get multiple people to complete a task the fastest, but rather to assist the trainer while facilitating the training process. Speed of training is a secondary consideration. Additionally, the VR-VR case removes real world situational awareness that is mandatory to the trainer, which is a dynamic not present in Grandi's work. Thus Grandi (Grandi et al., 2019) demonstrates the VR-AR case does perform well as a collaborative platform. Furthermore, Minamizawa (Chan & Minamizawa, 2017) provides an example of asymmetric collaboration increasing the presence of a non-HMD user.

Galati (Galati et al., 2020) utilized Microsoft's Mixed Reality Spectator View for the Microsoft Hololens (AR HMD) as a solution for asymmetric collaboration. Spectator View provides a similar experience to the prototype introduced in this paper. Both methods allow the use of a mobile AR device to view the virtual environment of an HMD user. Spectator View provides both marker-based and markerless options whereas the author's approach is markerless. The method proposed by this paper relies solely on body tracking to properly align the AR scene. A markerless solution means less setup is required by trainers and the trainees don't need to do any extra initialization. Spectator View's markerless solution utilizes Microsoft Azure spatial anchors in its implementation. More research will need to be done to test the viability of spatial anchors for training scenarios. Another difference between the proposed method's approach and Spectator View is that Spectator View only supports Hololens devices while the proposed method currently supports VR HMDs and has the possibility of expanding to support any VR/AR HMD.

METHODS

Prototype

The prototype introduced by this paper intends to solve the existing communication gap between a VR trainee and trainer by providing an AR experience for the trainer. This AR experience provides the trainer a real-time view of the virtual environment overlaid onto the real world. This is accomplished by utilizing mobile body tracking to locate a VR trainee and place the virtual environment around them. The trainer can move around the area with a mobile device and see an accurate representation of the virtual environment in 3D space as well as the VR user's current position within it. This prototype intends to create a middle ground between full immersion and real-world awareness. This is accomplished by providing the trainer with an immersive view that allows greater awareness of the virtual environment as well as real-world awareness to complete training facilitation tasks.

Hardware and Software

Development of this application extended the functionality of a preexisting VR trainer created by Dodoo (Dodoo et al., 2018) for the Oculus Rift CV1, where a user was guided in VR through a mock wing assembly. In addition to updating the VR application for newer VR HMDs using UnityXR, a separate AR version of the application was developed. The AR application used body tracking from Apple's ARKit 3 which required an Apple device with an A12 Bionic Chip and at least iOS 13. The final VR application was tested on an Oculus Quest 2 with a link cable, while the final AR application was tested on an iPad Pro 11-inch (2nd generation, iOS 14). Evaluation of the application required locating the iPad in relation to the VR HMD, so a Vive Tracker 3.0 was attached to the iPad. To get precise tracking data, evaluation was completed on a Valve Index with Vive Pro 2 Controllers utilizing the tracking information provided by the SteamVR Base Station 2.0's rather than the inside-out tracking of the Oculus Quest 2.

Multi-User Networking

The original training application only allowed for use with a single VR HMD user. In order to enable the AR trainer to see into the VR trainee's scene in real time, a multi-user network was set up using Photon Unity Networking (PUN). The VR application stores the current training progress state and the user position to a multiplayer server. This allows the AR application to connect to the VR user's session at any point and access the latest information about the VR scene's current progress. The multi-user networking also allows for VR-VR and VR-AR image tracking multi-user interactions that are implemented in the application, but this paper will focus on the VR-AR body tracking.

Figure 1. A view of the original VR scene.

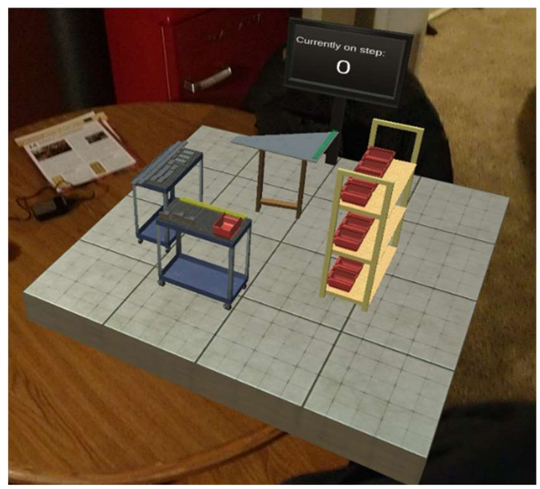

Figure 2. The AR version of the assembly scene simplified for use on a mobile device.

AR Scene Simplification

One limitation of the AR application is that mobile devices are not as powerful as VR capable machines. The original scene used in the VR application incorporated extra content to contribute to the immersiveness of the virtual environment, as seen in Figure 1. Extra geometry not vital to the trainer both obscured vital content and had a negative impact on performance on a mobile application. The AR scene did not require the full VR scene to be displayed. As a result, a digital twin of the VR scene for mobile AR was created that was simplified down to only encompass the necessary components of the assembly. Components added solely to increase immersion for the VR trainee (i.e., warehouse environment and additional workstations) were removed. Additionally, complex geometry such as the screws and washers filling the red bins seen in Figure 2 were removed to keep the vertex count down. While complex geometry can be displayed on mobile devices, the author's focus was on minimizing stress on the device to enable the best conditions for tracking. The changes to the digital twin still allow the AR trainer to maintain awareness of the vital portions of the assembly while balancing performance for mobile devices.

Body Tracking

With the AR scene completed, a method to position and rotate the digital models, in real-time, relative to the user was needed. A tracked real-world location is required by the AR application to dictate where the AR digital twin is placed. Image and object tracking can allow for identification of a real-world location. However, this requires setting up a physical marker such as a QR code or a specified 3D object. While AR applications can identify real world markers, as used in Spectator View (Galati et al., 2020), not all VR HMDs are able to track markers so calibration between the AR and VR devices would have to be completed manually. The VR scene allows constant access to information about where the VR HMD and tracked hand controllers are in the virtual environment. However, using this information combined with an AR marker placed on the VR user doesn't work in practice due to the constant movement and rotation of the VR user. Constant reorientation of a marker makes it difficult for an AR application to maintain tracking. Instead of trying to track a marker, body tracking allows the tracking of the actual VR user no matter their orientation. The body tracking functionality provided in ARKit 3 provides markerless access to the location of the VR trainee in real world 3D space via a tracked skeleton that is generated when the AR camera is directed at a person. Reimer (Reimer et al., 2021) found that while ARKit body tracking did not provide a precise range of motion values, it was able to reliably track motion. Aligning the HMD location of the VR trainee and the tracked skeleton's head provides easy positional alignment of the scene in AR around the VR user. Head tracking provided by ARKit 3 offers accurate positional data, but there is no real-world point of reference to sufficiently determine the rotation values to accurately set the AR digital twin rotation.

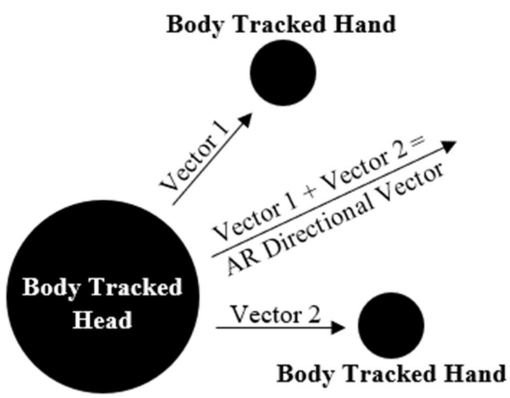

Figure 3. A diagram of how the AR directional vector is calculated from the body tracking data.

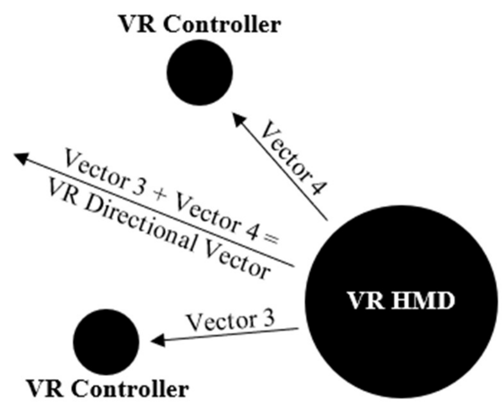

Figure 4. A diagram of how the VR directional vector is calculated from the VR HMD data.

To obtain accurate rotational data for scene alignment two vectors are calculated, one between the body tracked head and left hand and the other between the body tracked head and right hand. These vectors are added together to get a single directional vector (see Figure 3). This process is repeated for the VR head and hands (see Figure 4). The VR and AR scene directional vectors are compared, and the difference found is used to rotate the AR digital twin around the AR body tracked head location's up axis. The full sequence can be seen in Equations 1-8. This process is done over a set of initialization frames before the median rotations and mean positions are locked in. If the rotation calculation process is continually run then the resulting experience is unstable, finding the median after an initialization period provides a smoother overall experience.

$$V_{AR} = AR\ Directional\ Vector \tag{1}$$

$$V_{VR} = VR\ Directional\ Vector \tag{2}$$

$$V_{up} = Up\ Vector\ or\ (0, 1, 0) \tag{3}$$

$$Q_{AR} = A\ quaternion\ with\ V_{AR}\ forward\ direction\ and\ V_{up}\ up\ direction \tag{4}$$

$$Q_{VR} = A\ quaternion\ with\ V_{VR}\ forward\ direction\ and\ V_{up}\ up\ direction \tag{5}$$

$$Y_{difference} = Euler\ Angle\ Y(Q_{AR}) - Euler\ Angle\ Y(Q_{VR}) \tag{6}$$

$$Q_{Scene\ Rotation} = A\ quaternion\ with\ rotation\ Y_{difference}\ around\ V_{Scene\ Up} \tag{7}$$

Body Occlusion

After the AR digital twin's position and rotation were set, the AR scene obscured most of the VR trainee's body. To overcome this limitation an avatar was created that encompasses the ARKIT 3 skeleton and thus the outline of the VR trainee in the AR view. Since ARKit 3 body tracking provides 3D positioning information, the depth of the VR user in the scene can be utilized to accurately place them within the virtual environment. A shader was applied to the avatar that hides all geometry behind the avatar while still allowing geometry in front of the avatar to be visible. This creates a see-through effect that allows the mobile device's camera view of the VR trainee to show through where the AR avatar is, as seen in Figure 5.

Figure 5. A view of the final AR application showcasing both the AR digital twin and body occlusion. The trainee is wearing a VR Oculus Quest 2 HMD

RESULTS & DISCUSSION

Application Performance

An evaluation of the performance of the application found that it runs at a consistent frame rate of 120 frames per second on a 2nd generation iPad Pro. The AR application displayed a maximum of 923,050 vertices while viewing the digital twin in real-time, compared to the VR application which displayed a maximum of 18,305,338 vertices while viewing the VR scene in real-time.

Despite smooth framerates, the body tracking experiences some stability issues. These stability issues become more pronounced when one of the VR user's arms is blocked by their body, which impacts ARKit 3's body tracking algorithm. The scene alignment algorithm is impacted when one of the VR user's hands is obstructed from the camera view. Additionally, the movement of networked objects in the scene is affected by latency so an internet connection capable of supporting online gaming (at least 10mb per second) is advised.

In order to assess the accuracy of the final result of the scene alignment algorithm, a version of the application was set up on a Valve Index with a Vive Tracker attached to an iPad Pro. This allowed the capture of the AR user's position within the context of the VR scene. The AR scene alignment values were then compared against the expected alignment values given the VR scene data with the tracked iPad Pro location. Two poses were used during evaluation, a still pose and a moving pose. The still pose required the VR user to hold their arms up perpendicular to their body and a shoulders width apart, as seen in Figure 6. This pose was chosen for ideal initialization as it creates the largest VR and AR directional vector magnitudes as the hands are close together and as far from the head position as possible. The moving pose involved the VR user going through the assembly training actions. The two-user poses, along with two initialization periods, combine to create four different conditions for evaluation. Five trials were conducted for each of the conditions. The two initialization periods were the first 600 and 1200 frames. The same users posed for all trials in VR and AR and stood far enough from each other to see their entire body in the AR view but remained within the VR tracked space (approximately 5-10 feet from the VR user position). The AR user began in random positions in front of the VR user and reoriented their position throughout each trial to retain a view of the front of the VR user as much as possible to produce more accurate body tracking results. The longer initialization period and still pose conditions were expected to produce better results. The 600-frame still pose condition experienced an average position error of 0.600 meters and rotation error of 37.921 degrees. The 600-frame moving pose condition experienced an average position error of 0.802 meters and rotation error of 7.193 degrees. The 1200-frame still pose condition experienced an average position error of 0.632 meters and rotation error of 7.627 degrees. The 1200-frame moving pose condition experienced an average position error of 1.168 meters and rotation error of 53.644 degrees. The full results can be seen in Tables 1 and 2.

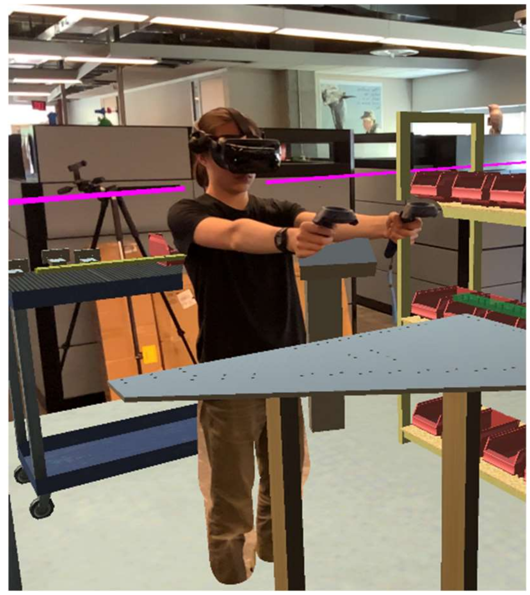

Figure 6. The ideal initialization pose used during the still pose trials.

Neither the final rotation or position final values conclusively show that a certain pose or initialization period produces better results. For example, the best two rotation conditions were the opposite condition pairs (600/moving and 1200/still). The inconsistent nature of these results exhibits the current instability of ARKit's body tracking whilst also showing its ability produce accurate results given the right conditions. While the current results demonstrate too much error for practical use, they also show feasibility of this approach. The randomization of the AR user's position during the trials likely picked up on certain tracking angles that produced better results and impacted the overall results. Fine tuning the position and rotation calculations to remove outliers could provide more consistency between different tracking angles. A set of trials with a consistent AR user position could show better separation between the different conditions but would not be representative of an actual AR user's experience.

Table 1. Rotation Evaluation Results

	Rotation Error (in degrees)			
Initialization Period/Pose	**600/Still**	**600/Moving**	**1200/Still**	**1200/Moving**
Trial Results	38.1166	11.59315	8.056954	61.37426
	43.97235	6.436779	8.275978	60.9529
	39.13592	7.12598	7.561094	61.22197
	59.05196	5.591099	7.62464	61.69018
	9.329796	5.216011	6.614883	22.97977
Average Result	37.9213252	7.1926038	7.6267098	53.643816

Table 2. Position Evaluation Results

	Position Error (in meters)			
Initialization Period/Pose	**600/Still**	**600/Moving**	**1200/Still**	**1200/Moving**
Trial Results	0.552908	0.690155	0.565561	1.948599
	0.841839	0.696322	0.564943	1.24569
	0.404127	0.868378	0.520061	1.022677
	0.734682	0.867478	0.683977	1.188664
	0.468203	0.885185	0.824315	0.434928
Average Result	0.6003518	0.8015036	0.6317714	1.1681116

Mobile AR Viability

The main challenge with scene alignment was determining the scene's initial orientation. After initialization, ARKit's tracking was able to keep the scene stable enough for a viewing experience with less than 1.2 meters of positional error and less than 60 degrees of rotational error. These values can improve to less than .7 meters of positional error and less than 10 degrees of rotational error given the best conditions. Thus, the overall precision of the application mainly depends on the outcome of the initialization process. Imprecise positioning and latency meant that dynamic components and small-scale interactions (such as a hand pressing a button or picking up a small bolt) were visibly off by the error values discussed above. Actions that involved changing positions in the scene and interaction with larger components provided a much better viewing experience within the AR application.

In the event that someone other than the VR user gets picked up by ARKit's body during initialization, it is possible that the scene orientation or body occlusion could be applied to the wrong person. VR occludes the trainee's view, so most VR setups already require an open space free of objects and people to avoid collisions, so this is not a major issue.

Asymmetric Collaboration Viability

Preliminary feedback suggests that with some adjustments the introduced prototype could be implemented into future VR training scenarios with success. A trainer can view content all around the trainee, including content behind them. This provides the trainer with information lacking from a 2D display showing the trainee's point of view. Since there are no extra trackers for trainers to set up or complicated calibration processes, the VR trainee need only worry about the VR training material and the trainer can simply access the AR supplement whenever desired. The process of grabbing the iPad, opening the application, and pointing it at the VR trainee is quick and intuitive. This allows the trainer to pick up the AR application and start tracking at any point throughout the training without requiring an excessive initialization sequence. Accurate initialization of the scene alignment does require the camera to have a view of both trainee's hands so communicating with the VR user to stand still for a couple of seconds while the AR application gets the correct orientation is the most arduous portion of the setup. This could be improved in future iterations by making the orientation time quicker or utilizing improved body tracking systems.

One drawback for asymmetric collaboration with the proposed prototype is the reliance on verbal communication. Verbal communication is the only current method for the trainer to interact with the trainee, so instructions from the trainer must include orienting information as well since there is no visual component. For example, the AR application allows the trainer to the next assembly piece located behind the trainee, but they must verbally tell the VR trainee that the next part is to their back left. The more intricate the location, the harder it is for the trainer to convey 3D orientation information. Additionally, if audio is a vital component of a VR training application, then trainer-trainee communication will be challenging. Possible solutions to this could involve giving the trainer a microphone to speak directly to the trainee or incorporate a visual system that allows the trainer to place points of interest into the virtual environment for the trainee. This could differ depending on the training topic. A commercial assembly trainer may want to be able to tell a trainee to turn around and walk five feet or a military fighter jet trainer may just want to make a specific button glow on a control panel. Currently information is only sent from the trainee to the trainer about the virtual environment's current state but enabling communication from the trainer back to the trainee would not be difficult and is a logical next step.

Confirming that tracking is oriented correctly is another aspect that relies on verbal communication. After orientation has completed, the AR user has no way of knowing the accuracy of the scene alignment without asking the VR user what direction they are facing or waiting for the user to interact with the virtual environment and see how far off the user is from the action happening in the scene. To avoid verbal communication the trainer can also look at the desktop point of view of the trainee and try to determine if orientation is correct, but this removes the benefits that are offered by the mobile AR method until the orientation is confirmed to be correct. Outside of observing how far off interactions in the scene are, none of these options provide a precise method of confirming that scene alignment has been initialized successfully.

CONCLUSION

The current state of the prototype shows that current commodity hardware is capable of supporting the asymmetric collaboration method introduced in this paper. By utilizing AR combined with body tracking on a mobile device, this prototype was able to successfully identify a VR user and display a digital twin of the virtual environment around them. This enables the trainer to see more information about the virtual environment than what is provided from the VR trainee's point of view. Higher-level tasks such as movement between different stations and placing large components are easy for a trainer to follow along with. Additionally, training that focuses on sequences and larger actions currently fit in the capable realm of this prototype. Military and commercial training programs could use this method to bolster existing and future VR training applications by providing trainers an enhanced experience.

The current limiting factors of this prototype are ARKit body tracking and the requirement to create a digital twin of the virtual environment capable of running on a mobile device. Cutting content out of the AR digital twin can only be completed on an individual application basis depending on what is vital to the training material. Future improvements to mobile body tracking would greatly improve the stability of the application and the feasibility of using this method of asymmetric collaboration in military and commercial training environments.

Future Work

The authors plan to refine the AR scene alignment system as well as add smoothing to ARKit 3's body tracked skeleton to increase overall stability of the AR experience. More work will need to be done to improve precision to enable smaller scale training interactions to be viewable. ARKit 3 also contains a built-in body occlusion function that would provide better fidelity if it was able to work simultaneously in a body tracked scene. Additionally, the Valve Index version of the application that involved attaching a Vive Tracker to the iPad Pro could be used to create a version of the application that relies more on the networking data but produced more accurate tracking results, this approach was not discussed as this paper intended to study the viability of mobile body tracking for VR training scenarios with as minimal user setup as possible. An intriguing possibility involves replacing the trainer's iPad Pro with an AR HMD such as a Microsoft Hololens that would provide both 3D perspective of the VR user's scene as well as a hands-free experience. The ability for the trainer to send information back to the trainee as discussed previously is another proposed goal.

The ultimate culmination of this work would involve running a user study that compares the differing levels of immersion a trainer can experience. These levels would consist of a traditional desktop view, mobile AR as introduced in this paper, HMD AR, and VR. This would allow insight into which level of immersion provides the most benefit to a trainer.

REFERENCES

2021 Training Industry Report (2021). Retrieved June 2021 from https://trainingmag.com/2021-training-industry-report/

Cavallo, M., Dholakia, M., Havlena, M., Ocheltree, K., & Podlaseck, M. (2019). Immersive insights: A hybrid analytics system for collaborative exploratory data analysis. In Proceedings of the ACM Symposium on Virtual Reality Software and Technology, VRST. Association for Computing Machinery. https://doi.org/10.1145/3359996.3364242

Chan, L., & Minamizawa, K. (2017). FrontFace: Facilitating communication between HMD users and outsiders using front-facing-screen HMDs. In Proceedings of the 19th International Conference on Human-Computer Interaction with Mobile Devices and Services, MobileHCI 2017. Association for Computing Machinery, Inc. https://doi.org/10.1145/3098279.3098548

Dodoo, E. R., Hill, B., Garcia, A., Kohl, A., MacAllister, A., Schlueter, J., & Winer, E. (2018). Evaluating commodity hardware and software for virtual reality assembly training. In IS and T International Symposium on Electronic Imaging Science and Technology. Society for Imaging Science and Technology. https://doi.org/10.2352/ISSN.2470-1173.2018.03.ERVR-468

Endsley, M. R. (2017). Toward a theory of situation awareness in dynamic systems. In Human Error in Aviation (pp. 217–249). CRC Press. https://doi.org/10.4324/9781315092898-13

Galati, R., Simone, M., Barile, G., De Luca, R., Cartanese, C., Grassi, G. (2020). Experimental setup employed in the operating room based on virtual and mixed reality: Analysis of pros and cons in open abdomen surgery. Journal of Healthcare Engineering, 2020, 1–11. https://doi.org/10.1155/2020/8851964

Garcia Fracaro, S., Glassey, J., Bernaerts, K., & Wilk, M. (2022, April 1). Immersive technologies for the training of operators in the process industry: A Systematic Literature Review. Computers and Chemical Engineering. Elsevier Ltd. https://doi.org/10.1016/j.compchemeng.2022.107691

Goldstein, P. (2021, January 12). The Army uses AR to make training more dynamic. Technology Solutions That Drive Government. Retrieved May 9, 2022, from https://fedtechmagazine.com/article/2021/01/army-uses-ar-make-training-more-dynamic

Grandi, J. G., Debarba, H. G., & Maciel, A. (2019). Characterizing asymmetric collaborative interactions in virtual and augmented realities. In 26th IEEE Conference on Virtual Reality and 3D User Interfaces, VR 2019 - Proceedings (pp. 127–135). Institute of Electrical and Electronics Engineers Inc. https://doi.org/10.1109/VR.2019.8798080

Gugenheimer, J., Stemasov, E., Frommel, J., & Rukzio, E. (2017). ShareVR: Enabling co-located experiences for virtual reality between HMD and Non-HMD users. In Conference on Human Factors in Computing Systems - Proceedings (Vol. 2017-May, pp. 4021–4033). Association for Computing Machinery. https://doi.org/10.1145/3025453.3025683

Gugenheimer, J., Stemasov, E., Sareen, H., & Rukzio, E. (2018). FaceDisplay: Towards asymmetric multi-user interaction for nomadic virtual reality. In Conference on Human Factors in Computing Systems - Proceedings (Vol. 2018-April). Association for Computing Machinery. https://doi.org/10.1145/3173574.3173628

Hoover, M., Miller, J., Gilbert, S., & Winer, E. (2020). Measuring the performance impact of using the Microsoft hololens 1 to provide guided assembly work instructions. Journal of Computing and Information Science in Engineering, 20(6). https://doi.org/10.1115/1.4046006

Huang, W., & Roscoe, R. D. (2021, September 1). Head-mounted display-based virtual reality systems in engineering education: A review of recent research. Computer Applications in Engineering Education. John Wiley and Sons Inc. https://doi.org/10.1002/cae.22393

Kaplan, A. D., Cruit, J., Endsley, M., Beers, S. M., Sawyer, B. D., & Hancock, P. A. (2021). The effects of virtual reality, augmented reality, and mixed reality as training enhancement methods: A meta-analysis. Human Factors: The Journal of the Human Factors and Ergonomics Society, 63(4), 706–726. https://doi.org/10.1177/0018720820904229

Lele, A. (2013). Virtual reality and its military utility. Journal of Ambient Intelligence and Humanized Computing, 4(1), 17–26. https://doi.org/10.1007/s12652-011-0052-4

Liu, X., Zhang, J., Hou, G., & Wang, Z. (2018). Virtual Reality and Its Application in Military. In IOP Conference Series: Earth and Environmental Science (Vol. 170). Institute of Physics Publishing. https://doi.org/10.1088/1755-1315/170/3/032155

Lohre, R., Bois, A. J., Athwal, G. S., & Goel, D. P. (2020). Improved complex skill acquisition by immersive virtual reality training. Journal of Bone and Joint Surgery, 102(6). https://doi.org/10.2106/jbjs.19.00982

Military Simulation and Training Market. (2022). Retrieved from June 30, 2022 from https://www.marketsandmarkets.com/Market-Reports/military-simulation-virtual-training-market-661.html

Miller, J., Hoover, M., & Winer, E. (2020). Mitigation of the Microsoft hololens' hardware limitations for a controlled product assembly process. The International Journal of Advanced Manufacturing Technology, 109(5-6), 1741–1754. https://doi.org/10.1007/s00170-020-05768-y

Ouverson, K. M., Hoover, M., Gilbert, S. B., Dorneich, M. C., & Winer, E. (2020). Situational Awareness Methods in Virtual Reality Training: A Scoping Review. Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC), (20336), 1–13.

Reimer, L. M., Weigel, S., Ehrenstorfer, F., Adikari, M., Birkle, W., & Jonas, S. (2021). Mobile Motion Tracking for Disease Prevention and Rehabilitation sing Apple ARKit. In Studies in Health Technology and Informatics (Vol. 279, pp. 78–86). IOS Press BV. https://doi.org/10.3233/SHTI210092

Renganayagalu, S. kumar, Mallam, S. C., & Nazir, S. (2021). Effectiveness of VR head mounted displays in professional training: A systematic review. Technology, Knowledge and Learning, 26(4), 999–1041. https://doi.org/10.1007/s10758-020-09489-9

Seymour, N. E., Gallagher, A. G., Roman, S. A., O'Brien, M. K., Bansal, V. K., Andersen, D. K., … Blumgart, L. H. (2002). Virtual reality training improves operating room performance results of a randomized, double-blinded study. In Annals of Surgery (Vol. 236, pp. 458–464). https://doi.org/10.1097/00000658-200210000-00008

Stone, A. (2021, October 7). US Army makes headway on Synthetic Training Environment. Defense News. Retrieved May 9, 2022, from https://www.defensenews.com/training-sim/2021/09/30/us-army-makes-headway-on-synthetic-training-environment/#:~:text=Known%20as%20the%20Squad%20Immersive,soldiers'%20view%20of%20their%20surroundings.&text=The%20SIVT%20training%20software%20is,the%20Integrated%20Visual%20Augmentation%20System.

Synthetic Training Environment. USAASC. (2022). Retrieved May 9, 2022, from https://asc.army.mil/web/portfolio-item/synthetic-training-environment-ste/