Collaborative Helmet and Weapon Tracking for Augmented Reality Based Training

Supun Samarasekera, Rakesh Kumar, Taragay Oskiper, Zhiwei Zhu, Glenn Murray, Kevin Kaighn, Nicholas Vitovitch, Andy Coppock, Ali Chaudhry

SRI International

Princeton, NJ

supun.samarasekera@sri.com, rakesh.kumar@sri.com

Frank Dean, Pat Garrity

U.S. Army Research Laboratory, Human Research and Engineering Directorate, Advanced Simulation Technology Division (ARL-HRED-ATSD)

Orlando, FL

frank.s.dean6.civ@mail.mil, patrick.j.garrity4.civ@mail.mil

ABSTRACT

There is a need within the military to have increased levels of warfighter proficiency and readiness by providing realistic training scenarios for complex urban combat at forward and home locations. Currently, elaborate infrastructure and supporting actors are needed to create training scenarios, and record and review training sessions. Live ammunition training is limited to Force on Target training with extremely limited scenarios (no movers, same old targets); while laser-based training does allow Force on Force training, it is limited by the scheduling of exercises, range time availability and scenarios possible with live forces.

The key emerging innovative technology that addresses these shortcomings is precision mobile Augmented Reality (AR). The AR system precisely tracks actions, locations, and head and weapon pose of each trainee in detail so the system can appropriately position virtual objects in the trainee's field of view. Synthetic actors, objects and effects are rendered by a game engine on the eyewear display. Synthetic actors respond in realistic ways to actions of the trainee, e.g., taking cover, firing back, or milling as crowds. The AR-weapon can be used to fire simulated projectiles at real or synthetic entities.

This paper describes improvements made to a prototype AR system based on live testing with warfighters at a Military Operations in Urban Terrain (MOUT). We present a method for joint tracking of the helmet worn and the weapon attached sensors in a collaborative fashion in which the wearable unit on the helmet aids the weapon unit by sharing visual landmarks along with 3D location estimates in the scene. These are stored in a dynamic map on the weapon processor and continuously matched against to obtain weapon poses consistent with the head pose to provide accurate aiming capability. We also present solutions to miniaturize the system using mobile processors and smartphone sensors.

ABOUT THE AUTHORS

Mr. Supun Samarasekera is currently the Technical Director of the Vision and Robotics Laboratory at SRI International. He received his M.S. degree from University of Pennsylvania. Prior to joining SRI, he was employed at Siemens Corporation. Mr. Samarasekera has 15+ years of experience in building integrated multi-sensor systems for training, security and other applications. He has led programs for robotics, 3D modeling, training, visualization, aerial video surveillance, multi-sensor tracking and medical image processing applications. Mr. Samarasekera has received a number of technical achievement awards for his technical work at SRI.

Dr. Rakesh "Teddy" Kumar is currently the Director of the Center for Vision Technology at SRI International, Princeton, NJ. Prior to joining SRI, he was employed at IBM. He received his Ph.D. in Computer Science from the University of Massachusetts at Amherst in 1992. His technical interests are in the areas of computer vision, computer graphics, image processing and multimedia. Dr. Kumar received the Sarnoff Presidents Award in 2009 and Sarnoff Technical Achievement awards in 1994 and 1996 for his work in registration of multi-sensor, multi-dimensional

medical images and alignment of video to three dimensional scene models respectively. He was an Associate Editor for the Institute of Electrical and Electronics Engineers (IEEE) Transactions on Pattern Analysis and Machine Intelligence from 1999 to 2003. He has served in different capacities on a number of computer vision conferences and National Science Foundation review panels. Dr. Kumar has co-authored more than 50 research publications and has received over 35 patents.

Dr. Taragay Oskiper is a Senior Principal Research Scientist at SRI International. He received his Ph.D. in Electrical Engineering from Princeton University. His main area of research is in visual-inertial navigation for simultaneous localization and mapping. Dr. Oskiper has over 15 years' experience in developing vision-aided motion estimation and multi-sensor fusion algorithms for navigation and AR applications for both video-see-through and optical-see-through platforms. He has been the lead algorithm developer for numerous augmented reality projects, most recently the Office of Naval Research Augmented Immersive Team Training (AITT) program.

Dr. Zhiwei Zhu is a Principal Scientist at SRI International. He received his Ph.D. in Electrical Engineering from Rensselaer Polytechnic Institute, Troy, NY. His main research focus is in the area of Computer Vision and Human Computer Interaction. He has published over 40 journal and conference papers. Dr. Zhu received the Best Transaction Paper Award from IEEE Transactions on Vehicular Technology in 2004 for his driver fatigue monitoring work. He received another Best Paper Award at IEEE Virtual Reality Conference in 2011 for his co-authored work in the high-precision localization and tracking for the large-scale infrastructure-free augmented reality applications.

Dr. Ali Z. Chaudhry is a Program Director at SRI International. In this capacity, Dr. Chaudhry has managed several AR and training programs—most notably the Integration of Technologies into the Future Immersive Training Environment Augmented Reality (FITE-AR) System with the Office of Naval Research and, Future Immersive Training Environment Seamless Indoor/Outdoor Tracking of Marines and Weapons, for the Marine Corps Warfighting Labs. He has also served as Program Manager for several key aerial programs, including Night Eagle, Desert Owl and the DARPA HART program. Dr. Chaudhry is a member of the PMI and has extensive experience applying earned value project management.

Mr. Frank Dean is an Engineer and Science & Technology Manager at the U.S. Army Research Laboratory- Human Research and Engineering Directorate, Advanced Simulation Technology Division (ARL-HRED-ATSD), Orlando, FL. He currently works in the Ground Simulation Environments Division conducting R&D in the area of dismounted soldier training & simulation. Mr. Dean is a former U.S. Army signal officer and has over 30 years of military and government civilian service. Prior acquisition assignments have included managing technical programs for PM IEW/RSTA, PM Army Air Traffic Control, and STRICOM's (PEO STRI) Live Simulation Systems Division. His current interests revolve around researching augmented reality techniques and their potential application in the live training environment. Mr. Dean has earned a B.S. in Electrical Engineering from the University of Miami and his Masters of Engineering Management (M.E.M.) from George Washington University.

Mr. Pat Garrity is the Chief Engineer for Dismounted Soldier Technologies at U.S. Army Research Laboratory - Human Research and Engineering Directorate, Advanced Simulation Technology Division (ARL-HRED-ATSD), Orlando, Fl. He currently works in Ground Simulation Environments Division conducting R&D in the area of dismounted soldier training & simulation where he is the Army's Science & Technology Objective Manager for the Augmented Reality for Training Science and Technology Objective (STO). His current interests include Human-In-The-Loop (HITL) networked simulators, virtual and augmented reality, and immersive dismounted training applications. He earned his B.S. in Computer Engineering from the University of South Florida in 1985 and his M.S. in Simulation Systems from the University of Central Florida in 1994.

Collaborative Helmet and Weapon Tracking for Augmented Reality Based Training

Supun Samarasekera, Rakesh Kumar, Taragay Oskiper, Zhiwei Zhu, Glenn Murray, Kevin Kaighn, Andy Coppock, Ali Chaudhry

SRI International

Princeton, NJ

supun.samarasekera@sri.com, rakesh.kumar@sri.com

Frank Dean, Pat Garrity

U.S. Army Research Laboratory, Human Research and Engineering Directorate, Advanced Simulation Technology Division (ARL-HRED-ATSD)

Orlando, FL

frank.s.dean6.civ@mail.mil, patrick.j.garrity4.civ@mail.mil

INTRODUCTION

To train warfighters for modern warfare, live exercises are held at various Military Operations on Urban Terrain (MOUT) facilities. This training may also happen near the battlefield. However, setup and configuration of an instrumented training site is time-consuming, laborious and costly. For effective training, commanders need to have situational awareness of the entire mock battlefield and also the individual actions of the dispersed units and various warfighters. Instructors must be able to provide instant feedback and play through different actions and what-if scenarios with the warfighters. There is a need for accurate measurement, capture and analysis of warfighter movements at a detailed level. Additionally, providing a wide range of training scenarios with different emphasis and different difficulties tailored for individual teams and individual warfighters is critical for improving training efficiency. Realistic training requires large numbers of actors to role-play opposing forces and crowds in the environment. Logistics of gathering such a large group of people is difficult and costly.

Concept planning for the Army's vision for future training capability (Synthetic Training Environment (STE) and the follow-on Future Holistic Training Environment Live/Synthetic (FHTE-L/S)) has highlighted augmented reality (AR) as a solution to address a major gap in past approaches to integrating live, virtual, constructive and gaming environments. This shortfall is the fact that live players participating in Live-Virtual-Constructive-Gaming (LVCG) events cannot (1) observe, (2) react to, nor (3) execute appropriate actions and maneuvers in response to events emanating from the virtual or constructive domains without assistance from observer-controllers (O-C). As such, this workaround introduces varying levels of "negative training" [Dean 2016]. Negative training is practicing procedures in a manner inconsistent with how an action would be performed in combat, which results in the development of bad habits [Report 2002].

Maintaining a proper azimuth for AR's overall development in support of STE and FHTE-L/S requires periodic field technology demonstrations; including warfighter participation and input. Early and periodic input and recommendations are critical in aiding developers in focusing in on key capabilities that are necessary for the eventual delivery of an effective training system that provides a positive training experience. With this in mind, Army researchers and contractor support personnel participated in the Army Warfighting Assessment (AWA) 17.1 event during the period of 12-16 September 2016 (Figure 1). The purpose of this event was to demonstrate future training capabilities that address Army Warfighting Challenge 8 – "Enhance Realistic Training."

Figure 1. Testing of Dismount AR Training System at AWA 17.1.

Following brief instructions and familiarization with the AR equipment, the unit was able to conduct platoon-level situational training exercises (STX) involving a dismounted assault on a MOUT facility. Fire team sized elements conducted room clearing drills, with the mounted elements in support and overwatch. Following the completion of each training scenario, participants provided feedback through interviews and written surveys. Several comments expressed by warfighters, in the written surveys, regarded the M4 weapon's interface and sighting capability. Observation of the training and video playback of warfighter/weapon views convinced researchers that the interplay and synchronization between the warfighter and the weapon must be more robust and accurate. This paper describes improvements made to a prototype AR system (Dismounted AR Training system) based on live testing with warfighters at AWA 17.1 exercises.

PREVIOUS WORK

Most dismounted training systems today rely on physical targets or role players to represent opposing forces during exercises. In live fire training, range systems use physical targets (e.g., paper pop up silhouettes) that lack realism as targets react in predictable ways to the actions of trainees. Laser-based systems have also been widely used in live training. Role players and warfighters use laser weapons and laser detectors to determine when someone is hit by weapons fire. These systems rely on line-of-sight pairing of the laser to determine engagement. As such, it cannot be used for effecting fixed infrastructure like buildings or for engaging targets through building, walls, windows, etc. Similarly, it is difficult to mimic injuries and reaction of the opposing forces to determine their true reactions in these systems. Such systems also require human actors as role players and as such typically require large numbers of support personnel to run training exercises. Recently a few Mixed Reality Systems such as the Infantry Immersive Trainer [Muller 2010] and the Automatic Performance Evaluation and Lessons Learnt (APELL) system [Cheng 2009] have been deployed at Camp Pendleton and other Marine Corp's MOUTs (Military Operations on Urban Terrain). These systems use video projectors that project images of virtual actors on walls of rooms within a training facility. However, these systems are limited to indoor exercises and require significant infrastructure.

Existing systems also have a limited ability to track trainees during exercises, and to adapt virtual actions to the movements of the trainees. Current systems used for tracking trainees at a MOUT require significant infrastructure to be installed beforehand. The systems also require time-consuming procedures for preparing the environment. There are very few systems that can track trainees both indoors and outdoors. Global Positioning System (GPS)-based systems [Saab, 2010] may be used for providing location outdoors. However, the performance of these outdoor-only systems decreases in challenging GPS-limited situations. .

Recently there has been a number of experimental AR systems developed for training warfighters for different functions, e.g., observers [Kumar 2013, Schaffer 2015] and vehicle operators [Brookshire 2015]. Experiments conducted with the Army at AWA 17.01 in 2016 for training dismounts for close quarter battle were based on the technology presented in [Samarasekera 2014]. In this paper, we present design and evaluation results of an improved system based on feedback received from those tests. A new system has been designed to overcome shortcomings identified of the previous design. In the subsequent sections, we present the overall approach and describe the system modules. For each module, we discuss shortcomings of the previous approach and rationale for the new design. We then present evaluation results and final conclusion and impact to the Army for training.

OVERALL APPROACH

The AR system precisely tracks the actions, locations, and head and weapon pose of each trainee in detail. As such, the system can appropriately position virtual objects in the trainee's field of view. The helmet- and weapon-mounted sensors are used to locate the trainee and his gaze and weapon pose with respect to the pre-mapped 3D environment. Both the trainee's head pose and weapon's 6-degrees of freedom (DOF) pose are fed to a simulation game engine. Within the simulation engine, synthetic avatars and objects are rendered to enhance the activity observed in the real-environment. Stereo-based 3D reasoning is used to occlude all or parts of synthetic entities obscured by real world 3D structures based on the location of the synthetic entity. These synthetic entities, avatars, and effects are inserted into the live view on the eye-wear or Head Mounted Display (HMD) for real-time engagement with the trainee. Synthetic actors respond in realistic ways to actions of the trainee, e.g., taking cover, firing back, or milling as crowds. The AR-weapon can be used to fire simulated projectiles at real or synthetic entities. Finally, the AR system is

designed to be infrastructure free. The primary hardware needed to implement the solution is worn by the individual trainees.

Software Architecture

Figure 2 below shows the software architecture of the AR system. A modular architecture is used to interconnect the core modules. The primary module is the dismount body-worn unit. In this module, all sensor data is connected to the rest of the system through a sensor abstraction layer. This allows for upgrades and hardware modification without modifying the algorithm modules. The stereo module runs primarily on the GPU and computes depth from a pair of images. The localization module will run on the mobile ARM CPU and generates 6-DOF poses of the trainees' head. The localization module is used to compute low-latency high-rate poses for the renderer. The localization module also provides visual feature updates and poses to the weapon module and assists computing of relative aiming information of the weapon. The pose data and depth images are sent to the Unity graphics engine for rendering AR content. The dismount weapon uses the same sensor head and mobile ARM processor and same localization engine to track the weapon position. Weapon poses and triggers are then wirelessly transmitted to the AR rendering engine. The rendering engine runs on the same machine as the dismount-worn system. It is connected to a central game server that coordinates the entity actions across the multiple dismount units. Low-latency poses of the dismount, depth estimates, and weapons poses are used by the AR-renderer to render synthetic entities and effects correctly onto the HMD. The game-server provides coordinated actions to the dismount system. The system uses a scenario scripting mechanism that is used by the game-server to coordinate synthetic character actions. Based on the dismount and weapon poses, the trainer can set up triggers that will activate the actions of synthetic entities.

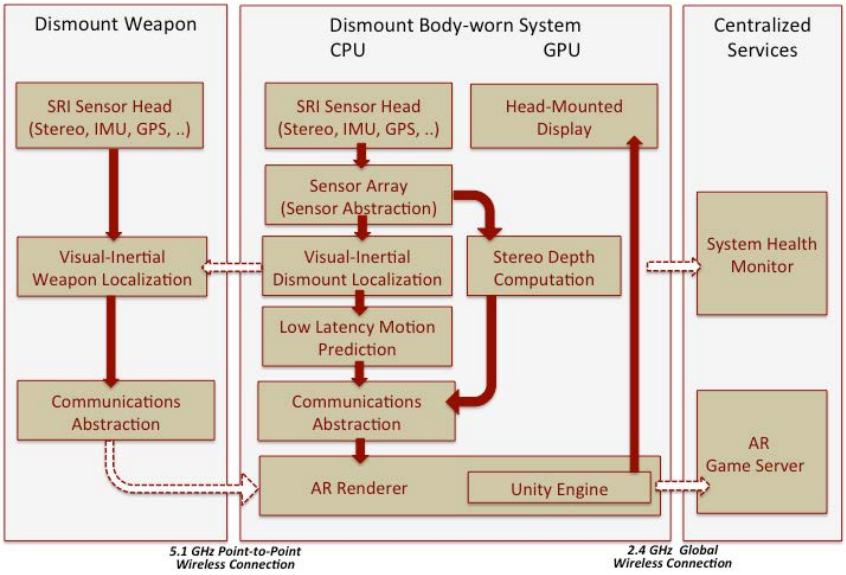

Figure 2: Architecture for Dismount AR Training System.

Dismount Head Tracking System

The approach to dismount tracking uses a visual-inertial, multi-sensor navigation framework for precisely locating trainees at a training site (Figure 3). This includes two key components: (1) Precise relative localization that tracks the dismount movement in 6-DOF, and (2) A global-position system that uses GPS and pre-built visual-landmark database for locating the dismount accurately in the global coordinate frame of the training site. These two capabilities run concurrently and are fused within an Error-State Kalman filter[†] in real-time. The framework developed by SRI can support use of a monocular camera, stereo-cameras or multiple cameras for the tracking system. The system can generate 6-DOF poses in real-time at > 100 Hz with < 10-ms latency. This is critical for AR systems.

[†] A Kalman filter is an algorithm that uses a series of measurements observed over time, containing statistical noise and inaccuracies, to produce estimates of unknown variables that tend to be more accurate than those based on a single measurement, by using Bayesian inference and estimating a joint probability distribution over the variables for each timeframe.

We have identified key issues with the tracking system described in [Samarasekera 2014] when testing at the MOUT in AWA 17.01. We describe the solution for each of these issues.

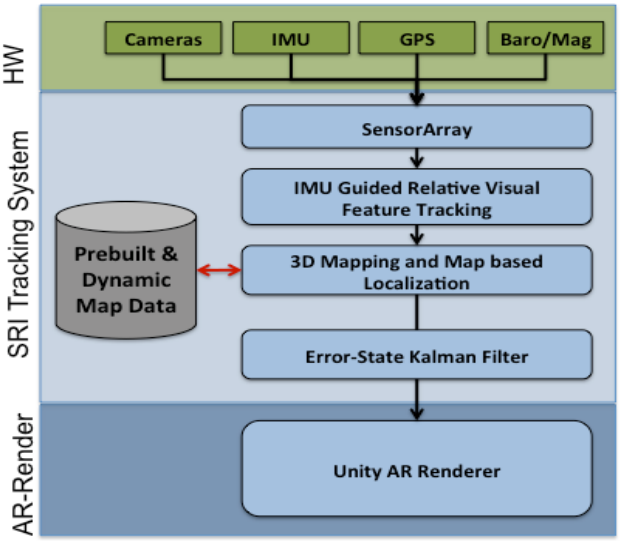

Figure 3. Dismount Head Pose Tracking using Multiple Sensors.

Handling variations in scene lighting and appearance: Significant variation in lighting affected the video that is captured and processed for the vision-based tracking (significantly saturated or dark and noisy images). This affected the quality of visual feature tracking and visual-landmark matching processes. To address these challenges, we made a number of changes. First, we significantly sped up the camera dynamic aperture adaptation when moving from bright outdoors to dark indoor areas (and vice versa) to ensure faster reacquisition of pose. This greatly improves capture of good images and reduces the amount of time we have either dark images or overly saturated images during transition from indoor to outdoor or vice versa. Second, we updated the pre-build landmark database approach with the addition of visual landmarks to account for lighting and appearance changes in the environment from when the database was recreated. We implemented a comprehensive simultaneous-localization and mapping (SLAM) system to update the pre-built databases. The database can be updated rapidly before the exercise.

The SLAM system (Figure 4) executes a concurrent pipeline in which odometry and mapping threads are operated in parallel to take advantage of the multi-core processor architecture in mobile computers, which have become ubiquitous. SLAM is used both for updating the 3D model and for navigation during execution on an exercise. During map updating steps, loop closures are used to reset drifts in navigation paths due to visual odometry alone. Figure 4 shows a high level system block diagram with bidirectional interaction between the odometry and mapping blocks.

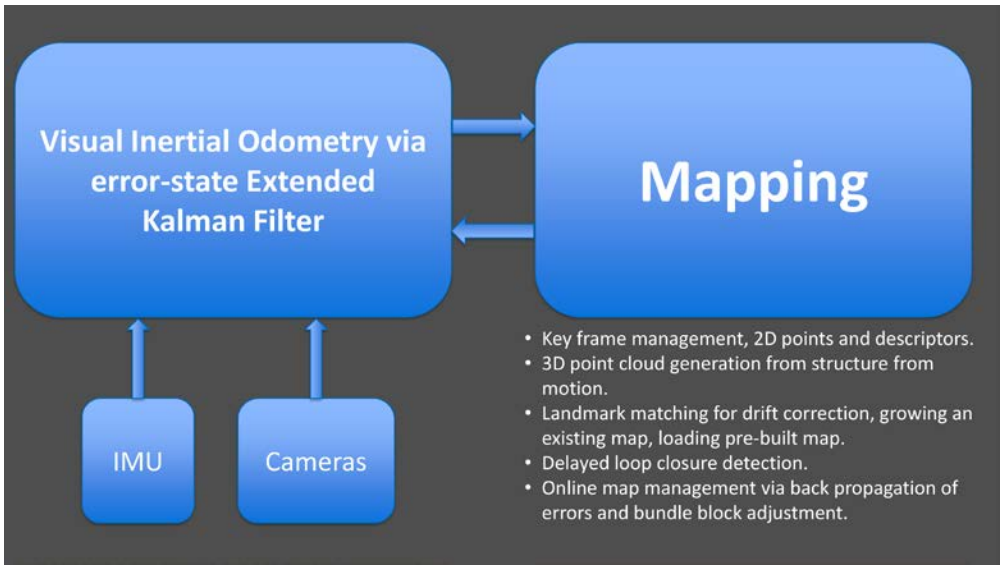

Figure 4. SLAM System for Simultaneous Navigation and Mapping.

In the map creation stage, main information flow is mostly from the odometry block to the mapping engine in terms of feature tracks and camera pose data, except for loop closure events at which point it is in the reverse direction and the Error-State Kalman Filter is reset after each loop closure to maintain a consistent coordinate system with the mapping side. During map updating, the feature tracks and poses in a detected loop are fed to a bundle block to update the 3D location of tracked scene points and poses of key frames. During the execution stage with map-based

navigation using preloaded map data, the information flow is mostly bidirectional. The odometry block provides query frame pose and feature point data. The mapping engine returns 3D to 2D feature correspondences. The current query frame key points and the 3D point cloud are stored in the map, which are used as global measurements in the Extended Kalman Filter (EKF) to contain the odometry drift. During the map creation stage, a new key frame is inserted into the map every 10 cm of travel since the most recent key frame. For each key frame we store the pose, 2D normalized feature point coordinates along with their descriptors, and pointers to their 3D positions in the map database. To keep the map size small, we only store points that have been tracked over the past three frames indicated by the odometry module. The initial 3D position of each key point is again supplied by the odometry module, which are then refined over time by the mapping module.

Handing dynamic clutter when multiple trainees are in a closed area: In our previous experiments we observed that when multiple dismounts were moving close to each other in limited spaces that they occlude each other's view (from video). This poses a challenge to visual-inertial navigation systems. Close proximity and visual blockage can delay initialization and reacquisition of the landmark being matched. Specifically, in room clearing operation where the whole movement through the building is very short and extremely fast, such delays can have significant impact on the AR content being displayed. To address this challenge we implemented a system using a stereo pair of fisheye, wide field of view cameras. This ensures more of the background scene is visible at any time even in crowded environments. Finally, the optical track system automatically switches between stereo processing to monocular processing, if one of the cameras in a stereo pair is occluded.

Interaction System with Synthetic Entities for Augmented Reality System

The Interaction system involves many modules, including modules for occlusion reasoning, rendering, weapon interaction and game engine. Occlusion reasoning is performed to not render synthetic entities that are occluded by static and dynamic structures. Occlusion by static structures is performed by using the 3D model of the site. Occlusion of dynamic entities is performed by dynamically estimating the 3D depth from the vantage of the warfighter helmet view-point.

The AR-enabled small arms surrogate weapons is used for kinetic engagement of virtual targets. Weapon interaction module tracks the trainee's weapon and state. It knows when the trainee fires the weapon and does geo-pairing to know if a synthetic or real entity has been shot. The rendering system renders the synthetic entities onto the HMD display. The location and state of the synthetic entities are provided by the game engine. The game engine runs on a separate server and connects wireless to each AR system. We use a Unity-based rendering and game engine system. The system also supports interactions with avatars based on live triggers. Artificial intelligence (AI) interaction with the avatars is scripted based on triggers from the trainee's location, weapon location and orientation, and weapon firing. Finally, all systems including rendering are ported to run on the smartphone processor system. We now discuss in more detail our technical approach for occlusion and weapon interaction.

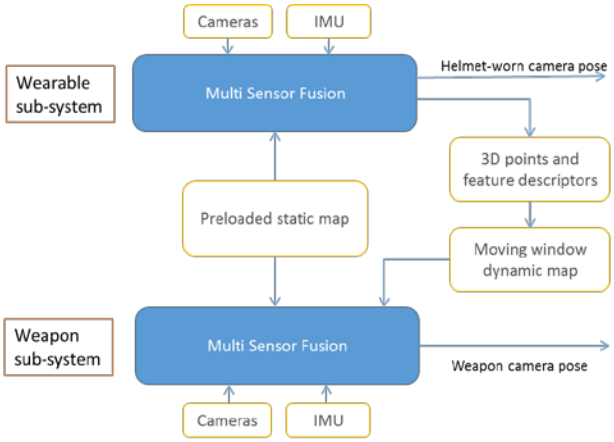

Figure 5. Head Pose and Weapon Pose Tracking.

Weapon Pose Estimation

Second critical aspect we addressed is the weapon pointing accuracy. This allows a dismount to engage virtual targets accurately. During the previous effort we focused on the accuracies at the longer distances and we achieved higher longer-range accuracies. However, in closed quarters the solution was less stable. Our post analysis indicates the accuracies can be improved by better estimating the weapon sensor position and improved calibration of the weapons barrel with respect to the sensor head. To improve estimation of the sensor position we moved our estimation methods from a purely 2D matching method (between the helmet and the weapon) to a 3D/2D estimation. This provides improved positional accuracy. We also improved tracking of the weapon by its sensors alone using the improvements made in previous section using fisheye cameras etc. Final pose estimate for the weapon is achieved by combining the

inference from sensors measurements from weapon sensors and matching of weapon video to helmet video (Figure 5, above).

Occlusion Reasoning

A key component in avatar interaction is occlusion reasoning. The synthetic entities in an AR system must appear correctly occluded with real infrastructure and dynamic entities such as other dismounts. While real-infrastructure-based occlusions can be extracted using pre-built 3D models of the training site, occlusion reasoning for dynamic objects needs to happen in real-time. To enable dynamic occlusion reasoning the dismount needs to be instrumented with a 3D sensor. SRI's approach to 3D sensing uses stereo cameras in which depth is computed in software. We used the same stereo pair that is used for navigation for the depth sensing, allowing for a reduced hardware footprint.

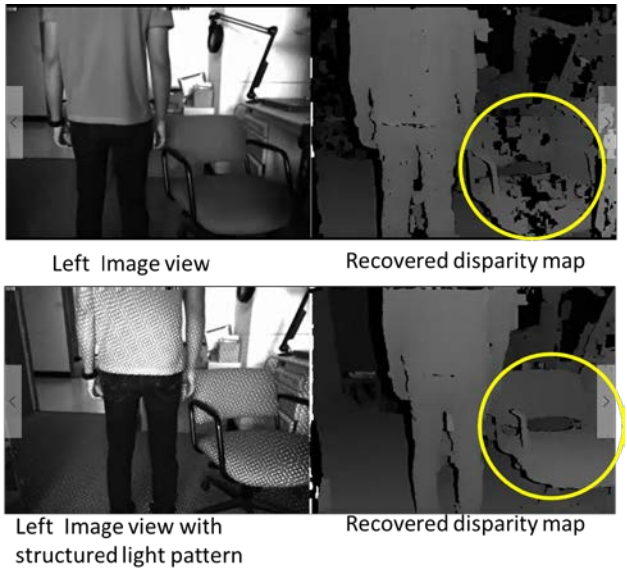

Figure 6. Stereo-based Depth Recovery with and without Structured Light.

However, there are two key deficiencies that were observed during the previous work: (1) when in darker areas or near uniform walls the depth estimates are less reliable, and (2) depth estimates using a software approach introduces latency in the pipeline. The lack of visual information is a fundamental issue for stereo algorithms and there are very limited real-time options available for improvement. We have engineered a solution by introducing IR structured light into our sensor package. The structured light provides synthetic texture in dark and uniform areas providing data for features that can be exploited in the stereo depth computation (Figure 6). We have a separate navigation sensor and depth sensor for each user and use IR filters on the navigation sensors to eliminate interference from the projected IR patterns. Figure 7 shows the 3rd party depth sensor with structured light source we are using, which will cover out up to ~5m.

Figure7. Sensor-Processor Hardware Board with Smartphone-type Sensors and FPGA.

Computation System and Hardware Integration

The previous AR hardware included off-the-shelf Gigabyte mini-computers with Intel processors and COTS cameras and IMU units. However, these systems consumed significant amounts of power and space. The previous sensor system weighed 12 oz. and the processor package was about 120 oz. Mobile processors have evolved significantly and hold the promise of providing the processing bandwidth required by dismount AR systems. Similarly, compact MIPI camera modules on these mobile systems have significantly smaller footprint including the lenses. In the system discussed here, we use smartphone sensors and a mobile processor, and reduced the weight of the sensor system to be < 8 oz. and for the processor system to be < 48 oz. For the sensors we have created a custom Navigation sensor board that integrates stereo-pair of global shutter MIPI camera modules, a VectorNav INS (IMU, Magnetometer) (Figure 7). This allows us tightly synchronize the camera and IMU data while moving to a significantly lighter more compact

sensor head. All data is transmitted via a high-speed USB-3 cable for each sensor to a processor that has minimal latency. We use the Intel NUC processors with Linux OS for running the navigation methods with the sensor and HMD package (Figure 8).

Figure 8. Sensor Head with HMD on NVG Mount, Full AR System with AR-weapon, Intel NUC Processor.

EVALUATION AND RESULTS

We present evaluation results for both helmet tracking and for weapon pointing accuracy. For collecting ground truth for helmet tracking, we had surveyed our property and received 3D coordinates of selected points (Figure 9). For collecting test data with ground truth, we built a custom sensor ground truth rig (Figure 9, right hand side). The ground truth rig has the same sensors as described earlier and shown in Figure 7. The ground-truth collection senor rig that is easy to hold and move around. The sensor rig is pre-calibrated with the tip of rig at the bottom, which is placed on top of the ground-truth surveyed point on the ground. A high accuracy level is used to ensure the sensor rig is level with the ground-truth point. During test collects, we walked through each ground-truth point one by one.

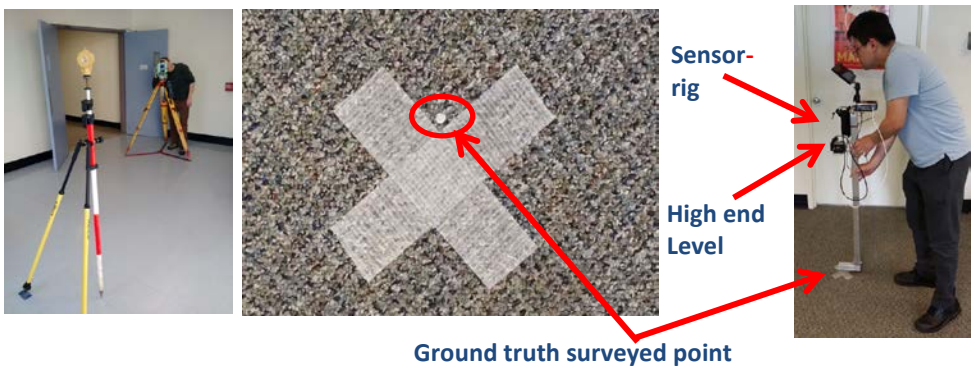

Figure 9: Survey of Ground Truth data and Sensor Rig for Collecting Test Data.

For testing navigation accuracy, we first built 3D landmark databases of different locales in the test site. We then used this 3D landmark database to match to while doing navigation. Figure 10 shows a typical collect performed on the 3rd floor of the SRI Princeton main building. The accuracy of building the landmark map database and the accuracy of tracking are shown in Table 1. In the table, we first show accuracy of building the landmark maps. We show results for each stage of map construction, including open loop, on the fly map refinement with loop closures, and final map accuracy after doing a bundle block. As can be noted from the table, the final median errors for both 1st floor and 3rd floor range from 0.6m to 0.3m. We then show the tracking results using the maps from both the 1st and 3rd floors. We find that the navigation accuracy very nearly approximates the landmark map construction accuracy.

Figure 10: (Left) Stereo Image and (Right) Landmark Map Database Collected on 3rd Floor.

Table 1. Accuracy of Localization.

Data		time	distance	mean	median	90%	gt points
		min's	meters	meters	meters	meters	
Mapping Accuracy Results							
3rd floor map							
SA_2017.06.12_14.29.15*	open loop	17.0034	681.08	4.0302	2.0905	10.7035	24
SA_2017.06.12_14.29.15*	on fly map	17.0034	691.60	0.6691	0.5502	1.5306	24
SA_2017.06.12_14.29.15*	Final Map	17.0034	683.17	0.3359	0.3112	0.6329	24
1st floor map							
SA_2017.06.12_15.00.30*	open loop	19.5576	621.82	1.3069	1.2292	1.7266	16
SA_2017.06.12_15.00.30*	on fly map	19.5576	636.00	0.9802	0.8923	1.3641	16
SA_2017.06.12_15.00.30*	Final map	19.5576	624.95	0.6811	0.6011	1.0245	16
Navigation Accuracy Results							
3rd floor							
SA_2017.06.05_13.18.03	Match to map	14.857	608.74	0.3628	0.3448	0.5468	23
SA_2017.06.05_13.34.35	Match to map	12.7395	513.37	0.4499	0.4887	0.6583	20
1st floor							
SA_2017.06.12_15.20.36	Match to map	9.1911	434.22	0.6316	0.4603	0.9679	16

For accuracy of weapon tracking, we designed a mount to hold a laser pointer along the axis of the left camera (Figure 11), which is weapon navigation pose output; we then spatially transform it to get the pose of the actual weapon barrel. For estimating error in determining weapon pose, we look at the intersection position of the virtual weapon line in the augmented helmet image and compare against the location of the laser point. We use that to compute the aiming error both in pixels and angles. The left bottom insets in Figure 12 shows the estimated pose of both the head (in red) and weapon (in black) for different positions of the user holding the weapon. Figure 13 shows an augmented image captured from the helmet camera. Both the laser pointer dot (white) and the weapon pointing dot (red) are visible. We use the difference in position of these two dots to estimate the error in weapon accuracy. Table 2 shows the calculated errors. Note the median angular error is about 1.2 to 1.7 degrees. Note the laser is offset from the camera barrel. There is also an error between the orientation of the 3D model using for occlusion reasoning and landmark database. We haven't corrected for these factors as of yet in the error estimate. Our estimate is that this offset is causing about 1 degree of error.

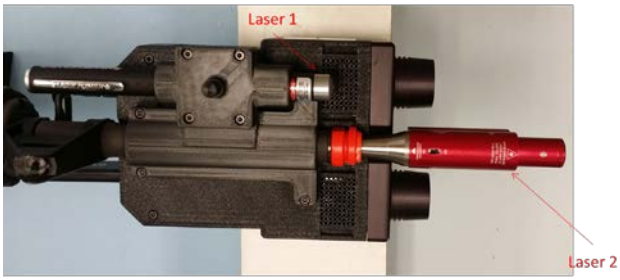

Figure 11. Laser Mounted on Weapon Aiming Camera to Test Accuracy.

Figure 12: Tracking of Weapon and Head Pose Rendered on 3D Model of the Scene.

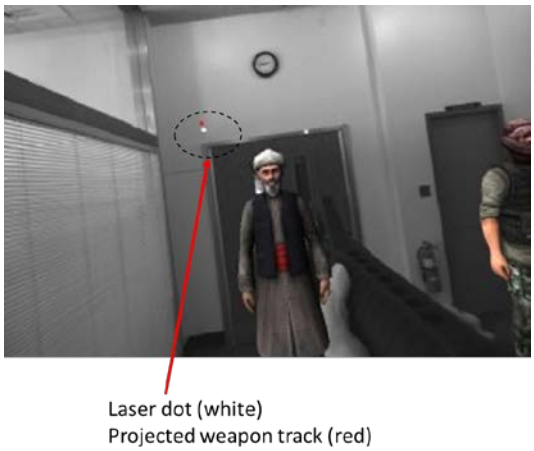

Figure 13: Augmented Helmet Camera Image Showing Both Laser Dot (White) and Projected Weapon Track (Red). Synthetic Characters are also Inserted into the Scene.

Table 2: Accuracy of Tracking Weapon Pose.

	Error between Laser Point and Center of Weapon Pointing						
data set	mean		median		90%		gt. points
	pixels	deg.	pixels	deg.	pixels	deg.	
arsimulator 2017-06-20 18-46-12-48	8.33	1.79	5.61	1.23	17.72	3.78	14
arsimulator 2017-06-20 18-39-34-92	9.11	1.96	8.06	1.78	13.60	2.93	15

CONCLUSIONS

The U.S. Army's future training capability (STE and FHTE-L/S) has highlighted AR as a solution to address a major gap in past approaches to integrating live, virtual, constructive and gaming environments. That shortfall is the fact that live players participating in LVCG events cannot observe, react to, or execute appropriate actions and maneuvers in response to events emanating from virtual or constructive domains without assistance from O-C's. As such, this workaround introduces varying levels of "negative training."

The research presented in this paper highlights the importance of periodically demonstrating research prototypes to trainees in live training environments. Under these conditions, critical shortfalls can be identified that may have been

overlooked in a laboratory environment. Man-wearable AR training is cutting-edge technology that is currently not available to warfighters, except for very limited use. The goal of Army researchers and support contractors is to develop an AR system that would push a combat training center (CTC)-in-a-box training capability down to the squad level. Realistic dismounted training, anywhere, anytime. Finally, lessons learned from this technology demonstration and user surveys are being used as the model for planning and implementation of ARL-HRED-ATSD's Augmented Reality for Training - Science and Technology Objective (ART STO), which began in early FY17.

ACKNOWLEDGEMENTS

The research reported in this paper was performed in connection with contract W911QX-13-C-0052 with the U.S. Army Research Laboratory through the University of Central Florida. The views and conclusions contained in this paper are those of the authors and should not be interpreted as presenting the official policies or position, either expressed or implied, of the U.S. Army Research Laboratory, the U.S. Government, or the University of Central Florida unless so designated by other authorized documents. Citation of manufacturer's or trade names does not constitute an official endorsement or approval of the use thereof. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation hereon.

REFERENCES

[Brookshire 2015] Jonathan Brookshire, Taragay Oskiper, Vlad Branzoi, Supun Samarasekera, Rakesh Kumar, Sean Cullen, Richard Schaffer. Military Vehicle Training with Augmented Reality, *Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC)*, Orlando, FL, 2015.

[Cheng 2009] H. Cheng, R. Kumar, C. Basu, F. Han, S. Khan, H. Sawhney, C. Broaddus, C. Meng, A. Sufi, T. Germano, M. Kolsch, and J. Wachs. An Instrumentation and Computational Framework of Automated Behavior Analysis and Performance Evaluation for Infantry Training. In *Proceedings of 2009 Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC-2009)*, Orlando, FL, 2009.

[Dean 2016] F. Dean, M. Belen'kii, L. Sverdrup, Y. Taketomi. Optical See-Through HWDs – Technical Challenges and Design Evolution, *Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC), 2016.*

[Kato 1999] H. Kato and M. Billinghurst, Marker Tracking and HMD Calibration for a Video-based Augmented Reality Conferencing System. *Int'l Workshop on AR*, pp.85-94, 1999.

[Kumar 2013] Rakesh Kumar, Supun Samarasekera, Mikhail Sizintsev, Taragay Oskiper, Vlad Branzoi, Richard Schaffer, Sean Cullen, Nikhil Krishnaswamy. Augmented Reality Training for Forward Observers, Interservice/ Industry Training, Simulation and Education Conference (I/ITSEC), Orlando, Fl. Dec 2013.

[Muller 2010] Muller, P. The Future Immersive Training Environment (FITE) JCTD: Improving Readiness Through Innovation. *Intraservice/Industry Training, Simulation & Education Conference*, 2010.

[Oskipe r2011] T. Oskiper, H. Chiu, Z. Zhu, S. Samarasekera, R. Kumar. Stable Vision-Aided Navigation for Large-Area Augmented Reality. *IEEE Virtual Reality*, March 2011.

[Reitmayr 2006] G. Reitmayr and T. Drummond. Going Out: Robust Model-based Tracking for Outdoor Augmented Reality. In *International Symposium on Mixed and Augmented Reality*, 2006.

[Report 2002] Report to the Chairman, Subcommittee on Readiness and Management Support, Committee on Armed Services, U. S. Senate. Military Training – Limitations Exist Overseas but Are Not Reflected in Readiness Reporting, *GAO Report GAO-02-525*, *17*, 2002.

[Saab 2010] http://saabtraining.com/PDF/PTD_3.pdf.

[Samarasekera 2014] Supun Samarasekera, Rakesh Kumar, Zhiwei Zhu, Vlad Branzoi, Nicholas Vitovitch, Ryan Villamil, Frank Dean, Pat Garrity. Live Augmented Reality based Weapon Training for Dismounts, *Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC)*, Orlando, FL, 2014.

[Schafer 2015] R. Schaffer, S. Cullen, L. Cerritelli, R. Kumar, S. Samarasekera, M. Sizintsev, T. Oskiper, V. Branzoi. Mobile Augmented Reality for Force-on-Force Training, *Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC)*, Orlando, FL, 2015.

[Yarnall 2015] Louise Yarnall, Sara Vasquez, Anna Werner, Rakesh (Teddy) Kumar, Supun Samarasekera, Girish Acharya, Glenn Murray, Michael Wolverton, Zhiwei Zhu, Vlad Branzoi, Nicholas Vitovitch, & Jim Carpenter. Human Performance in Content Design for Interactive Augmented Reality Systems, *Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC)*, Orlando, FL, 2015.

[Zhu 2008] Z. Zhu, T. Oskiper, S. Samarasekera, R. Kumar, and H. S. Sawhney. Real-Time Global Localization with A Pre-built Visual Landmark Database. In *IEEE Conference on Computer Vision and Pattern Recognition, CVPR*, 2008.