Augmented Reality Training for Forward Observers

Rakesh Kumar, Supun Samarasekera, Mikhail Sizintsev, Taragay Oskiper, Vlad Branzoi

SRI International

Princeton, NJ

{rakesh.kumar, supun.samarasekera, mikhail.szinitsev, taragay.oskiper,vlad.branzoi}@sri.com

Richard Schaffer, Sean Cullen, Nikhil Krishnaswamy

Lockheed Martin Corporation

Burlington, MA

{richard.l.schaffer, sean.cullen, nikhil.krishnaswamy}@lmco.com

ABSTRACT

Observers of all types are major force multipliers on the modern battlefield. However, with today's constrained budgets, the range time and training resources to support observer training, including munitions, supporting artillery and mortar units and aircraft sorties, are increasingly scarce. Augmented reality is an innovative technology that can supplement live training to address the challenge of affordably training Forward Observer, Joint Forward Observer, Joint Tactical Air Control and similar skills. We present the system design, hardware, software, algorithms and initial field results for a prototype augmented reality training system.

As part of this research, augmented reality devices for the unaided eye, binoculars, laser rangefinders and designators were developed. The augmented reality devices interface with real tactical equipment including the Defense Advanced GPS Receiver (DAGR) and the StrikeLink digital CAS system. The augmented reality system enables long range high precision live augmentation of both unaided eye and magnified imagery with aerial and terrain based synthetic objects, vehicles, people and effects. The inserted objects appear stable in the augmented reality device as the user pans through the battle-space. We present the navigation algorithms that use cameras in combination with an IMU and GPS to provide jitter free, robust and real-time 6-DOF pose estimation for precise augmentation.

We also present details of the rendering and simulation modules. We have developed a Unity based augmented reality rendering system that is part of an HLA federation that includes the USMC Deployable Virtual Training Environment, JSAF and an tablet-based Instructor Tablet. We will present the initial results of the system operating on an USMC range in which augmented fixed and rotary wing aircraft, ground vehicles, and weapon effects are combined with real world scenes.

ABOUT THE AUTHORS

Dr. Rakesh "Teddy" Kumar is currently the Director of the Center for Vision Technologies at SRI International, Princeton, New Jersey. Prior to joining SRI International Sarnoff, 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. Rakesh 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 received the University of Massachusetts Amherst School of Computer Science, Outstanding Achievement and Advocacy Award for Technology Development (2013). 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 (NSF) review panels. He has co-authored more than 50 research publications and has received over 50 patents.

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 Corp. Supun Samarasekera has 15+ years' experience in building integrated multi-sensor systems for training, security & other applications. He has led programs for robotics, 3D modeling, training, visualization, aerial video surveillance, multi-sensor tracking and medical image processing applications. He has received a number of technical achievement awards for his technical work at SRI.

Dr. Mikhail Sizintsev is currently a Computer Scientist at SRI International, Princeton. He received his Ph.D. degree in Computer Science from York University, Toronto, Canada under Prof. Richard P. Wildes. His doctoral thesis received the Canadian Image Processing and Pattern Recognition Society award in 2012. Mikhail Sizintsev has extensive experience in designing and implementing real-time stereo and motion estimation algorithms, developed visual navigation solutions and participated in numerous augmented reality projects.

Dr. Taragay Oskiper is currently a Senior Principal Research Scientist at SRI International, Princeton. He received his Ph.D in Electrical Engineering from Princeton University. He has over nine years' experience in developing vision-aided motion estimation and multi-sensor fusion algorithms for navigation and augmented reality for both video-see-through and optical-see-through platforms. He has acted as the lead algorithm developer for numerous augmented reality projects, most recently the Office of Naval Research AITT program, at Sarnoff and now SRI International Princeton.

Mr. Vlad Branzoi is currently a Computer Scientist at SRI International. He received his M.S. in Computer Science from Columbia University under Prof. Shree Nayar. Vlad Branzoi has over 10 years' experience in building novel sensors, integrated multi-sensor systems for training, robotics and mobile applications.

Mr. Richard Schaffer is currently a Lockheed Martin Fellow and Principal Investigator at Lockheed Martin Mission Systems and Training (MST). He leads the Human Immersive Simulation Lab at MST's Advanced Simulation Centers. Richard received his S.B. degree from the Massachusetts Institute of Technology and has over 28 years of experience in modeling and simulation research and development. His areas of research have included distributed simulation, environment modeling, and immersive simulation. In 2010 he received the NTSA's lifetime achievement award.

Mr. Sean Cullen is currently a Sr. Staff Software Engineer at Lockheed Martin MST. He received his B.S. in Computer Science from Middle Tennessee State University. Sean Cullen has 15+ years' experience in military based modeling and simulation. He has been the Project Engineer on multiple augmented reality programs at Lockheed Martin and has extensive experience in 3D graphics.

Mr. Nikhil Krishnaswamy is currently a Software Engineer Associate at Lockheed Martin MST, with 3 years of video game and simulation development experience. He received a B.S. in Computer Game Development from DePaul University and a Master's degree in Computational Linguistics from Brandeis University, where he is currently pursuing his Ph.D. in Computer Science.

Augmented Reality Training for Forward Observers

Rakesh Kumar, Supun Samarasekera, Mikhail Sizintsev, Taragay Oskiper, Vlad Branzoi

SRI International

Princeton, NJ

{rakesh.kumar, supun.samarasekera, mikhail.sizintsev, taragay.oskiper,vlad.branzoi}@sri.com

Richard Schaffer, Sean Cullen, Nikhil Krishnaswamy

Lockheed Martin Corporation

Burlington, MA

{richard.l.schaffer, sean.cullen, nikhil.krishnaswamy}@lmco.com

INTRODUCTION

Observers of all types are major force multipliers on the modern battlefield. However, with today's constrained budgets, the range time and training resources to support observer training, including munitions, supporting artillery and mortar units and aircraft sorties, are increasingly scarce. Augmented reality is an innovative technology that can supplement live training to address the challenge of affordably training Forward Observer, Joint Forward Observer, Joint Tactical Air Control and similar skills. We present the system design, hardware, software, algorithms and initial field results for a prototype augmented reality training system.

As part of this research, augmented reality devices were developed for the unaided eye, binoculars, laser rangefinders and designators. The augmented reality devices interface with real tactical equipment including the Defense Advanced GPS Receiver (DAGR) and the StrikeLink digital CAS system. The augmented reality system enables long range high precision live augmentation of both unaided eye and magnified imagery with aerial and terrain based synthetic objects, vehicles, people and effects. The inserted objects appear stable in the augmented reality device as the user pans through the battle-space. We present the navigation algorithms that use cameras in combination with an IMU and GPS to provide jitter free, robust and real-time 6-DOF pose estimation for precise augmentation. We also present details of the rendering and simulation modules. A Unity based augmented reality rendering system was developed and is part of an HLA federation that includes the USMC Deployable Virtual Training Environment, JSAF and an Apple iPad-based Instructor Tablet. We will present initial results of the system operating on an USMC range in which augmented fixed and rotary wing aircraft, ground vehicles, and weapon effects are combined with real world scenes. Finally we present the conclusions, acknowledgements and references.

PREVIOUS WORK

Most dismounted training systems today rely on physical targets to represent opposing forces during exercises, or human actors with laser weapons and laser detectors to determine when someone is hit by weapons fire. Systems using physical targets (e.g., paper pop up silhouettes) lack realism, as targets react in predictable ways to the actions of trainees. Systems using human actors 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 [Hsu, 2009] have been deployed at Camp Pendleton and other Marine Corp's MOUTs (Military Operations on Urban Terrain). These systems use video projectors to project images of virtual actors on walls of rooms within a training facility. 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 which can track Marines both indoors and outdoors. Global Positioning System (GPS) based systems may be used for providing location outdoors. However, the performance of these outdoor-only systems decreases in challenging GPS limited situations. Ultra-wideband (UWB) based systems have been used for indoor tracking of trainees to foot (30 cm) level accuracies [Fontana, 2002] but do not provide orientation information. Finally none of these systems meet the challenging requirement for augmented reality [Kato, 1999, Reitmayr, 2006] where both location and orientation of the users head must be tracked to cm level accuracy and less than 0.01 deg. accuracy for orientation. Overall, providing high accuracy tracking over large indoor and outdoor areas (multiple square miles) is a very challenging problem.

OVERALL APPROACH

We present an augmented reality based system for training forward observers. The major architectural elements of this training system are an augmented human system interface wirelessly connected to an instructor station. An additional, unusual element of the training system is the real world itself. As an augmented reality system, all components must to be capable of accurately incorporating the real world environment in which they operate, principally comprising terrain, natural lighting and weather. The augmented human system interface is comprised of a head-worn display, sensors, and commercial computer hardware running sophisticated navigation and rendering software. In addition, the trainee may use a number of augmented reality props. By an augmented reality prop we mean a device that functionally closely resembles real piece of military gear, such as a pair of binoculars and Portable Lightweight Designator Rangefinder (PLDR), but which are modified to portray the augmented world. The instructor station provides a means for the instructor to configure, monitor and control the training as well as a means for a role player to participate in training exercise.

Figure 1 shows a high level diagram of the overall augmented reality training system. The system is configured with multiple trainee-worn sensing, display and computational packages that connected over a wireless network. The trainee-worn sensors include monocular camera, IMU (Inertial Measurement Unit) and GPS devices. The trainee-worn display is eyewear or an HMD (Head Mounted Display). The computational package is a commercial Intel i7-based laptop. It includes computational modules for trainee 6- Degree of Freedom (6-DOF) pose estimation which interacts with the landmark database modules. The computational package also includes the rendering engine which interacts with the world model and local avatar interaction and weapon effects modules. We describe each of these modules in more detail in the subsequent sections.

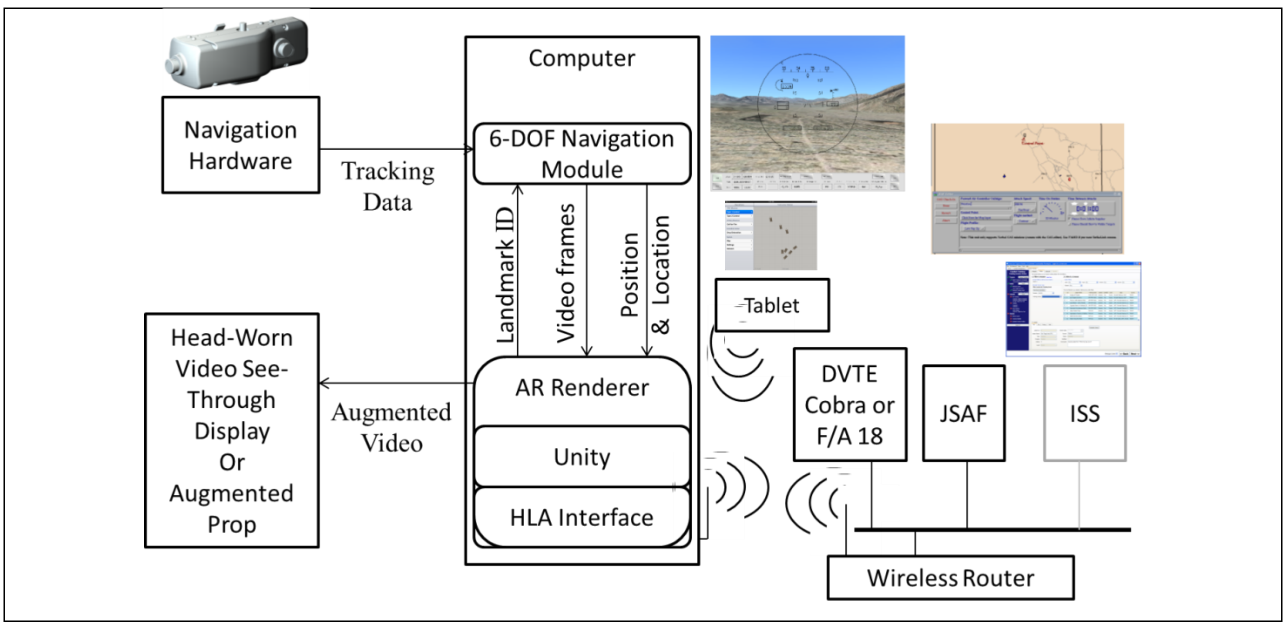

Figure 1. Overall Augmented Reality Training System Diagram.

AR HARDWARE and PROPS

Besides their normal eye vision, forward observers use multiple pieces of equipment in order to perform their missions. To conduct training in an augmented environment the trainee must be able to use this equipment in the same manner as he would be able to in the field and also be able to view the virtual entities while using them. In the

AR training system this was accomplished by creating props that had the same form, fit, and function as the real equipment but that have been built to contain the hardware described in the AR hardware section. The initial hardware chosen to replicate consists of a Vector 21B laser range finder/binocular system and the Portable Lightweight Designator Rangefinder (PLDR) system. It is important that the hardware the Warfighter wears has reasonable size, weight and power (SWaP) to enable training. There need to be helmet mounted light-weight sensor package and reliable processors unit that has (a) low weight, (b) supports low-latency processing needs and (c) has a sufficient power to support realistic training scenarios. We describe the augmented reality hardware that was built for the unaided eye, binoculars and PLDR, as depicted on Figure 2.

The compact sensor head for “unaided eye” augmentation (Figure 2-a) is made of USB camera and an X-Sense MTI-G unit. The MTI-G has an IMU, magnetometer, GPS and barometer. The sensors are instrumented such that the MTI-G unit triggers the cameras at a programmable interval. As such the system is designed to provide precise time synchronization between all the sensors. The back of the sensor package has an adaptable HMD attachment. For the experiments, both Vuzix and Intevac i-Port 75 HMDs were integrated with the sensor package.

The Vector 21B is an optical binocular with an embedded laser rangefinder, heading tracker, and azimuth tracker. The real device is capable of providing ranges to lased points, determining the distance between two different targeted points, and it can provide add/drop calculations for adjusting artillery fire among other things. The prop Vector 21B replaces the optical path with a video camera and embedded display with the augmented rendering added. The binoculars contains an X-Sens MTI-G unit, and have two cameras: a wide (80 degrees) field of view monochrome GiG-E camera and a narrow (6.3 or 4.75 degrees) field of view color GiG-E camera. For visualization, integrated Vuzix and Intevac i-Port 75 HMDs are used.

The AR Vector 21 prop (Figure 2-b) is capable of all the same complex operations as the real equipment. This includes the ability to be connected to a military GPS device. The real equipment allows a DAGR or PLGR to be connected via a special connector. Then the user can target a location and trigger a signal to be sent to the GPS device. The GPS uses its current location and the range, azimuth, and elevation angle information sent from the Vector 21B to calculate an accurate location for the targeted position. The prop supports the use of the same GPS device with the same connector cable and likewise can calculate a location but instead of firing a laser it uses a stored 3D map of the training area that the rendering system uses for occlusion reasoning along with the data from the tracking system. Additionally, using the same connector cable the prop Vector 21B can also be connected to a military StrikeLink™ computer allowing the target locations to be injected in the same manner as trainees would do in the field with the real equipment. For this effort, two versions of the Vector 21B device were built with the second device used to simulate the Vector 21B with the 10x optical zoom adapter installed.

The other prop built was the PLDR (Figure 2-C). In the field, this equipment is used to fire a powerful laser designator for guiding specialized munitions. The Augmented Reality PLDR has the same sensor package as the AR-binoculars. The display for the PLDR uses a monocular eyepiece with a small LCD screen. The AR prop PLDR contains much of the functionality of the actual PLDR and allows the operator to set laser codes and control the firing of the laser in the same manner as the actual device. The benefit of this device is that instead of dangerous laser energy going downrange the prop send information out on the network so any virtual entities that can detect the energy know where the device is pointed and when it is emitting. The AR prop PLDR was designed to sit on the same tripod as the real device and even contains a small speaker to emulate the sound the real device makes when the laser is active. The actual device contains a small camera that feeds a secondary LCD display. The LCD display is also used to show the user the menu for configuring the device. The AR prop PLDR implemented the menu portion of this display but skipped the video display as feedback from end users was that this functionality was rarely used.

The processor package consists of a PC-104 quad-core Intel i-7 processor board that computes localization solution and performs augmented reality rendering using an integrated graphics card. The computation unit (Figure 2-d) is packaged in a 10”x7”x3.75” rugged enclosure and utilizes approximately 90 Watts. The system is designed to work with two hot-swappable battery packs to accommodate continuous training and minimize downtime during operations in the field.

Figure 2. (a) Unaided Eye Sensor Head, (b) Binoculars (c) PLDR (d) Compact Processor.

Navigation Module

Figure 3 shows the core modules that make up the Navigation module for tracking in all devices.

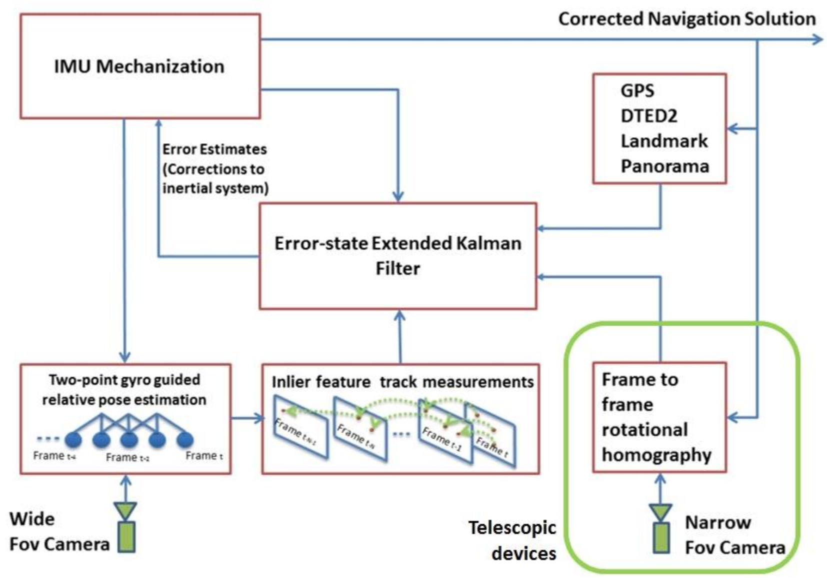

Figure 3. Navigation Subsystem Modules. Service module for Narrow field of view (FoV) cameras (framed in green) is specific to telescopic systems only (binoculars and PLDR).

Two types of augmented reality devices were developed as part of our research: “unaided-eye” helmet-mount devices and “telescopic” high-zoom devices in the form of binoculars and PLDR. Their core implementation modules are depicted in Figure 3 and are largely similar except telescopic devices have narrow field of view (FoV) cameras for zoomed-in augmentation and precise localization.

Precise navigation is key to an augmented reality system. The inserted objects must not jitter or drift as the user pans or moves around while viewing the scene with the augmented reality devices. The inserted objects must also appear at the same physical location when viewed from different locations and with different devices. From previous experiments, we have observed that 1 pixel in-accuracies in insertion are noticeable to humans and result in jitter and drift as the user moves around. Table 1 shows the precision required in estimating the orientation component of navigation state, based on less than 1 pixel of insertion accuracy. The insertions also must be done with very low latency (order of 5-10 milli-seconds) as the user may move his head and hands very rapidly.

Central to the localization solution is IMU centric extended Kalman filter. The Kalman filter uses the 3-axis accelerometer and Gyro to derive an IMU mechanization state and evaluates “Error-States” of other sensors (e.g. video, GPS) with respect to this mechanization. Video-based reasoning provides the high-fidelity localization to the solution. GPS provides initial location in global coordinates and diminishes the long-term drift in localization. Real time visual odometry is used within the Kalman filter for providing high-quality relative pose inputs. The visual landmark and panorama matching module enables low-latency accurate drift corrections in orientation.

Table 1: Navigation Orientation Accuracy required for insertions to have less than 1 pixel of jitter.

Device	Field of View	Video Resolution	Navigation Orientation Accuracy
Unaided eye	80 deg.	640x480	0.125 deg./ frame/
Unaided eye	80 deg.	1024x960	0.078 deg./ frame
Binoculars	6.75 deg.	640x480	0.011 deg./ frame
Binoculars	6.75 deg.	1024x960	0.005 deg./ frame
Binoculars	4.725 deg.	640x480	0.007 deg./ frame
Binoculars	4.725 deg.	1024x960	0.004 deg./ frame

IMU Mechanization and Extended Kalman Filter

The navigation module employs an IMU-centric error-state Extended Kalman Filter (EKF) approach (Figure 3) to fuse IMU measurements with external sensor measurements that can be local (relative), such as those provided by visual odometry, or global, such as those provided by GPS and landmark matching. This filter replaces the system dynamics with a motion model derived from the IMU mechanism. The filter dynamics follow from the IMU error propagation equations, which evolve smoothly and therefore are more amenable to linearization. This allows for better handling of the uncertainty propagation through the whole system. The measurements to the filter consist of the differences between the inertial navigation solution as obtained by solving the IMU mechanization equations and the external source data. The final filter estimate can automatically remove spurious measurements from external sensors, such as visual odometry when vision fails. Global position measurements are provided by GPS and are fed directly to error-state EKF filter. Global orientation measurements are provided by matching the current image to a panorama landmark database. Given a query image, landmark matching returns the found landmark shot from the database. These correspondences are applied as measurement equations in the error-states of the error-state EKF filter. The landmark database is built and online matched in real-time to enable the global constraints and promptly recover from the drift that occurs after rapid movements or interrupted visual feed (e.g. when binoculars are pointed at the ground).

Relative positioning from Visual Odometry

Simple image features are first extracted from each video frame using Harris corner detection algorithm. These features are correlated and tracked over time, and their motion relative to one another is measured. Each feature track, lasts as long as it is matched in the new frames that are acquired. As old tracks vanish, new ones are established with features that have not been previously observed. We use geometry checks to remove outliers from these feature tracks (Figure 4). From these 2D image features and corresponding 3D scene points the system calculates a precise estimate of the cameras' location and pose, in 6-DOF.

In case of telescopic optics, as the video frames from both narrow and wide FoV cameras are received, feature extraction and frame-to-frame matching is performed simultaneously in both cameras. Due to its very narrow field of view, the zoomed-in camera is able to capture only small motions and, in such cases, visual tracking is performed using the wide FoV camera only, as depicted in Figure 5. Furthermore, narrow FoV tracks are only used to correct the 3-DOF estimation of the orientation. The error-state EKF combines these solutions with rotation and acceleration readings from the IMU [Oskiper, 2012]. When fused with GPS, the result is a geo-localized position and orientation. We have observed experimentally that the visual odometry has a location drift rate of 0.1% of distance traveled and average orientation error less than 0.01 degrees.

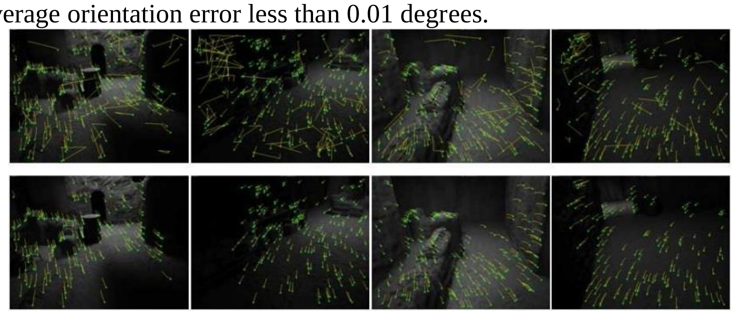

Figure 4. Raw Feature Tracks over nearby frames (Top row). Note there are tracks with wrong directions . Feature tracks that pass geometry checks (Bottom row).

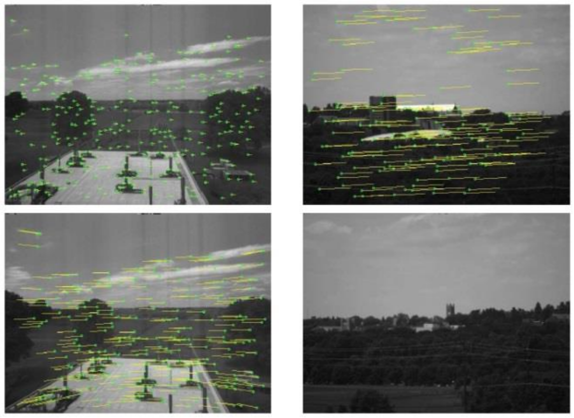

Figure 5. Inlier feature tracks obtained from wide (Left column) and narrow (Right column) FoV cameras. Small camera motion allows successful feature tracking on both cameras (Top row), while faster motions are captured by wide FoV camera only (Bottom row).

Drift Correction via Landmark Databases

In order to determine the global orientation of the device after startup, a landmark based initialization procedure similar to [17] is used, where a 3D to 2D tie-point is established between an image frame and a world location. It is assumed geodetic coordinates of the initial location and the coordinates for an easily discernible landmark in the scene are known. In case of telescopic AR, the narrow FoV camera image is used for highest angular accuracy. After initialization is performed, the video frames form wide FoV are collected to build an online panorama mosaic database as depicted in Figure 6. A new image is added to the database if its overlap with current mosaic is less than a certain threshold. Use of the extendable panorama database allows for global orientation estimation updates virtually for every video frame that results in drift-free tracking.

Matching of the query to the panorama database mosaic is performed using Harris corner interest points and Scale-Invariant Feature Transform (SIFT)/ Histogram of Oriented Gradients (HOG) [SE 2006, Zhu 2008] descriptors, as depicted in Figure 7. Importantly we only consider interest points near the horizon band (inferred from roll and pitch orientation estimates of the filter) because the sky pattern above the horizon line tends to continuously change over time, while close objects that normally reside below the horizon line can cause large visual parallax while moving laterally.

AR RENDERING

The rendering and simulation engine utilizes the Unity3D game development platform as the base platform. The choice of using Unity as a base platform stems from several major benefits of the system (overall far exceeding open source alternatives). It provides low level access to underlying subsystems that are critical for the development of AR simulations (unlike systems such as VBS2 or RealWorld). It is a cutting edge high fidelity game engine with continuous active development and improvement, as well as a huge established development community. It provides a rendering engine, physics integration, spatial audio, animation system, network messaging framework, Graphical User Interaction (GUI) elements, and an integrated development environment with a configurable scene editor. It supports deployment on multiple platforms such as PC, Mac, iOS, Android, and Web browsers. It also provides a highly reasonable licensing model, as opposed to the often massive licensing costs associated with other commercial engines.

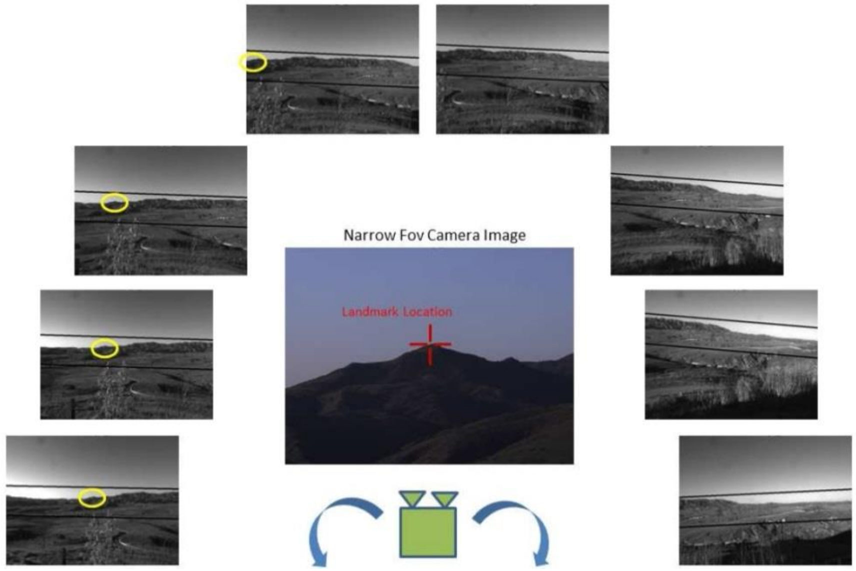

Figure 6. System initialization procedure is followed by online panorama generation. After initial geo-located landmark is clicked, and device pans across the scene during the exercise, system creates panoramic database of the area of interest covered up to that moment. Yellow oval shows the location of initial landmark and black lines show the horizon band.

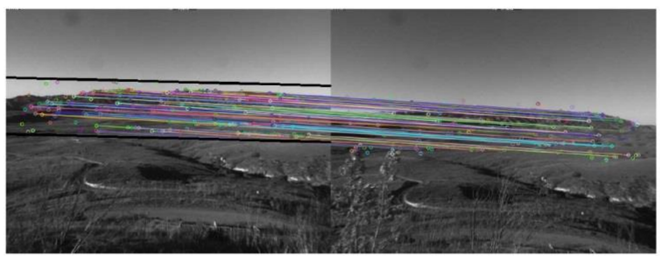

Figure 7. Inlier matches between panorama database image and query image. Horizon bands are shown by black lines

SIMULATION AND SCENARIO GENERATION

The ability to augment reality is only part of the system. In order to conduct observer training we must be able to augment reality with the appropriate entities and battlefield effects. These entities must be easily controlled with the right types of behaviors. In order to accomplish this we use a federated approach that puts the AR Rendering software on a wireless simulation network allowing interoperation with standard simulation products like the Deployable Virtual Training Environment's (DVTE) Combined Arms Network (CAN) and Joint Semi-Automated Forces (JSAF). In addition, we have developed a portable instructor tablet based app that connects to the simulation network for monitoring system status as well as controlling scenarios.

The DVTE CAN software allows users to conduct training in a pure virtual environment. This training system added a terrain database to the DVTE CAN software that allowed users to perform this training in the same location as the augmented reality trainees. By having the DVTE CAN software join the same federation as the AR training system, instructors can now pilot virtual aircraft that can be observed in the augmented reality environment. The CAN software allows in-depth simulation of many military vehicles to include helicopters with laser guided munitions. With the AR training system PLDR hardware, a simulated laser designator can be employed by the forward observer that is detectable by the virtual aircraft and which is then able to fire simulated munitions at the target with detonation effects visible by both parties. Now by putting the trainee in the AR system on a radio channel with the instructor flying the CAN helicopter, end-to-end JTAC/JFO missions can be accomplished without the need for live aircraft and active ranges. In addition to DVTE CAN, the AR training system can connect to semi-automated force simulators like JSAF.

In order to easily monitor the AR system's operational status as well as provide instructors scenario control, we developed a tablet (iPad) based application capable of joining the federation. Each instrumented prop and piece of support equipment publishes a status object that sends performance statistics to the tablet. This can be very helpful during training to allow the instructor to verify that each piece of equipment is ready before starting a training session. The iPad application serves another critical function in that it allows the instructor to insert entities and battlefield effects into the scenario. The app was designed to load a military map overlay that also displays any friendly or enemy vehicles on the screen. As required, the instructor can use a simple point and click interface to set off detonations when and where the instructor desires. This type of quick injection of effects allows the instructor to adjust the complexity of the scenario based on the ability of the trainee. The app was also designed to allow an instructor to setup and control fixed-wing bombing missions. The app does this by sending control messages to the JSAF simulation that triggers the appropriate entity behaviors on the aircraft. This high level of control allows the instructor to control a flight of two F/A-18 Hornets but without consuming so much of his time that he would miss errors the trainee might be making.

TRAINING AREA MAPPING AND MODELING

Augmented reality requires building an accurate 3D model of the site that place the players in a common coordinate frame that is shared with the 3D model. Currently, the 3D model of the scenario location is generated from air-borne LIDAR measurements to match the real terrain as close as possible. All vehicles and effects are rendered and controlled based on the terrain profile.

EXPERIMENTAL RESULTS

In this section, we report on a number of experiments aimed at evaluating different aspects of performance of the augmented reality based training system.

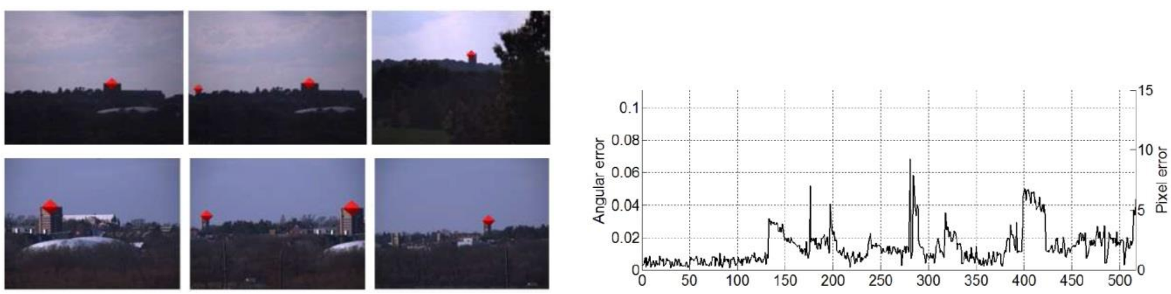

Figure 8. Augmentation of buildings in hand-annotated sequence (Left) and corresponding angular error of estimated poses in degrees and pixels in narrow FoV image (Right).

Pose Estimation for Augmented Reality

To demonstrate accurate and robust tracking for augmented reality, we conducted a number of qualitative and quantitative experiments. The "unaided eye" helmet system has been extensively evaluated previously in [Oskiper, 2012]. The unaided eye system exhibits average global positional error under 0.5 meter and orientation accuracy of less than 0.1 degrees, which are suitable for high-quality compelling AR augmentation. The current work puts more emphasis on evaluating telescopic augmented reality devices, which have a higher requirement for tracking precision. Example of AR augmentation of telescopic system (7x and 10x binoculars) is depicted in Figure 10, where tall distinct towers are labeled by red markers. The geo-located markers are been directly selected in Google Earth and the tabs appear at the correct spots when the buildings are in the binoculars view. Also, tags do not jump from one another, jitter or drift as the device moves around. Furthermore, several similar sequences have been hand-labeled for the true location of the towers in the image and angular errors have been analyzed (c.f. Figure 8). The average AR insertion error of only few pixels allows for very compelling AR perception when synthetic entities appear in correct physical locations, do not jitter and blend with the real environment.

Training scenarios examples

To test the system, we conducted several augmented reality combined arms scenarios where forward observers directed simulated artillery and aircraft fire. The exercises were conducted at two independent locations – the roof of a building on the US east coast and mountainous terrain in the US west coast. 3D terrain models were constructed from air-borne LIDAR data. Example renderings of virtual vehicle avatars positioned on a 3D model registered to the terrain and attacked by a virtual helicopter are depicted in Figure 9.

Figure 9. Augmented reality insertion example for AR binoculars. Scene is augmented with helicopter attacking tanks. Tanks and helicopters cast shadows.

Actual scenario examples for both locations are depicted in Figure 10 (unaided eye together with 7-x binoculars) and Figure 11 (all devices: unaided eye, 7x binoculars, 10x binoculars, PLDR). Position of all vehicles and effects match very well in both environments when vehicles are moving or stationary. No distracting jitter is visible in any augmented views. Virtual shadows cast by avatars significantly improve the perception of them being part of the scene. Furthermore, inserted locations are consistent across multiple devices used simultaneously, which allows to reliably detect and track a virtual entity by multiple trainees in a cooperative fashion. All these demonstrate that the system can provide compelling augmentation over extended time period and over multiple devices using underlying drift and jitter-free pose mechanisms.

Figure 10. Augmented reality insertion examples for unaided eye (top row) and binoculars (bottom row).

CONCLUSIONS

We presented technical modules and experimental results from an infrastructure free augmented reality system for training forward observers. The augmented reality system can be used for *live* training of forward observers at home stations or deployed locations. The system enables the forward observers to train live outdoors without requiring real aircraft to be flown, real targets to be placed or real munitions to be dropped. Synthetic targets, vehicles, avatars for opposing and friendly forces and civilians are automatically rendered by a simulation engine onto the Head Mounted Display (HMD) and other augmented reality props used by the trainees. The rendered avatars, aircraft, vehicles, targets and effects appear as if they are part of the live scene. The simulation engine is used to control the behavior of the synthetic elements. Scenarios with different levels of difficulty can be generated for different exercises and adaptive training.

Finally the augmented reality props mimic the real devices used by forward observers, so no special training is required to use the augmented reality devices. The system was tested at Camp Pendleton in Feb 2013. Almost no coaching was necessary when warfighters used the system to execute several scenarios. Though not involved in the setup and maintenance of the system, warfighters reported that the system was easy to use with no training, an effective simulation of scenarios and would be a great piece of future training equipment.

Figure 11. Augmented tanks and explosion insertion examples. From left to right: unaided eye, 7x binoculars, 10x binoculars, PLDR. Same tank are highlighted by yellow and orange oval across several views.

ACKNOWLEDGEMENTS

This material is based upon work supported by the Office of Naval Research Contracts N00014-11-C-0433 and N00014-11-C-0064. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the US government/ Department of Defense.

REFERENCES

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

Eduardo Gudis, Gooitzen van der Wal, Sujit Kuthirummal, Sek Chai, Supun Samarasekera, Rakesh Kumar and Vlad Branzoi, Stereo Vision Embedded System for Augmented Reality *IEEE CVPR Embedded Vision Workshop*, Providence, RI, June 2012

Robert J. Fontana and Steven J. Gunderson, Ultra-Wideband Precision Asset Location System, *2002 IEEE Conference on Ultra Wideband Systems and Technologies*, Baltimore, MD, May 2002.

E. Foxlin and L. Naimark. Vis-tracker:a wearable vision-inertial selftracker. In *IEEE Virtual Reality*, 2003.

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.

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

T. Oskiper, H. Chiu, Z. Zhu, S. Samarasekera, R. Kumar, "Multi-Modal Sensor Fusion Algorithm for Ubiquitous Infrastructure-free Localization in Vision-impaired Environments", *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS*), October 2010.

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

T. Oskiper, S. Samarasekera, and R. Kumar. "Multi-sensor navigation algorithm using monocular camera, IMU and GPS for large scale augmented reality", *IEEE International Symposium on Mixed and Augmented Reality (ISMAR), 2012*

[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.

S. Se, D. Lowe, and J. Little. Vision-based global localization and mapping for mobile robots. *IEEE Transactions on Robotics*, 21(3), 2006.

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*