Mixed and Augmented Reality for Marine Corps Training

Richard Schaffer, Sean Cullen, Phe Meas, and Kevin Dill

Lockheed Martin Mission Systems & Training, Burlington, MA, USA
{richard.l.schaffer,sean.cullen,phe.meas,kevin.dill}@lmco.com

Abstract. The United States Marine Corps faces numerous challenges in preparing Marines for current operations; among them are the cost of specialized training environments and the difficulty of realistically representing the deployed environment. This paper reports on two Office of Naval Research efforts to address these challenges. The first employs Mixed Reality, which combines real-world and virtual elements to create a Hollywood-set-like representation of an Afghan village where Marines can train prior to deployment. The second explores the use of Augmented Reality to train USMC observers. Observers are responsible for directing artillery and mortar fires and aircraft attacks in the proximity of friendly forces. While the live environment has numerous advantages, the costs of supporting troops, ammunition, and equipment are considerable. Augmented Reality can replace live supporting forces, resulting in lower cost use of training areas during down time and enabling almost any area to become an augmented training area.

Keywords: Mixed Reality, Augmented Reality.

1 Introduction

Training our military forces is challenging. Effective training requires the construction of specialized environments, expenditure of ammunition, use of high-cost equipment such as aircraft or artillery, and supporting personnel (including military personnel, civilian role-players, and staff). In addition, it can be difficult to accurately portray conditions in the deployed environment, and even more difficult to rapidly convert training environments from one remote location to another.

In this paper we report on two Office of Naval Research (ONR) efforts that explore techniques for combining live training with virtual elements in order to create more realistic, more immersive training environments at lower cost and with decreased requirements in terms of materiel and manpower.

1.1 Mixed Reality at the Infantry Immersion Trainer

Since its inception in 2007 at US Marine Corps Camp Pendleton, California, the goal of the Infantry Immersion Trainer (IIT) has been to recreate as accurately as possible

R. Shumaker (Ed.): VAMR/HCII 2013, Part II, LNCS 8022, pp. 310–319, 2013.

a small village similar to what the Marines could expect to face in deployed environments, first in Iraq and more recently in Afghanistan. This includes not only the physical features – the sights, the sounds, the smells – but also the socio-cultural factors – the language, the culture, and the behaviors of the locals [1]. The latter is typically achieved through the use of civilian role-players who are first-generation Afghan immigrants, and who play roles such as the village elder, the fertilizer salesman, the insurgents, and so on. Immersive training has become increasingly popular in recent years due to its effectiveness in preparing Marines for contemporary combat operations. Since 2007, ONR has continued to enhance the system and IIT facilities have been constructed at two additional USMC sites, although the Camp Pendleton IIT remains the site for ONR experimentation.

The creation of an immersive training experience requires careful construction of the physical environment and selection of appropriate role-players, but also includes a variety of Mixed Reality (MR) technologies, where an MR technology is anything that combines real-world and virtual elements into a single system. Among the MR elements at the IIT are scent generators, speakers that provide ambient sounds, Hollywood-quality pyrotechnics for simulating IEDs and rocket-launched grenades, and digital avatars. In this paper we will focus on the digital avatars, which are virtual characters who are projected onto the wall using a standard video projector.

1.2 Augmented Immersive Team Training

Augmented Reality (AR) uses a head-mounted display (HMD) to overlay virtual characters onto the physical world. For example, the HMD might show the physical world but insert an aircraft, or an insurgent, or an explosion. The Augmented Immersive Team Training (AITT) program is a five-year program, currently in its third year, which is examining the viability of using AR for training USMC observers. Observers are responsible for directing artillery and mortar fires, and for controlling aircraft attacks, in the proximity of friendly forces. While the live environment has numerous advantages, the costs of supporting troops, ammunition, artillery, and aircraft are considerable, and training areas are not always available. We anticipate that the use of AR-generated forces will allow the Marine Corps to decrease costs and to use almost any area as an augmented training area when the Marines are deployed or physical training areas are unavailable.

2 Digital Avatars at the IIT

While the role-players do an excellent job of portraying culturally authentic characters, they require expenditures of thousands of dollars each day. In addition, there are certain roles that can't easily be portrayed with role-players. Appropriate personnel may simply not be available (as is often the case with the elderly), or there may be legal and moral prohibitions against using them (as is the case with children or badly wounded characters). Thus the digital avatars are not intended to replace the role-players, but rather to supplement them with additional characters in specific roles [2].

Fig. 1. Digital avatars of a father and son (left) and insurgent and hostage (right)

Figure 1 shows two examples of digital avatars. The image on the left is a screenshot of the virtual characters, while that on the right shows characters in the context of the surrounding room.

Originally, the digital avatars at the IIT were used for highly scripted "shoot-or-no-shoot scenarios" in which the Marines would see the avatars upon entering a room and immediately decide whether or not to open fire. These scenarios typically lasted only a few seconds, so there was limited need for complex decision making; with the exception of detecting hits from hostile fire, high-level decisions could be handled by an instructor/operator who was controlling the character from behind the scenes. Our more recent work strives to make the avatars capable of a broader, more expressive set of behaviors. Consequently we have implemented a new Artificial Intelligence (AI) system. The characters are still semi-autonomous, capable of responding to both their own AI decisions and operator control, but their degree of autonomy has significantly increased.

The digital avatars present us with three distinct HCI challenges. First, if the avatars are to respond appropriately to the situation in the physical world then they must be able to sense the actions of the trainees. Second, even the best machine sensing and AI algorithms are not as good as those of a human. Consequently, while we want our avatars to be as autonomous as possible, they should still be responsive to direction from the instructor/operators. Finally, we need to enable the instructor/operators to modify the characters' performances and to create new scenarios. This is particularly challenging because, while the instructor operators are quite capable and are experts in their domain, they are not deeply knowledgeable with respect to computer programming or artificial intelligence (AI).

2.1 Sensing the Environment

The first step of nearly every approach to AI is to sense the current state of the environment, so that appropriate actions can be selected. The AI that we see in video games and training simulations has improved significantly over the last decade or two, but these AIs have the advantage of perfect knowledge of everything that happens in the game or simulation. In a mixed reality scenario, on the other hand, the trainees take their actions in the physical world. As a result, we need to be able to sense the trainees' physical actions in order to allow our characters to respond.

One form of sensing technology, included at all IITs, is a shot-tracking system that was implemented by NAVAIR Orlando. This system uses an infrared (IR) camera to detect hotspots on the screen. These hotspots correspond to the impact of a DITS laser (which is part of the training system used by the Marines to register hits when weapons are fired with blank rounds), so we can use them to determine not only that the Marines have shot at the screen, but also whether or not they hit one of our avatars.

More recently we have integrated the Microsoft Kinect, which is a motion sensing input device originally developed for use with the Xbox 360. The Kinect includes a visible light camera, an active IR sensor, and a multi-array microphone. It provides:

- The position and orientation of each Marine in range
- Skeletal animation data for each Marine in range
- Speech recognition, as well as the direction of any sound

Using this data, we are able to provide the AI with the following sensory information:

- Whether there are any Marines in the room.
- If Marines are in range of the Kinect, we can determine their location and the direction that they are facing. If we have multiple Kinects in the room, we can use their microphone arrays triangulate estimated positions of any Marines that are speaking even when they are not within the Kinect's usual detection range.
- We can watch for specific poses, such as whether a Marine's weapon is raised or lowered. We can couple this with the orientation of the Marine's torso in order to get a rough idea of the direction that they are aiming their weapon (so avatars can play appropriate responses when the Marines aim at them), although obviously this is inexact. Figure 2 shows a user with their weapon raised, the skeleton detected by the Kinect, and the avatar's response (inset in the bottom right).
- We can use the Kinect's speech recognition capabilities to allow the avatars to respond to specific utterances (such as "surrender!" or "get down!"). In addition, even if we don't have a response to a particular utterance, we can still recognize the fact that the Marines are speaking and play a generic response (perhaps something like "No speak English! Don't shoot! Don't shoot!").

Both sensing systems (the NAVAIR shot tracker and the Kinect) are imperfect. They may fail to detect valid events or falsely detect events that didn't occur. In particular, the Kinect can be very sensitive to placement (it has a fairly limited sensing area), and when there are more than two humans in the room it will only detect skeletons for two of them (though it will provide position and orientation for the rest). Frustratingly, we don't get to select which users have their skeletons tracked. In addition, because the Kinect uses an active IR sensor, if two Kinects are placed in close proximity or if the Kinect is placed too close to the shot tracker, the systems may conflict. With that said, these systems are good enough to allow the avatars to display realistic responses to many trainee actions – and the operators can take control when they don't.

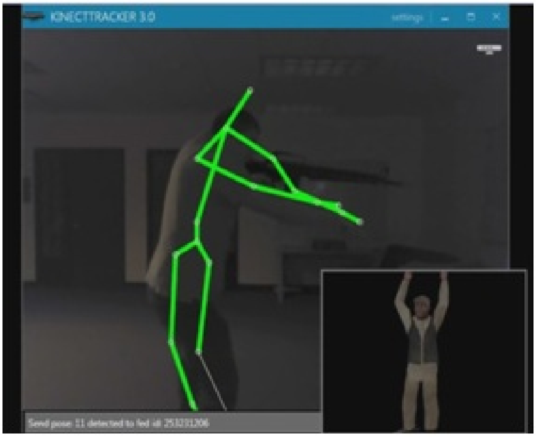

Fig. 2. Pose detection with the avatar's response

2.2 Semi-autonomous Control

Although the Kinect gives us a far greater ability to sense the Marines' actions than we had previously, it is still limited. Furthermore, even if we had perfect sensing the best current AI techniques are not at the level of human intelligence. Consequently, it makes sense to retain the ability for instructor/operators to take control.

Unfortunately, humans are also imperfect. The IIT instructor/operators have two interfaces that they can use to directly control the avatars, each of which has limitations. One is via the instructor/operator station (IOS), shown in Figure 2, which is located in a remote command center where the instructors observe and react to situations viewed in video and audio feeds available throughout the facility (this is far more difficult than one might think). The other is via a standard TV remote control programmed to a limited set of commands and controlled by an instructor who is moving with the trainee unit. Either way, their attention is often divided, and their reaction time can never be as fast as that of a machine. While we do want to provide the instructors with high-level directorial control over the avatars, we certainly don't want them to have to micro-manage every action that the characters take. Thus the ideal approach is to have an autonomous AI which can respond to most likely actions on the part of the trainee and maintain moment-to-moment believability, while allowing the operator to provide high-level commands that change the overall direction of the performance.

With all of that in mind, each scenario specifies both scenario-wide commands and entity-specific commands. The scenario-wide commands change the behavior of all entities in the scenario, while the entity-specific commands are sent to a specific individual. Examples of common commands include things like "fire at the Marines," "surrender," or "wave your weapon and rant." The instructor/operator still retains ultimate override capability using the IOS and remote control interfaces, but the AI augments the instructor's ability to manage a realistic, immersive scenario.

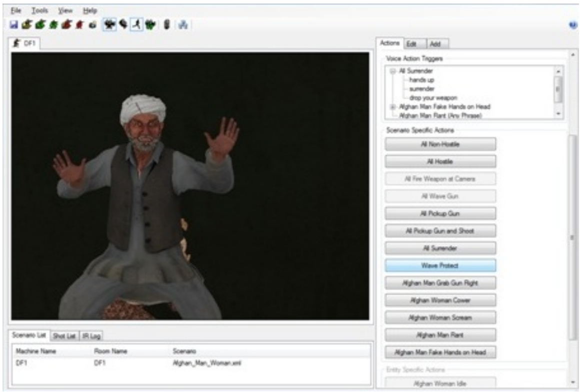

Fig. 3. The Instructor Operator Station. The actions of the digital avatars are shown on the left. The voice triggers appear on the top right, and buttons for action selection appear below them.

A full description of the AI architecture is beyond the scope of this paper, but the general approach is similar to that taken by the Angry Grandmother character created as part of ONR's Future Immersive Training Environments (FITE) Joint Capabilities Technology Demonstration (JCTD) [2], and is built on top of our Game AI Architecture (GAIA) [3]. That work has been extended with the notion of *roles*. Roles are assigned to characters when the scenario is created, and define the purpose that each character serves in the scenario. Currently, most scenarios assign the characters to "hostile" and "non-hostile" roles. The roles limit the actions that will be available to each character, preventing the instructor/operators from unintentionally selecting a nonsensical action. They also define the avatars' actions when under AI control or when they receive a scenario-wide command.

2.3 Scenario Creation

In addition to having overarching control of the digital avatars during scenario execution, the instructor/operators also want to be able to create their own training scenarios. They may do this on their own or in conjunction with the unit commanders for the Marines being trained. We can and do support them in this endeavor, but they shouldn't have to turn to us every time they need a new scenario constructed.

GAIA's modular architecture was designed in large part to make it easier to build the AI by "plugging together" large, conceptually meaningful pieces [4]. Nevertheless, GAIA components still require significant configuration, making the architecture most appropriate for use by somebody with a strong technical background. With that in mind, we have worked to find standard configurations for the GAIA modules, so that they can be dropped into a character's AI without further tuning [5]. This is made simpler because of the relative lack of complexity in our inputs. Thus a GAIA configuration for one of our digital avatars contains a dual utility reasoner. The reasoner's options each contain one or more considerations that are used to determine which option to select. The considerations that we support are:

- **Operator Command**: Activates when the appropriate command is received from the instructor/operator.
- **Track Detected**: Activates if Marines have been detected in the avatar's room.
- **Shot Detected**: Activates if shots are detected by the shot tracking system.
- **Utterance Detected**: Activates if a specific phrase is uttered by the Marines. Alternately, "any phrase" can be selected, in which case this consideration will activate any time a Marine speaks.
- **Pose Detected**: Activates if a particular pose is detected (e.g. if the Marine raises his weapon).
- **Cooldown**: Prevents the option from running twice in close succession.

In the terminology we previously proposed [5], all of these are "opt-in" considerations except for the Cooldown consideration, which is "opt out." A full discussion of dual utility reasoners and GAIA configuration can be found in our previous papers [2-5].

3 Augmented Reality

"Augmented reality" (AR) is a term that has been used in a variety of ways. Often, it simply refers to the overlay of information (such as intelligence data) on top of the physical world. Under the FITE JCTD, and more recently as part of AITT, we have been working with our teammates at SRI International to develop technology that, rather than overlaying information, can overlay virtual characters and events. So, for example, you might be looking at a physical landscape but we would insert a virtual plane, or a virtual tank, or a virtual explosion. This requires:

- A Head-Mounted Display (HMD) or augmented prop
- A 6-axis head/prop tracker
- A 3D model of the static physical environment
- A man-wearable computer
- A Remote-Enhanced Instructor Station (R-EIS)

In brief, the data from the head tracker is used to determine the position and orientation of the user's head in the physical environment. From this, we can determine where on the HMD to draw the image of our virtual objects. The model of the physical environment is used to determine which portions of the virtual objects would be visible to the user, and occlude them as appropriate. All computation and rendering occurs on the man-wearable computer, and the R-EIS controls the training scenario. A more detailed description of each subsystem appears below.

The Head-Mounted Display. Two approaches exist to combining a view of the physical world with virtual objects on the HMD. The optical see-through approach uses a translucent HMD. In this case the user can see the physical world directly, and we simply augment their view by displaying the virtual objects in the appropriate positions, appropriately occluded. This approach would generally be preferable, but the technology is not as advanced. The video see-through approach uses a head-mounted camera to display the physical world, along with augmentation, on an opaque HMD.

The ideal HMD would support the resolution and field of view of normal human vision. In addition, it would support near and far vision, allow us to attach the head-tracking system directly to the HMD, and would be compatible with arbitrary external optical devices (such as binoculars). Unfortunately, currently available HMDs fall far short of this. For example, where human vision is roughly 160x135 degrees, typical commercial HMDs only support a 60 degree diagonal display. The resolution is lower as well – typically about 2 arc minutes per pixel, as opposed to 50 arc seconds for 20/20 vision or 1.5 arc minutes for 20/30 vision [6].

Augmented Props. Observing the enemy, determining, and marking his location is often accomplished through the use of optical devices. Consequently our training needs to support these as well. The ideal solution would be to use HMDs with the standard optical devices, but this is impractical. Instead we have created props that mimic the form, fit, and function of the real devices but replace the optical paths with cameras and miniaturized displays. Currently we support two common devices used by the Marine Corps: the Vector 21 (sophisticated binoculars with orientation and laser range-finding capabilities in both 7x and 10x magnifications) and the Portable Lightweight Designator Rangefinder (PLDR), which is used to mark a target for attack. The Vector 21 prop is fully functional and can be connected to a Defense Advanced GPS Receiver (DAGR) for calculating the location of the virtual target in the exact same manner that the trainee would use with a live target.

The Head Tracker. Our head-tracking technology is provided by SRI International. It incorporates numerous sensors, including cameras for performing both visual feature landmark tracking and visual odometry, an inertial measurement unit, a GPS, a magnetometer, and a barometric pressure sensor. This system combines these inputs and then communicates the position and orientation of the user's head, along with the image from the forward looking camera [7].

The 3D Model of the Physical Environment. The rendering system needs to ensure that it only displays objects that are observable from the trainee's current position. To accomplish this, a 3D terrain skin is developed using the best data available, and this is supplemented with real-time models of dynamic objects (such as vehicles, structures, furniture, and other humans). The better the data, the better the system will be at properly occluding objects in the scene (such as a helicopter flying over a hill, or an insurgent peering around a corner). Currently occlusion works poorly for nearby targets or when high precision is required, but reasonably well on distant targets.

The Man-Wearable Computer. This computer runs both the tracking system software and the rendering software. We currently use a high end laptop with a modern graphics card, but the goal for the program is to transition to a single board solution utilizing embedded mobile graphics chips. This would allow the computer to be carried in a lightweight backpack or embedded directly into the props. Because the graphics system is only displaying a limited number of entities and weapons effects (and not a detailed environment with complex terrain and numerous characters) the typical rendering load should be manageable on these less powerful systems.

The Remote-Enhanced Instructor Station. The R-EIS provides the ability to inject friendly, enemy, and neutral entities into the scenario, create Combat Air Support aircraft, and perform Call for Fire missions with simulated artillery units. The interface is built on an iPad (Figure 4), allowing the instructor to move around with the trainees while controlling both friendly and enemy forces.

Fig. 4. The Remote Enhanced Instructor Station

3.1 Augmented Immersive Team Training

In theory, we could alter reality in whatever ways we want. For example, we could build a simple concrete training environment and then “paint the walls” with whatever materials we want (wood, stucco, wallpaper, etc.) to change the appearance to match the target environment. We could create virtual opponents who move in the physical world and are indistinguishable from physical opponents. This level of fidelity has the potential to truly transform the ways in which we do training. Unfortunately, the accuracy of the head tracker and the 3D model of the physical environment are not yet good enough to make this possible – although they have been rapidly improving.

In contrast, Forward Observer/Forward Air Controller (FO/FAC) training is easily conducted with this system. A typical scenario requires only a Vector 21 prop and an R-EIS. The trainee would use the Vector 21 to observe the enemy units placed by the instructor, and then connect his DAGR to the Vector 21 to calculate the real-world location of the virtual enemy (or use a paper map to determine the location by hand). Next, a fire mission is developed and communicated to the instructor. The instructor uses the R-EIS to insert virtual artillery detonations at the appropriate time and location. The student uses the Vector 21 to observe the incoming rounds and provide corrections back to the instructor. This entire scenario is possible with minimal staff and resources, and yet is still conducted outside in a realistic working environment, where the trainee has to deal with real-world weather situations.

Forward Air Controller training is likewise possible with the system. The FAC can track/locate the enemy in a similar fashion, and then use the HMD to observe and clear attack aircraft in bombing missions. By adding the PLDR laser designator and flight simulator, a much more complex scenario can be employed that has the trainee designating the target and directing the pilot who is flying on the simulator.

4 Conclusion

Mixed reality and augmented reality are new paradigms for Human-Computer Interaction with the potential to transform the ways in which we use computers. The supporting technologies for them are available now, and are improving by leaps and bounds every year, so now is the time to start thinking about how we can make use of them. In this paper we described two efforts that are currently using these technologies in the context of training US Marines.

References

1. Fry, D.J., Meas, P., Cerritelli, L., Schaffer, R.L., Pair, J.: Mixed-Reality Urban Combat Training with the Infantry Immersive Trainer Program. In: 2008 SISO Spring Interoperability Workshop (2008)
2. Dill, K.: A Game AI Approach to Autonomous Control of Virtual Characters. In: 2011 Intraservice/Industry Training, Simulation and Education Conference (2011)
3. Dill, K.: Introducing GAIA: A Reusable, Extensible Architecture for AI Behavior. In: 2012 SISO Spring Interoperability Workshop (2012)
4. Dill, K., Pursel, E.R., Garrity, P., Fragomeni, G.: Achieving Modular AI through Conceptual Abstractions. In: 2012 Intraservice/Industry Training, Simulation and Education Conference (2012)
5. Dill, K., Pursel, E.R., Garrity, P., Fragomeni, G.: Design Patterns for the Configuration of Utility-Based AI. In: 2012 Intraservice/Industry Training, Simulation and Education Conference (2012)
6. Australian National University, College of Engineering and Computer Science, `http://escience.anu.edu.au/lecture/ivr/sight/eyeMainCharacteristics2.en.html`.
7. Kumar, R., Samarasekera, S., Chaudry, A., Zhu, Z., Chiu, H.P., Oskiper, T., Villamil, R., Branzoi, V., Hadsell, R., Pursel, E.R., Dean, F., Garrity, P.: Implementation of an Augmented Reality System for Training Dismounted Warfighters. In: 2012 Intraservice/Industry Training, Simulation and Education Conference (2012)