Implementation of an Augmented Reality System for Training Dismounted Warfighters

Rakesh Kumar, S. Samarasekera, A. Chaudhry, Zhiwei Zhu, Han-Pang Chiu, Taragay Oskiper, Ryan Villamil, Vlad Branzoi, Raia Hadsell	Eugene Ray Pursel	Frank Dean, Pat Garrity
SRI International Sarnoff	Marine Corps Warfighting Laboratory	U.S. Army Research Laboratory's Simulation and Training Technology
Princeton, NJ	Quantico, VA	Orlando, FL
{rakesh.kumar, supun.samarasekera, ali.chaudhry, zhiwei.zhu, han-pang.chiu, taragay.oskiper, vlad.brnazoi, ryan.villamil, raia.hadsell}@sri.com	eugene.pursel@usmc.mil	frank.dean@us.army.mil, pat.garrity@us.army.mil

ABSTRACT

There is a need within the military to enhance its training capability to provide more realistic and timely training, but without incurring excessive costs in time and infrastructure. This is especially true in preparing for urban combat. Unfortunately the creation of facility based training centers that provide sufficient realism is time consuming and costly. Many supporting actors are needed to provide opponent forces and civilians. Elaborate infrastructure is needed to create a range of training scenarios, and record and review training sessions. In this paper we describe the technical methods and experimental results on building an Augmented Reality Training system for training dismounts doing maneuver operations that addresses the above shortcomings. The augmented reality system uses computer graphics and special head mounted displays to insert virtual actors and objects into the scene as viewed by each trainee wearing augmented reality eyewear. The virtual actors respond in realistic ways to actions of the Warfighters, taking cover, firing back, or milling as crowds.

Perhaps most importantly, the system is designed to be infrastructure free. The primary hardware needed to implement augmented reality is worn by the individual trainees. The system worn by a trainee includes helmet mounted sensors, see through eye-wear, and a compact computer in his backpack. The augmented reality system tracks the actions, locations and head and weapon poses of each trainee in detail so the system can appropriately position virtual objects in his field of view. Synthetic actors, objects and effects are rendered by a game engine on the eyewear display. Stereo based 3D reasoning is used to occlude all or parts of synthetic entities obscured by real world three dimensional structures based on the location of the synthetic. We present implementation details for each of the modules and experimental results for both day time and night time operations.

ABOUT THE AUTHORS

Dr. Rakesh "Teddy" Kumar is currently the Senior Technical Director of the Vision and Robotics Laboratory at SRI International Sarnoff, Princeton, New Jersey. 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. 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 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 35 patents.

Mr. Supun Samarasekera is currently the Senior Technical Manager of the Mobile Vision Group at SRI International Sarnoff. 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 number technical achievement awards for his technical work at SRI.

Mr. Ryan Villamil is a Senior Computer Scientist at SRI. He has a MS in Computer Science from Columbia University. He has over 12 years' experience in computer graphics, computer vision, multi-player networking, simulation and training systems development. He has developed and continues to develop high performance rendering, visualization and gaming systems for a number of mixed and augmented reality applications.

Dr. Zhiwei Zhu is a Senior Computer Scientist at SRI International. He has a 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, and has received one Best Transaction Paper Award from IEEE Transactions on Vehicular Technology in 2004 for his driver fatigue monitoring work and 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.

Mr. Vlad Branzoi is currently a Computer Scientist at SRI International Sarnoff. 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. Eugene Ray Pursel served over 23 years as a Marine in both enlisted and officer roles. His billets ranged from Aviations Operations Clerk to Helicopter Section Leader to Modeling and Simulations Officer. He earned a B.S. in Computer Science and Mathematics Minor from the Pennsylvania State University in 1995 and an M.S. in Modeling, Virtual Environments and Simulation from the Naval Postgraduate School in 2004. Now retired from active duty, he is serving as a Modeling and Simulations Analyst with the Marine Corps Warfighting Laboratory.

Mr. Frank Dean is an engineer and a science and technology manager at the U.S. Army Research Laboratory, Simulation and Training Technology Center (ARL-STTC), Orlando, FL. He currently works in the Ground Simulation Environments Branch conducting research and development in the area of dismounted soldier training and 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 Product Manager Intelligence and Electronic Warfare (PM-IEW)/ Reconnaissance, Surveillance, and Target Acquisition (PM-RSTA), PM Army Air Traffic Control, and Simulation Training and Instrumentation Command's (STRICOM) 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 from George Washington University.

Mr. Pat Garrity is the Chief Engineer for Dismounted Soldier Technologies at U.S. Army Research Laboratory's Simulation and Training Technology Center (ARL-STTC). He currently works in Ground Simulation Environments Division conducting research and development in the area of dismounted soldier training and simulation where he was the Army's Science and Technology Objective Manager for the Embedded Training for Dismounted Soldiers program. Prior to his involvement at ARL, he worked as the Project Director for the Advanced Concepts Research Tools program in Product Manager for Simulation Technology Distribution (PM-STI) at STRICOM. His current interests include Human-In-The-Loop 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.

Implementation of an Augmented Reality System for Training Dismounted Warfighters

Rakesh Kumar, S. Samarasekera, A. Chaudhry, Zhiwei Zhu, Han-Pang Chiu, Taragay Oskiper, Ryan Villamil1, Vlad Branzoi, Raia Hadsell	Eugene Ray Pursel	Frank Dean, Pat Garrity
SRI International Sarnoff	Marine Corps Warfighting Laboratory	U.S. Army Research Laboratory's Simulation and Training Technology Center
Princeton, NJ	Quantico, VA	Orlando, FL
{rakesh.kumar, supun.samarasekera, ali.chaudhry, zhiwei.zhu, han-pang.chiu, taragay.oskiper,vlad.brnazoi, ryan.villamil, raia.hadsell}@sri.com	eugene.pursel@usmc.mil	frank.dean@us.army.mil, pat.garrity@us.army.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 close to the battlefield. However, the 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 the 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.

In this paper we describe the technical methods and experimental results on building an Augmented Reality Training system for training dismounts doing maneuver operations that addresses the above needs in a cost effective manner. The augmented reality system is designed to be infrastructure free. The primary hardware needed to implement the solution is worn by the individual trainees. The system uses computer graphics and special head mounted displays to insert virtual actors and objects into the scene as viewed by each trainee wearing augmented reality eyewear. The virtual actors respond in realistic ways to actions of the Warfighters, such as taking cover, firing back, or milling as crowds using triggers and pre-scripted actions.

In order to achieve this, helmet mounted sensors are used to locate the Warfighter and his gaze with respect to the 3D environment. The Warfighter's 6-DOF head pose is 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. These avatars and objects are inserted into the live view on the eye-wear or Head Mounted Display (HMD) for real-time engagement with the Warfighter.

In the remainder of the paper, we first describe previous work and then present our overall approach, description of the technical modules and experimental results. The first technical module described is the Trainee Localization System which includes modules for IMU (Inertial Measurement Unit) mechanization and Kalman Filter, Integration of Visual Sensors, Relative Visual Odometry, Landmark Matching and Radio Frequency (RF) ranging. We then discuss modules for augmented reality display and avatar Interactions. These include modules for simulation and rendering, scenario generation, occlusion handling, and stereo based depth reasoning. Finally, we present the augmented reality hardware, experimental results, 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 [Saab, 2010] 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 0.05 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

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 dynamically configured wireless mesh network to each other and to a central After Action Review (AAR) server and a central game server. The only infrastructure required is the placement of the mesh network to ensure comprehensive coverage and the server systems for AAR and avatar interactions. The trainee-worn sensors include stereo cameras, IMU (Inertial Measurement Unit) and RF-ranging devices. The trainee-worn display is eyewear or an HMD (Head Mounted Display). The computational package is a processor worn by the trainee on his back-pack. It includes computational modules for trainee and weapon 6- Degree of Freedom (6-DOF) pose estimation which interacts with the ranging and landmark database modules. The computational package also includes the rendering engine which interacts with the depth reasoning/ world model and local avatar interaction modules. A Central Game Server directs the local avatar interactions. We describe each of these modules in more detail in the subsequent sections.

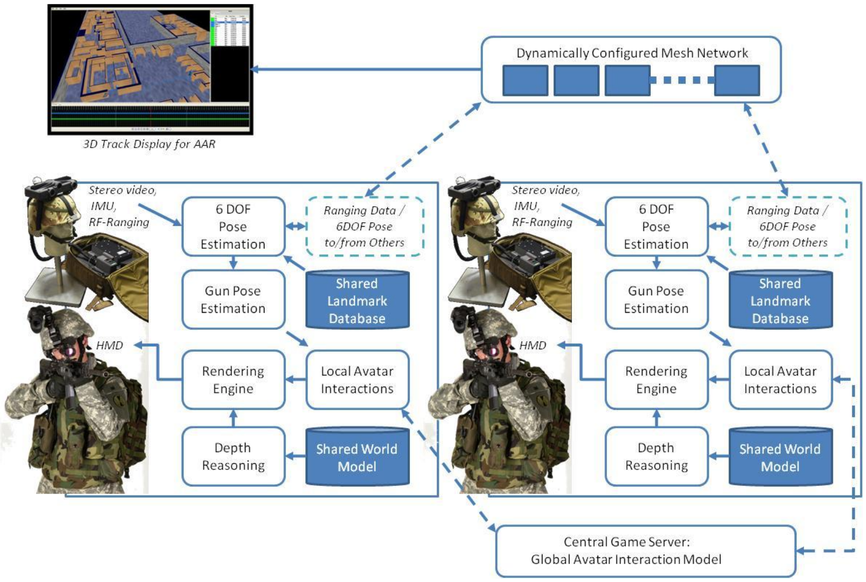


Figure 1. Overall Augmented Reality Training System Diagram.

TRAINEE LOCALIZATION SYSTEM

Figure 2 shows the core modules that make up the localization module.

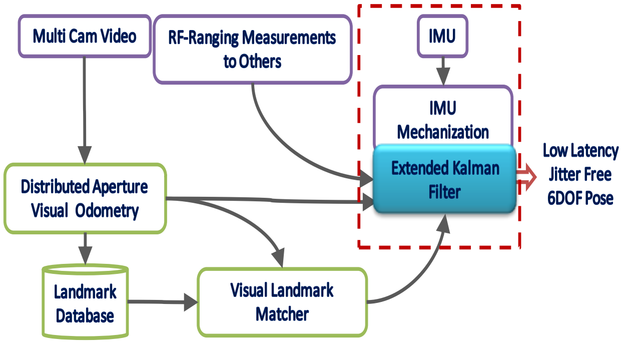

Multi Cam Video
RF-Ranging Measurements to Others
IMU
IMU Mechanization
Extended Kalman Filter
Low Latency Jitter Free 6DOF Pose
Distributed Aperture Visual Odometry
Landmark Database
Visual Landmark Matcher

Figure 2. Localization Subsystem Modules.

Central to the localization solution is IMU centric extended Kalman filter. We have developed a Kalman filter that uses the 3-DOF accelerometer and Gyro to derive an IMU mechanization state and evaluates "Error-States" of other sensors (e.g. video, RF ranging) with respect to this mechanization. Video-based reasoning provides the high-fidelity localization to our solution. We have developed real-time methods for doing visual odometry for providing high-quality relative pose inputs. A visual landmark matching module enables longer range drift (location and orientation) corrections.

IMU Mechanization and Extended Kalman Filter

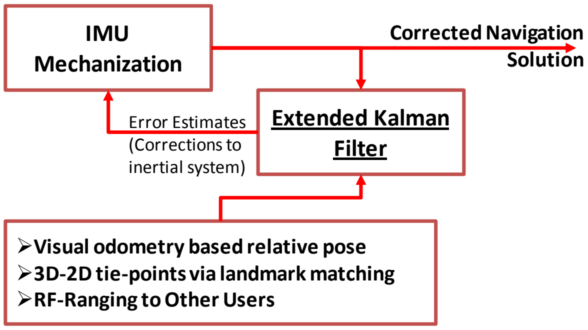

IMU Mechanization
Corrected Navigation Solution
Error Estimates (Corrections to inertial system)
Extended Kalman Filter
➢Visual odometry based relative pose
➢3D-2D tie-points via landmark matching
➢RF-Ranging to Other Users

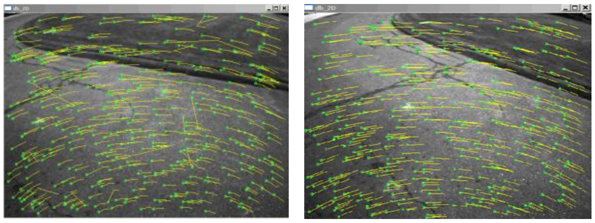
Figure 3. Error-state EKF and IMU mechanization.

We employ 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 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 measurements are provided by matching the current image to a landmark database. Given a query image, landmark matching returns the found landmark shot from the database. This match is used to establish 2D to 3D point correspondences between the query image features and the 3D world model. The 2D-3D correspondences are applied as measurement equations in the error-states of the error-state EKF filter. The landmark database is built offline and is matched in real-time to establish the global constraints.

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. Within our framework multiple video feeds can be integrated into the navigation solution. Tracks from individual video feeds are used as before but poses estimated are verified against the relative calibrations of the two cameras for consistency. Only the consistent hypothesis is fed to the EKF. This enables maintaining relative odometry while some of the cameras are occluded. The error-state EKF combines these solutions with rotation and acceleration readings from the IMU [Oskiper, 2011]. When fused with GPS (if available), the result is a geo-localized position and pose. We have observed experimentally that the visual odometry has a drift rate of 0.1% of distance traveled and the jitter estimates are under 1mrad/frame.



Raw feature tracks

Feature tracks that pass geometry checks

Figure 4. (Left) Raw features tracked over nearby frames. Note there are tracks with wrong directions. (Right) Feature tracks that pass geometry checks.

Dynamic Drift Correction of Landmark Databases

We use Landmark Matching technology for correcting long-term drift in the absence of a GPS signal. As the dismount goes through the complex terrain, a set of distinctive image features (visual landmarks obtained using Scale-Invariant Feature Transform (SIFT)/ Histogram of Oriented Gradients (HOG) methods, [SE, 2006, Zhu, 2008].) are stored in a database. These features are selected for their scale and rotational invariance in matching. If the dismount returns to the same area in the course of the exercise, the software searches the database via an efficient database indexing technique, to match the feature set on the newly observed image against feature sets on stored landmark database (Figure 5). A successful match results in a drift correction. This gives the system the ability to maintain the dismounts absolute location for extended periods of time even without GPS.

The landmark database is created from previously captured Light Detection and Ranging (LIDAR) and video imagery. We build a visual codebook for rapid database indexing and matching. The visual codebook is an efficient hashing scheme on visual feature description vectors that allow rapid indexing. Dynamically stored landmarks can be re-acquired to reduce longer range drift in both location and orientations. This approach allows production of landmarks that can be tracked accurately and robustly at video frame rate. We have successfully shown the use of view-point invariant point features as a robust representation for matching landmarks [Zhu, 2008]. These features can be matched across large viewpoint and scale changes (Figure 6).

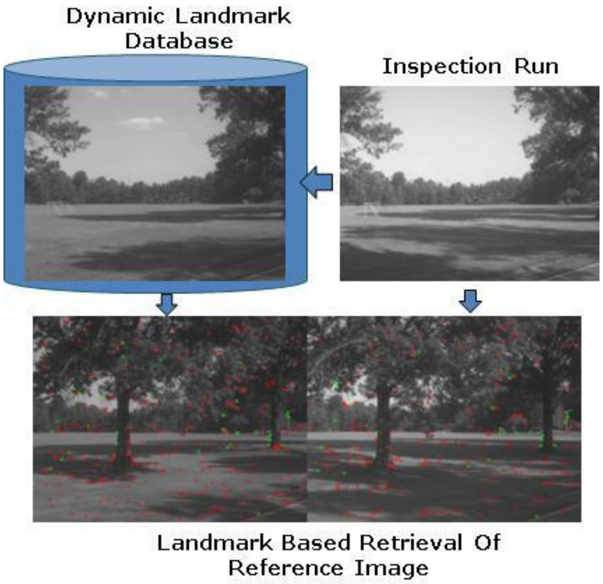

Dynamic Landmark Database
Inspection Run
Landmark Based Retrieval Of Reference Image

Figure 5. An example of a successful match between the new observed image and the reference image stored in the database.

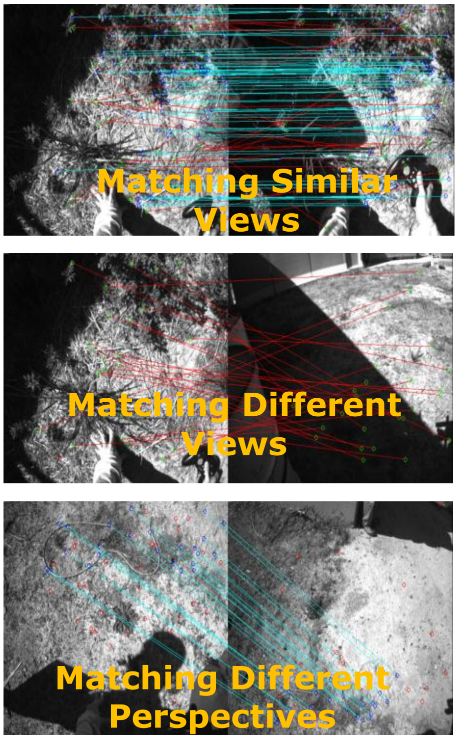
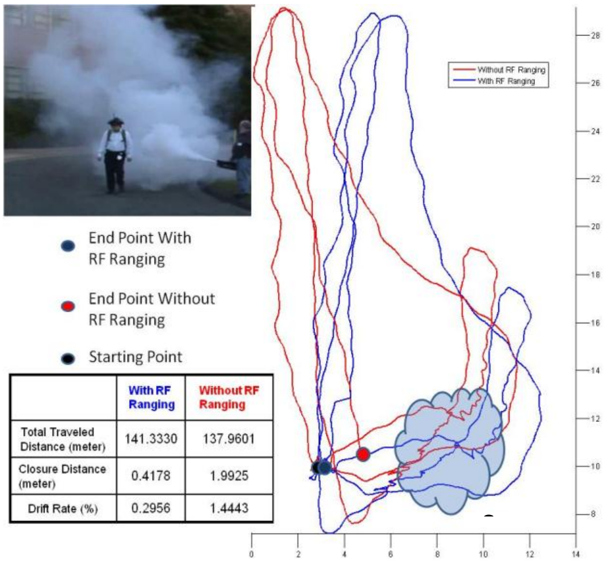


Figure 6. The features are robust to matching across similar views, different views, and even different perspectives.

Integration of RF-Ranging

RF-ranging techniques are particularly useful in the absence of GPS information and visual cues. The radio frequency waves can penetrate obstacles/ walls and do not require direct line of sight. The distance information between mobile humans and static RF nodes with known 3D locations in the transmission signal can be easily embedded with a unique ID for data association. These inter-nodal range readings are purely distance measurements, and do not contain sufficient information for 6-DOF pose computation. Thus, although they are also treated as global measurements, they constrain pose estimation to a lesser degree than landmark measurements. However, when integrated properly in a Kalman filter framework they contribute to limiting drift rate considerably, especially under certain situations where visual sensors fail due to lack of good features in the environment or low light conditions [Oskiper, 2010]. We compare the received RF measurements to the predicted value estimated by the current state in our filter. The measurements that exceed a difference threshold are rejected as outliers. The bias of each radio node, such as the gaps between the range measurements and the filter estimates, is also tracked as part of the filter states based on the accumulated measurements. The measurements that pass the outlier rejection procedure along with the bias states are used for updating the filter. Figure 7 shows the improvements due to RF ranging in smoky scenes.



	With RF Ranging	Without RF Ranging
Total Traveled Distance (meter)	141.3330	137.9601
Closure Distance (meter)	0.4178	1.9925
Drift Rate (%)	0.2956	1.4443

Figure 7: Tracking in clear and smoky areas using IMU, helmet worn video cameras and RF ranging. Trajectory with (blue) and without (red) using RF ranging. Note improvement using RF ranging due to smoke in part of scene.

AR RENDERING AND SIMULATION

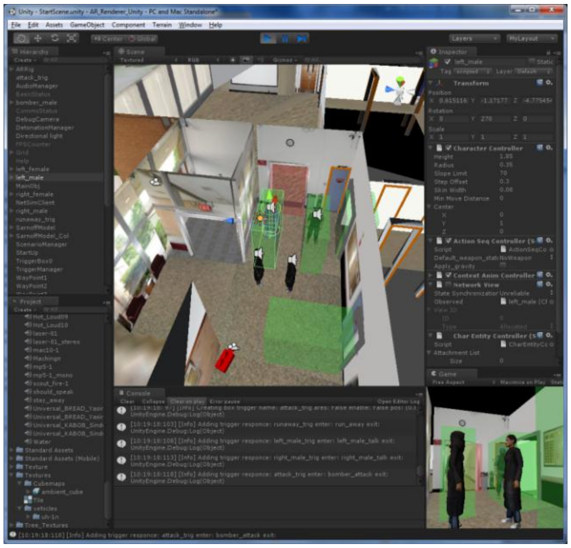
We are utilizing the Unity3D game development platform as the base layer of our rendering and simulation engine. We have developed a robust AR simulation framework upon Unity.

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.

Optimized techniques have been developed for input of augmented reality related data (such as head pose, depth pose, stereo background imagery, left/right stereo disparity, weapon pose, etc.) through a custom plugin interface over network or shared memory from our external localization and dense stereo process. Optimized techniques have been developed for efficient methods of background texture updates and a Graphical Processing Unit (GPU) based dense stereo re-projection (that is used for dynamic occlusion). These developments are critical to the usage of Unity as the rendering backbone of an augmented reality application.

Scenario Generation

We also customized Unity's provided scene editor in combination with an externally modifiable scenario specification to facilitate rapid scenario creation and modification (Figure 8). Utilization of the scene editor provides an intuitive layout of avatars, static waypoints, and trigger volumes. The external scenario specification provides a means of adjusting scenario layout and logic without necessitating the recompilation of the Unity player executable.



```
[10:19:18:97] [Info] Creating box trigger name: attack_trig ares: False enable: False pos: (0.
UnityEngine.Debug:Log(Object)
[10:19:18:103] [Info] Adding trigger responce: runaway_trig enter: run_away exit:
UnityEngine.Debug:Log(Object)
[10:19:18:108] [Info] Adding trigger responce: left_male_trig enter: left_male_talk exit:
UnityEngine.Debug:Log(Object)
[10:19:18:113] [Info] Adding trigger responce: right_male_trig enter: right_male_talk exit:
UnityEngine.Debug:Log(Object)
[10:19:18:118] [Info] Adding trigger responce: attack_trig enter: bomber_attack exit:
UnityEngine.Debug:Log(Object)
```

Figure 8: Scenario Generation.

Run-Time Avatar Control

The avatar control system for our simulation consists of a central simulation server and rendering clients running on each of the trainee worn processors. The central server controls all avatars, and is responsible for synchronizing avatar states (such as position, orientation, and current animation) amongst all renderer clients. All trainee locations are conveyed to the central server, providing multi-participant support. Currently the server contains a collection of pre-scripted high-level avatar action sequences and a collection of triggers based on trainee location and orientations. The actual locations and orientation of the trainees are compared against the trigger rules to initiate the avatar action sequences. Action sequences can also enable or disable triggers, providing cascading trigger conditional logic. Alternatively actions can also be executed via a visual server interface that can be used to explicitly manage the scenario by an exercise director.

Context sensitive automatic animation selection, adjustment, and mixing is used to simplify action specification, this automatic selection can also be overridden by alternative actions. The rules of the context animation selection can be specified by a user via an animation finite state machine. The avatar actions are executed on the server, via an avatar proxy, which broadcasts avatar state changes to all remote renderer clients residing on trainee units. The renderer client provides a detailed rendering of the synchronized avatar behaviors and effects (such as explosion, ammunition impact, etc.) from the correct player perspective.

OCCLUSION REASONING

For realistic augmented reality, the rendering engine must occlude all or parts of synthetic entities obscured by real world 3D structures based on the location of the synthetic. The pre-built 3D site model is used to do the occlusion reasoning for static structures in the scene. The low-latency trainee pose is used to render the site model into only the GPU depth buffer from the correct perspective. Subsequently when avatars are rendered to augment the view the model depth is automatically used to cull the rendered shapes for occlusions.

Stereo-based Depth Reasoning

In order to do the occlusion culling for dynamic objects, we use the depth map [Szinitsev, 2010] computed continuously using the helmet worn stereo cameras (same cameras used for pose estimation) to augment the model based reasoning using a pre-built model. The estimated pose with respect to the 3D model is used to render a range map of the background. The foreground range map is computed using a real time, robust, and accurate stereo matching algorithm, based on a coarse-to-fine architecture. At each pyramid level, we use non-centered windows for matching and adaptive up-sampling of coarse-level disparities. To minimize propagation of disparity errors from coarser to finer levels, we also perform an iterative optimization, at each level, that minimizes a cost function which preserves occlusion boundaries. An example of depth estimation for occlusion reasoning using stereo and model rendering respectively is shown in Figure 9.

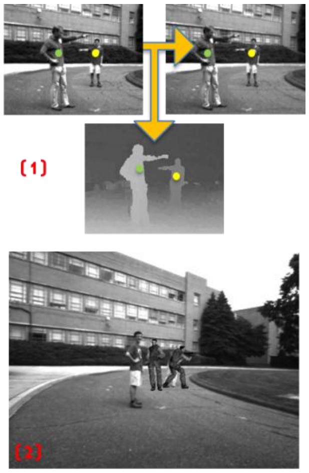


Figure 9. Depth estimation and occlusion reasoning using (1) stereo estimation and (2) rendering of 3D model.

TRAINING AREA MAPPING MODELING AND LANDMARK COLLECTION

Augmented reality requires building an accurate 3D model of the site and generating a set of visual landmarks that place the players in a common coordinate frame that is shared with the 3D model. Mapping the site is accomplished using a robot or sensor cart that is outfitted with visual and 3D sensors. As shown in Figure 10, the robot/sensor cart traversed the exercise site under remote operation, continually scanning the environment using calibrated video cameras and 3D LIDAR range scanner, building an integrated 3D point cloud and a set of landmarks which contain both 2D image features and 3D point locations. The output model is formatted to ingest into the rendering/simulation engines and landmark database is integrated onto the trainee worn system for accurate localization in 3D space.

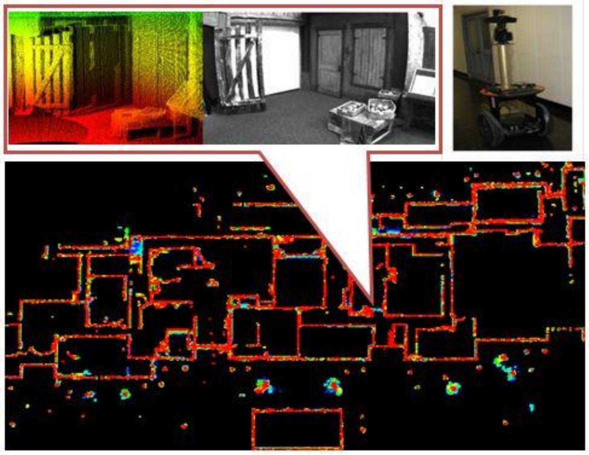


Figure 10. LIDAR and Video Based Mapping at Exercise Site.

AR HARDWARE

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 lightweight 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. Figure 11-a shows our current sensor design.

The compact sensor head is made of two USB cameras and an X-Sense MTI-G unit. The MTI-G has an IMU, magnetometer, GPS and barometer. The sensors are instrumented such that MTI-G unit triggers the cameras at a programmable interval. As such we ensure precise time synchronization between all the sensors. The back of the sensor package has an adaptable HMD attachment. We have successfully integrated Vuzix HMDs and Intevac i-Port 75 HMDs with the sensor package.

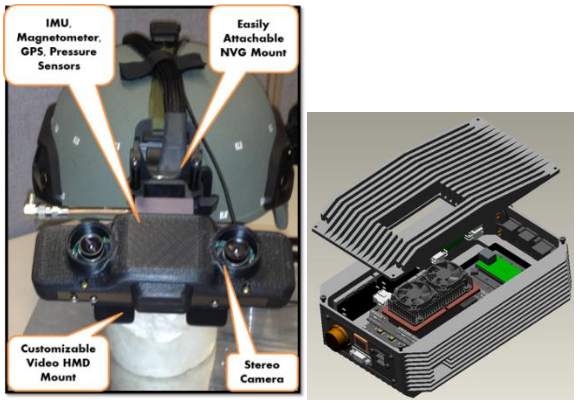

IMU, Magnetometer, GPS, Pressure Sensors
Easily Attachable NVG Mount
Customizable Video HMD Mount
Stereo Camera

Figure 11. (a) Sensor Head, (b) Compact Processor.

Our processor package consists of a PC-104 dual-core Intel i-7 processor board and custom PC-104 accelerator board. The board has an Altera Stratix IV Field Programmable Gate Array (FPGA). The FPGA was programmed to compute stereo depth maps with low latency in real-time. The current system uses the board to generate 640x480 depth images at 30Hz. However, the system is designed to run depth estimations at 2Kx2K resolution in real-time [Gudis, 2012]. The Intel processor then computes the localization solution, packages it with the stereo depth and sends it to the rendering engine that displays augmented reality entities with proper occlusion culling.

The new system (Figure 11-b) 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 for in-the-field switching. In previous applications the stereo was implemented on a laptop GPU’s and had a power utilization of over 250 Watts. In the future we plan to port the localization methods to the FPGA enabling further power reductions.

EXPERIMENTAL RESULTS

In this section, we report a number of experiments aimed at evaluating different aspects of performance of our tracking framework. We also demonstrate that our framework can provide highly accurate real-time tracking in both indoors and outdoors over large areas. Compared to [Oskiper, 2007] and [Zhu, 2008], we show our navigation system can provide more stable pose estimation to fulfill the demanding requirements for augmented reality applications.

Real-Time Tracking Over Large Areas

To demonstrate that our system can be used both indoors and outdoors over large areas, Figure 12 shows the automatically generated real-time camera trajectory corresponding to a 196.5 meter course within Exercise site completed by a user wearing our helmet, backpack system, and a video see-through HMD.

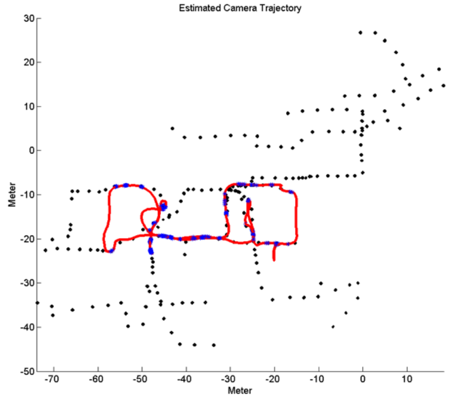


Figure 12. Real-time computed camera trajectory corresponding to a 196.5 meter completed in 3.25 minutes during an online exercise.

This user walked indoors and outdoors, moving through narrow hallways, rooms with white walls, etc. The entire trajectory shown in Figure 12 is within the pre-built landmark database capture range which is loaded in the beginning before the exercise takes place and landmark matches occur whenever a query image is within close proximity to a stored landmark shot in the database. As shown in Figure 13, the system locates itself inside the building immediately as the system turns on and estimates the travelled trajectories of the person, which is marked as red; the dark represents the locations where the landmarks were collected and the blue shows where successful landmark matching happened.

Figure 13 shows several screen shots corresponding to locations towards the beginning, middle and end of this exercise obtained from our visualization tool which we use to verify the accuracy of the camera pose outputs. This visualization tool uses the camera poses output by our system to render views from a 3D graphical model built upon the same visual data as the landmark database which also forms the global coordinate system. We compare the render views to the actual video images. It is observed that these views are in very good agreement which indicate how precisely the camera is tracked throughout the entire duration of the course.

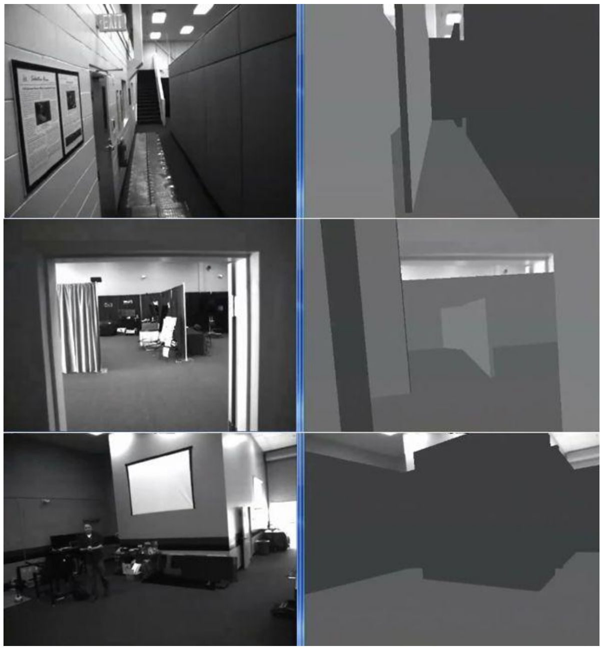


Figure 13. The views rendered from the model using the real-time camera pose estimates by our system for various locations throughout the exercise, together with the real scene views captured by the camera.

Pose Estimation for Augmented Reality

During the exercises of the augmented reality training scenarios, we inserted a group of virtual actors at particular locations based on the estimated pose and recorded the insertion video which was seen by the user from the video-see-through HMD. The scenarios are carefully designed, and each virtual person's 3D positions are pre-determined with the built 3D model, and a pre-configured action will be triggered once the trainee is approaching each virtual person. This video can verify whether the virtual actors are correctly aligned to the real scene based on real-time tracking. The pose estimation from our system needs to be very accurate and stable during the whole course, otherwise it will break the illusion of mixture between rendered and real world for the user. Figure 14 shows eight snapshots of the videos when the user went through two training scenarios: one indoor (top) and one outdoor (bottom). The positions of the inserted actor are mixed very well with the environment in these eight snapshots. This result demonstrates that our system is able to provide drift free pose estimation for an extended time period and distance travelled.

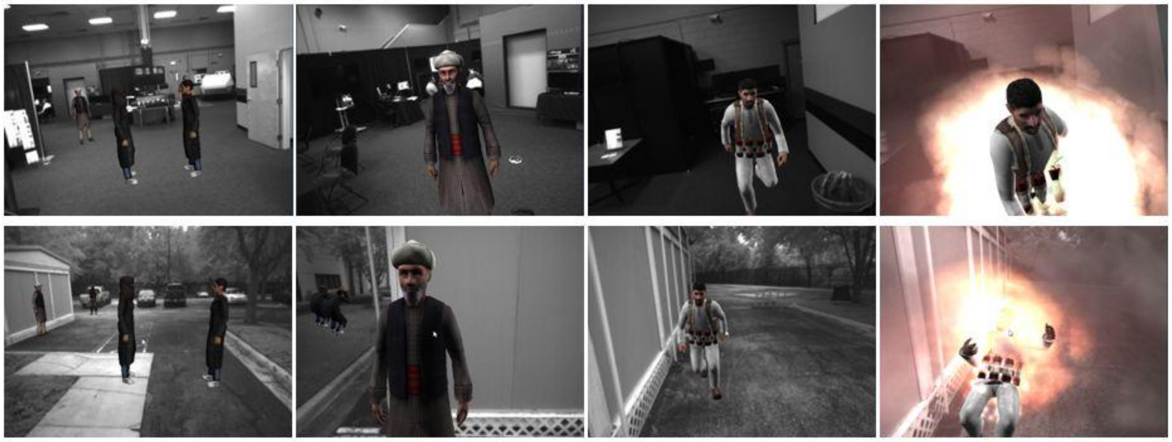


Figure 14. The snapshots of live exercises of training scenarios (both indoors (top) and outdoors (bottom)) .

CONCLUSIONS

We presented technical modules and experimental results from an infrastructure free augmented reality system. The augmented reality system can be used for *live* training of dismounts at home stations or deployed locations. The system reduces the need of hiring a large number of actors to role-play opposing forces and crowds in the environment. Avatars for opposing and friend forces and civilians are automatically rendered by a simulation engine onto the Head Mounted Display (HMD) of the trainees. The rendered avatars and effects appear as if they are part of the live scene. The simulation engine is used to control the behavior of the avatars. Scenarios with different levels of difficulty can be generated for different exercises and adaptive training.

The augmented reality system uses a 6 DOF tracking system based on unified Kalman filter framework using local and global sensor data from IMU, stereo cameras and range radios. Using a pre-built landmark database of the entire exercise area provides precise tracking and eliminates the problem of long term drift inherent in any inertial based navigation platform. We showed how to construct this landmark database automatically using sensors mounted on a robot/sensor cart system. We described the augmented reality occlusion reasoning and simulation and rendering and scenario generation modules. We described the augmented reality hardware and processing platforms that we have developed to conduct our exercises. We showed experimental results to illustrate the accuracy and robustness of our system for 6-DOF seamless indoor/outdoor tracking and augmented reality for dismount training. Finally the movements and activities of the trainees are captured in great detail and can be used for an after action review.

ACKNOWLEDGEMENTS

This material is based upon work supported by the Marine Corps Warfighting Lab under Department of the Air Force, Eglin AFB Contract FA9200-07-D-0045/0019 and the U.S Army Research Laboratory Simulation and Training Technology Center under Office of Naval Research Contract N00014-11-C-0433/0002 and 0004. 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

[Cheng, 2009] 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

[Gudis, 2012] 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

[Fontana, 2002] 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.

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

[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.

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

[Oskiper, 2007] T. Oskiper, Z. Zhu, S. Samarasekera, and R. Kumar. Visual odometry system using multiple stereo cameras and inertial measurement unit. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2007.

[Oskiper, 2010] 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.

[Oskiper, 2011] 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.

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

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

[Sizintsev, 2010] M. Sizintsev, S. Kuthirummal, H. Sawhney, A. Chaudhry, S. Samarasekera and R. Kumar, "GPU Accellerated Realtime Stereo for Augmented Reality", *In Proceedings of the 5th International Symposium 3D Data Processing, Visualization and Transmission (3DPVT)*, 2010

[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*