Mobile Augmented Reality for Force-on-Force Training

Richard Schaffer, Sean Cullen, Laura Cerritelli

Lockheed Martin
Burlington, Massachusetts
{Richard.L.Schaffer, Sean.Cullen, Laura.Cerritelli}@lmco.com

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

ABSTRACT

Live field training against a thinking human opposing force – force-on-force training – is highly valued by commanders. However, a limitation of current force-on-force training is the lack of battlefield effects, such as mortar or artillery detonations. This prevents fully employing indirect fires as part of combined arms operations in these exercises. In particular, forward observers have no means to adjust fire if they cannot observe impacts. We describe the development of a prototype system that provides mobile forward observers the visual feedback they need to conduct these operations.

The key emerging innovative technology that enables this training is precision mobile augmented reality. Augmented reality inserts virtual elements into views of real environments. In this application, a forward observer's position and look direction must be precisely tracked in real-time in order for battlefield effects to appear stably in the correct location. This precision must be maintained as the observer moves between positions. In addition, the effects must be rendered realistically, so they appear to be part of the environment and reflect local conditions, including wind and obscuration by terrain. Forward observers routinely use binoculars to locate targets and adjust fire. Consequently, augmented reality capable binoculars are also required for this task. As an additional challenge, the tracking and rendering for both naked eye and binocular views must be performed on a small, lightweight, body-worn computer compatible with field use. Finally, the system must integrate with an existing LT2 force-on-force training system.

This paper describes the key advances needed to produce the prototype system. We focus in particular on the challenges of extending an earlier prototype designed for use only from fixed positions and not connected to any live training system. The results of initial demonstrations at MCB Quantico are also presented.

ABOUT THE AUTHORS

Richard Schaffer is 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 30 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.

Rakesh Kumar is 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.

Sean Cullen is 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 over 17 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.

Laura Cerritelli is a Staff Software Engineer at Lockheed Martin MST. She received a B.S. in Computer Science from the Massachusetts Institute of Technology in 2003. Laura has over 8 years' experience in military simulation system research and development including work on Joint Terminal Attack Controller and Forward Observer trainer prototypes.

Supun Samarasekera is the Technical Director of the Vision and Robotics Laboratory 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 over 17 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.

Mikhail Sizintsev is 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.

Taragay Oskiper is a Senior Principal Research Scientist at SRI International, Princeton. He received his Ph.D. in Electrical Engineering from Princeton University. He has over ten 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.

Vlad Branzoi is 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.

Mobile Augmented Reality for Force-on-Force Training

Richard Schaffer, Sean Cullen, Laura Cerritelli

Lockheed Martin
Burlington, Massachusetts
{Richard.L.Schaffer, Sean.Cullen, Laura.Cerritelli}@lmco.com

Rakesh Kumar, Supun Samarasekera, Mikhail Sizintsev, Taragay Oskiper, Vlad Branzoi
SRI International
Princeton, New Jersey
{Rakesh.Kumar, Supun.Samarasekera Mikhail.Sizintsev, Taragay.Oskiper, Vlad.Branzoi}@sri.com

INTRODUCTION

Live, force-on-force training is one of the most valued forms of military training. It allows trainees to engage in the combat of wills between thinking, adaptive human opponents that is characteristic of warfare. Currently, such training is supported by systems such as the Marine Corps' Instrumented-Tactical Engagement Simulation System (I-TESS) and the Army's MILES-heritage systems. These systems enable direct fire engagements employing coded lasers and laser detectors. However, they only support limited indirect fire engagements, typically controlled from a central exercise control station. A key limitation of such systems is that trainees are unable to see battlefield effects such as direct fire impacts and indirect fire mortar and artillery explosions. This limitation is particularly critical for indirect fire, as it prevents forward observers from adjusting fire based on observed impacts. This in turn limits the ability to conduct full combined arms training.

Augmented reality is a rapidly-emerging technology that offers a solution to this problem. Augmented reality refers to technologies for inserting virtual elements into live views of the real world. In the context of military training, it promises to bring the best of virtual and live training together (Defense Science Board, 2013, pp. 60-65). The Office of Naval Research's Augmented Immersive Team Training (AITT) program is concluding a 5 year effort to apply augmented reality to USMC training needs (Schaffer, Cullen, Meas, & Dill, 2013). The first phase of that program focused on observer training from fixed locations (Kumar, et al., 2013). The second phase of the effort, reported here, is tackling the challenging task of bringing augmented reality to forward observers in a force-on-force training environment. A key challenge of force-on-force training is that the augmented reality hardware must be man-wearable and mobile. Hence there are much more stringent size, weight and power (SWaP) constraints on the system than the previous system. This required optimization and modification of existing algorithms to run within these constraints. In addition, significant new functionality was required in the form of precision tracking algorithms able to operate when moving freely over a training area as well as a means to interface to the force-on-force training system. A final challenge was enabling two previously separate augmented reality devices – an unaided eye head-worn display and a binocular / laser range finder – to run on the same SWaP-limited mobile hardware. This combined system is referred to as the "two-in-one" system (Figure 1).

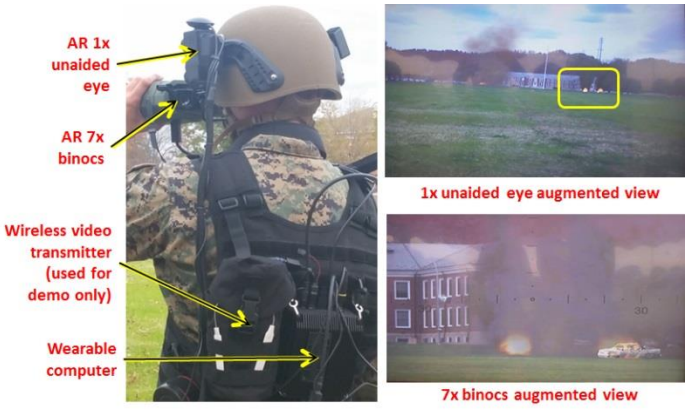

Figure 1. The two-in-one system displays the real world augmented with virtual elements such as targets and explosions. Unaided eye and magnified binocular views are shown on the right.

The two-in-one system uses video see-through augmented reality technology. In a video see-through system, a real-time video feed and associated precise camera position and orientation data (pose) are used by a rendering system to insert virtual elements into the real world view. These are displayed on an opaque ruggedized Head-Worn display (HWD) such as could also be used in a virtual reality system. The key hardware components of the system are the immersive head-worn display, the dual camera video head and associated sensors, the body-worn computer and the augmented reality

binocular prop. The HWD is a commercial Intevac I-PORT 75 with a 60 degree horizontal field of view, 100 percent binocular overlap, and a1280 x 1024 resolution. The dual camera video head is a custom unit comprised of commercial components described in more detail in the Navigation and Tracking section. The body-worn computer is a custom system based around a PC/104 single board computer with an Intel Core i7-4700EQ 2.4GHz CPU and 8 GB of RAM. The computer runs AITT navigation and tracking, rendering, simulation, and live training interface software. The augmented reality binocular prop is 3D printed in the form factor of the USMC's Common Laser Range Finder (CLRF). Software enables it to operate either as a pair of binoculars with a mil reticle or as a fully-functioning Vector 21B CLRF. For brevity, this multifunctional device is subsequently referred to simply as binoculars or binocs. The right side of Figure 1 shows the views displayed to a trainee. The upper view shows an unmagnified view as displayed on the HWD. The lower view shows a magnified view produced by the binoculars. The magnified area is marked with a yellow rectangle. In both views, inserted virtual vehicles and detonations are visible. The scene is from an event conducted at MCB Quantico.

This paper is divided into five sections. The first describes the Marine live force-on-force system into which the augmented reality system is being integrated. The second presents recent demonstrations and assessments. The next two provide details of the rendering and tracking system enhancements required to operate in the mobile force-on-force environment. Finally, exploratory work in applying an alternative, optical see-through, augmented reality display technology is described.

FORCE-ON-FORCE TRAINING INTEGRATION

I-TESS Increment II (I-TESS II) is a US Marine Corps program of record for live, force-on-force training. The Marine Corps Program Manager for Training Systems (PM TRASYS) is developing an enhancement to the I-TESS II system incorporating mortar team and forward observer training elements originally developed by the U.S. Army's OneTESS program. The OneTESS hardware includes sensors that attach to a real mortar allowing the system to determine the mortar's deflection and elevation. When a simulated mortar round is dropped into the mortar, the system computes the round's simulated impact location and time. In the original OneTESS system, this information was conveyed to a Forward Observer kit that included a tablet computer for displaying the impact. The impact was shown either on a 2D map or on a simple 3D display that was not tracked to trainee's viewing direction. The detonation information was also used to signal appropriate damage to participants in the live training system. This is the system into which we were challenged to integrate augmented reality technology. In particular, the goal was to replace the relatively limited, artificial and symbolic representation of mortar detonations provided by the OneTESS tablet with a more natural display of 3D virtual detonations on the trainee's view of the real terrain.

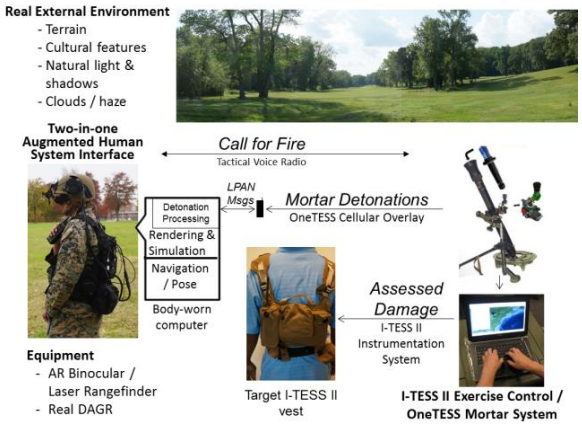

Figure 2. AITT Live Training System interfaces

The AITT system's interface to PM TRASYS's combined OneTESS/I-TESS II system is provided at the level of the player unit vest worn by the Forward Observer. It is based on Live Training Transformation (LT2) Live Player Area Network (LPAN) Standard (U.S. Army PEO STRI, 2015) messages conveyed over a Bluetooth connection. Due to the physical mortar round's time of flight, detonation information is received in advance along with a timestamp indicating when it should be rendered. Initial development was conducted with the OneTESS MILES Network Interface (OMNI) device. This was the interface used at the player unit in the original OneTESS system. Hence the AITT two-in-one system communicated via exactly the same protocol used by the OneTESS tablet. More recently, a new Android-based device, implementing an updated version of the LPAN standard, has been integrated and replaced the OMNI. It is the interface planned for the final system. The overall interface between AITT's two-in-one system and the OneTESS/I-TESS system is shown in Figure 2.

The initial, fixed location, phase of AITT employed an HLA interface and a Wi-Fi network to connect all exercise participants in a local area. This allowed these participants to share a common view of augmented environment as

well as interface with other DOD simulations. However, maintaining network connectivity limited the travel of participants to many tens of meters. The I-TESS II system uses a much longer range VHF radio and a cellular data link overlay for OneTESS communications is being added. Since the initial I-TESS II system did not have the ability to communicate detonation messages to the AITT system, a gateway was developed that carried detonation events directly from the I-TESS II exercise control system to the AITT system's HLA network. Any detonation message on the Common Training Instrumentation Architecture (CTIA) network was captured and transformed into an AITT system HLA detonation message. This initial capability allowed a trainee wearing an I-TESS II vest to be both damaged by a CTIA detonation event as well as observe the detonation through the AITT augmented reality system. This system retains the earlier capability to inject dynamic virtual targets as well as a wide variety of detonations. It supports demonstrations when a sufficient quantity of I-TESS II player units and trainees are not available to serve as targets for the forward observer.

TECHNOLOGY DEMONSTRATIONS AND ASSESSMENTS

The system has been developed in a series of 6-months build cycles, each concluding with a demonstration and opportunity for Marines to operate and provide feedback on the system. The technology was most recently demonstrated at MCB Quantico in November 2014 and May 2015.

Figure 3. Marine participating in November assessment

At the November 2014 event, the AITT system was demonstrated at Lejeune Field, a parade ground chosen for ease of access. The demonstration focused on the transition from an observer at a fixed position to a mobile system that allows the observer to move around a predefined area. The mobile system consisted of the two-in-one configuration that allowed the observer to view targets and detonations via both an HMD and an augmented binocular or Vector21 laser rangefinder. The system allowed mobility over an approximately 50 meter by 50 meter area. As part of this effort an assessment was conducted by a team lead by the University of Central Florida (Champney, Lackey, Stanney, & Quinn, 2015). Five Marines experienced in conducting Call For Fire (CFF) each conducted three types of CFF missions – Grid, Polar, and Shift from a known point. A written survey was conducted regarding the use of the system as a CFF trainer (Figure 3). Results highlights include that four of the five would recommend the system for use as a CFF trainer and that the prototype was rated highest in comparison with other CFF trainers with which the Marines were familiar. Participants agreed with the system's utility as a CFF trainer based on its ability to allow the execution of CFF tasks as well as their subjective impressions. It was rated highly for realistic visual representations of ground vehicles, people, smoke and explosions, the expected behavior of its simulated binoculars and ordnance, update rate and highly "believable" scenarios. Neutral responses were received regarding the visualization of damage, where more specificity was desired. There were some concerns on the adjustability and comfort of the equipment, the screening of ambient light, and the technical support required to set up the equipment.

In May 2015, the AITT system conducted a demonstration on the edge of the Quantico Medal of Honor Golf Course. The demonstration highlighted how the AITT system could be used to transform any area into a Forward Observer training range. The demonstration included both an updated version of the mobile training system and elements of the previously developed system. This included the ability to insert virtual targets in the live environment, which is not required in the transitioning mobile training system. The exercises showed virtual enemy vehicles advancing down the fairway while Forward Observers conducted Call-For-Fire training to suppress and destroy the enemy targets. A Mk19 grenade launcher from another effort, which included a gunner wearing an AITT-style unaided eye system, engaged BMPs with virtual 40mm grenade rounds. In addition, Joint Terminal Attack Controller (JTAC) training capabilities were demonstrated. In this scenario, a virtual F-18 dropped a GBU-12 guided bomb on 2 enemy APC vehicles. In a second scenario, the JTAC controlled a virtual AH-1W helicopter in attacks on technical vehicles. The virtual helicopter was flown by a human at a Marine Corps' Deployable Virtual Training Environment (DVTE) Combined Arms Network (CAN) laptop simulator connected to the JTAC via Wi-Fi.

Images of the system in use at the May event are shown in Figure 4. The top row shows two unaided eye and one binocular view of the target area from the AH-1W scenario. The bottom row of Figure 4 shows a scenario in which

targets are engaged with virtual mortar fire. The screenshots are extracted from real-time video recorded on-site. As in a real-life call for fire, the observer used his unaided eye in conjunction with a magnified (7x) view.

Figure 4. Images from Quantico field tests. Top row: Augmentation of 1x and 7x views with renderings of armored vehicles and an attacking helicopter. Bottom row: Augmentation of 1x and 7x views with renderings of technical vehicles and munition detonations.

For both of these demonstrations, Marines from MCB Quantico operated the equipment. The demonstrations were conducted outside in the open with a small portable generator providing power to the support equipment, including large screen displays wirelessly repeating the Marine's mobile view.

A final iteration of the system is in progress, with assessments and demonstrations planned at MCB Quantico between August and October 2015.

RENDERING ENHANCEMENTS FOR MOBILE OPERATIONS

The AITT system rendering system is built on the Unity 3D game engine. The tracking system provides position, orientation, and camera projection information matched to each frame received from the camera. This information is used to generate a precisely registered augmented scene. The camera image is drawn to the background and a 3D terrain model of the operating area is rendered to the depth buffer so that virtual entities and effects are properly occluded by terrain features.

The transition from fixed to mobile operations required implementing rendering system optimizations. Using the Intel integrated HD 5500 graphics, rather than a separate graphics card, allowed the system to run much longer on battery power. In exchange, we were limited to the lower graphics throughput of integrated graphics. Turning off shadow effects simplified the shader pipeline. The HMD used in the system is capable of displaying a 1280x960 video resolution but driving the corresponding number of pixels caused the lower power GPU's framerate to drop below 15Hz. By switching to a 640x480 resolution the system maintained 40Hz update rates. The tracking system only supplies position and orientation updates at 30Hz, so 40Hz was sufficient.

The two-in-one mobile system must support both the unaided eye and binocular devices. Fortunately, the trainee can only physically see through one of the two devices at a time. This allows a single rendering channel to be dynamically reconfigured between the different simulations required for each device. The user switches to the binoculars by pressing a device-mounted button and returns to the unaided eye view simply by lowering the binoculars. The system also supports a mode in which the binocular display is not used and the HWD changes to a binocular or Vector 21 display when that device is raised and activated. This has become the preferred mode of use.

Occlusion Editor

In real life situations, generally not all objects that can block the line of sight are represented in the occlusion database. This may be due to recent real-world changes or because of limitations in the original source material. Hence, we've found it valuable to have a means to insert new or missing occlusion elements on-site. In a stationary augmented reality system, two dimensional virtual occluders can work well. These are essentially "billboards" of

the appropriate size and distance to reflect the average first surface of blocking objects. An example is side of a building. However, in a mobile system, where the user may potentially observe an obstruction from any angle, a three dimensional representation of occluders is required. To support this capability, we added an Occlusion editor to the Vector 21 emulation software. The approach is to first define the distance of a vertical polygonal surface based on the intersection of the line-of-sight with the terrain model. Vertices of a vertical polygon are then entered on this surface. Volumes are built up by "extruding" such surfaces or linking multiple surfaces. This enables cultural features, such a light poles and other man-made structures to be entered and modified. The Vector is operated from a steady tripod at one or more locations to build up necessary three dimensional model of the object. An example of entering the occlusions volume for a light pole and a newly created building occluder obscuring an augmented artillery detonation is shown in Figure 5.

Figure 5. Above: Entering occluder for a light pole. Below: Dust from a virtual detonation behind an inserted building occluder.

NAVIGATION AND TRACKING ENHANCEMENTS FOR MOBILE OPERATIONS

In acting as a Forward Observer, a Marine or soldier must switch between observing the world with his unaided eyes and observing potential targets and impacts through binoculars or other magnified optics. To support mobile live training without adding excessive weight, it is highly desirable that both devices be supported by a single mobile computer. In addition, a key ingredient of the required stable high precision tracking is the use of vision-aided navigation algorithms. This is especially challenging for the binoculars, which will often hang around the user's neck pointed at the ground, providing little or no useful visual tracking data. To address this challenge, an approach was developed that allowed the unaided eye tracking device to rapidly initialize the binoculars' tracking state. The overall tracking approach as well as the combined unaided eye and binoculars tracking system is described in this section.

The overall tracking approach is based on a dual 6-Degree of Freedom (DoF) pose estimation algorithm combining inputs from an Inertial Measurement Unit (IMU), Global Positioning System (GPS), and wide and narrow Field of View (FoV) cameras using an error-state Extended Kalman Filter (EKF). We employ a joint tracking pipeline for both the unaided eye and binocular systems, with the helmet-mounted unaided eye system acting as the master and the more challenging neck-worn binocular package acting as the slave. Methods implemented to aid the hand-held binocular pose estimation using the unaided eye sub-system are described. The software operates on a single wearable computer in a PC/104 form-factor with low latency and high precision.

Hardware

Figure 6. Helmet attached unaided eye and hand-held binoculars each include a wide FoV monochrome camera for robust tracking and narrow FoV color camera for augmentation and precision tracking.

The two-in-one hardware includes unaided eye and binocular AR modules as well as a portable computer as depicted in Figures 1 and 6. The unaided-eye sub-system consists of MEMs type MicroStrain 3DM-GX4-25 IMU, USGlobalSat GPS BU-353S4 receiver, wide 67.4° FoV monochrome PointGrey Flea3-GE-13S2M camera pointed 20° down to capture nearby ground providing rich texture and good depth estimates essential for reliable 6 DoF monocular visual-inertial odometry. It also includes a narrow 45.5° FoV color PointGrey Flea3-GE-13S2C camera to closely match the field of view of Intevac I-Port 75 display. The IMU and cameras are

synchronized by external trigger made of Teensy 2.0 board with real time clock.

The binocular AR sub-system consists of XSens-MTi-G unit that includes MEMs type IMU, magnetometer, GPS and barometric pressure sensor, wide 51.2° FoV monochrome Prosilica camera for reliable visual tracking of large motions and narrow 6.3° (7x) FoV color Prosilica camera for augmentation and precise tracking. Xsens sync-out is used to trigger the cameras. Note that the wide FoV lens of the binoculars was chosen to have a similar FoV to the narrow FoV lens of unaided eye sub-system to enable the reliable matching of images between devices essential for cooperative tracking.

Unaided Eye Augmented Reality Tracking Overview

A navigation filter in a multi-sensor fusion framework is used to combine visual (cameras), inertial (IMU) and global (GPS, landmarks, etc.) measurements to estimate the navigation state (location, orientation, velocities, accelerometer and gyroscope biases) at a regular rate given sensor measurements arriving at different times. The total (full) states of the EKF consist of the IMU location, ground to IMU orientation, velocity in global coordinate frame and gyroscope and accelerometer biases (Oskiper, Samarasekera, & Kumar, 2012; Oskiper, Sizintsev, Branzoi, Samarasekera, & Kumar, 2013). The EKF error-state is based on the relation between the total state and its inertial estimate. During filter operation, ground-to-IMU pose is predicted prior to each update instant by propagating the previous estimate using all the IMU readings between the current and previous video frames via IMU mechanization. After each update, estimates of the errors (which form the error-states of the filter) are fed-back to correct the predicted pose before it is propagated to the next update.

IMU Guided Frame-to-Frame Feature Tracking

Visual feature tracking is the basic component of pose estimation as it is used to generate relative measurements for odometry. In the following, we describe how we make use of the gyro readings from the IMU in the Random Sample Consensus (RANSAC) (Fischler & Bolles, 1981) hypothesis generation step for rejecting outliers from the frame-to-frame raw feature matches obtained via normalized cross correlation using only 3 points instead of 5 as in (Oskiper et al., 2012). This significantly reduces the number of RANSAC iterations and, consequently, reduces the processing load of the tracking modules.

The main idea stems from the fact that if orientation is obtained from gyro readings between two consecutive frames, translation between frame pairs can be deduced with only two points in correspondence by exploiting epipolar geometry relations. The third point is required to resolve chirality and the ambiguity of all points being in-front or behind cameras (Hartley & Zisserman, 2004), which can be done in closed-form. Finally, each hypothesis formed by the triplet of points is evaluated using trifocal Samson error (Hartley & Zisserman, 2004). If the error is below the threshold (3 pixels by default), the hypothesis is retained and will further be evaluated on all the feature correspondence data as part of the RANSAC process. However, if the error is above the threshold, it is immediately dropped. The hypothesis generation step is terminated whenever we obtain 25 such good hypotheses out of the maximum 500 RANSAC trials, and these 25 hypotheses are sent for evaluation on the entire dataset to determine the one with the overall best score which is finally used to reject outliers.

This mechanism speeds up the RANSAC process due to its much faster hypothesis generation compared to the 5-point method and by the early pruning of bad hypotheses. It also ensures only a small set of good hypotheses are evaluated. Finally, it only accepts hypotheses that are consistent with IMU-sensed motion. In practice, whenever there is good texture and lots of correct matches across the frames, the hypothesis generation process usually will terminate right after the first 25 or so trials, well below the maximum trial count of 500.

Global Heading Correction

Our high quality dead-reckoning pose tracking requires absolute measurements in all 6 Degrees of Freedom to eliminate drift. While GPS provides an absolute 3D location and the accelerometer provides absolute roll and pitch, determining global absolute heading (or yaw) is more challenging. Global heading is especially important during the initialization procedure in order to have EKF start in a good state. While a magnetometer can be used to determine approximate global heading, it requires prior calibration, is very susceptible to external electro-magnetic fields and does not provide the necessary accuracy for our pixel-precise augmentation. A more effective yet complex procedure requires using visual landmarks to determine absolute heading. Potential visual landmark approaches range from terrain-model skyline matching to employing a set geographically registered landmarks. Building on our

previous work (Oskiper et al., 2013; Oskiper, Sizintsev, Branzoi, Samarasekera, & Kumar, 2014), we improved the robustness of our GeoLandmark matching module in handling different viewpoints, lighting conditions and the presence of occlusions.

GeoLandmark Matching For Global Heading Correction

The purpose of the Geolandmark matching process is first to localize a known landmark point to establish its 2D image coordinates in a given query frame, after which one can determine the global heading of the camera using the 3D geodetic coordinates associated with the landmark and its 2D image location. First we will expand upon how the landmark database is created and the method by which its image localization is performed and then we will derive the measurement model for global heading correction.

Multi-GeoLandmark Database Creation

As a first step, we developed an easy to use interface that allows an operator to conveniently create a landmark database. This step involves the operator obtaining images of several easy to identify landmarks in the world, possibly from different viewpoints. This is done by selecting a landmark point on the video frames displayed on a window followed by a mouse click to record its 2D pixel coordinates. For each such point, an image patch of 100x100 pixels in size is selected around the landmark point. Feature points, along with their corresponding descriptors, are extracted for this region. We use a Harris corner detector (Harris & Stephens, 1988) for keypoint detection and OpenCV Oriented Brief (ORB) descriptors. The dominant orientation for the ORB descriptor is set explicitly based on the current camera roll estimate available from the EKF for that frame. The corresponding geo-coordinates are obtained via a mouse click on a Google-Earth like map displayed in another window which shows the predetermined available landmark locations overlayed on the map. The camera pose corresponding to this landmark frame is also stored as part of the database and will be used during the run-time matching process for geometric verification.

As described further below, we developed a very robust and powerful landmark matching pipeline to minimize the number of different landmarks and viewpoints required to operate in a large area. Here the goal is to have each landmark provide a large area of coverage around the location from which it was created, meaning that the radius of the region from which the same landmark patch can be matched successfully against any query frame within that volume is in the order of 50-100 meters. Furthermore, query images, being successfully matched to the geo-landmark, are added automatically to the database when the number of inliers falls below a threshold of 5 points. This automatically grows the landmark database to cover different views and lighting conditions without user supervision. Finally, in all the experiments and demonstrations we present in this paper, *only one* landmark from a single viewpoint was used. This provided the user a simple, one-step initialization process while still providing the system the needed occasional heading correction.

IMU Guided Landmark Matching

At run-time, the collected landmark database is pre-loaded, and whenever a new frame is received the landmark engine performs a match between this query frame and the database to determine the pixel location of the landmark in the query frame. If the database contains multiple landmark instances obtained from different image patches, the first task of the landmark engine is to return the keypoint and descriptor data for the top landmark patch candidates (we use the top three) that are believed to be visible in the query image. After computing the feature correspondences between the query image and the landmark patch, we perform geometric verification using a rotation-only model in a RANSAC framework. The winning hypothesis that denotes the pseudo-estimate of the relative rotation between the landmark frame and query frame, due to the ignored parallax effects, is only used as a means to transfer the pixel coordinates from the landmark image to the query image using rotational homography. However, since we are performing wide baseline matching between a small landmark patch and the query frame, we have to use a relatively low inlier threshold and need further robustness to reject bad hypotheses. Inliers are determined by explicitly comparing the deduced roll and pitch component of rotational homography to the roll and pitch estimates of EKF filters, which are expected to be highly accurate.

Binocular sub-system overview and two-in-one interaction

The Binocular system offers free form mobility while the user is walking around with the binoculars worn around the neck. The user may pick up the binoculars at any time and see an AR augmented world. The navigation framework for the binocular sub-system is based on that of Oskiper (Oskiper et al., 2014). However, the previous version relied on a single panorama database built on the fly for global heading correction and only allowed motion

in a few square meter region. In the previous system, the binocular AR system operated independently from the unaided eye AR system. Hence the user needed to remain close to the location of the initial landmark selection. In the new system, the user can move freely. This is because the binocular AR system depends on the unaided eye system for global location and heading corrections. As a result, even when the binocular is looking down, its global location remains known.

In particular, the motion estimation pipeline running on the unaided-eye system acts as the master and continuously transmits the head position and image data to the binocular system, as depicted in Figure 7. As a result, while the user is on the move with the binoculars pointing down, the EKF on the binocular system relies on these position estimates as additional sensor updates on its own position (in place of its GPS readings, which are of inferior quality and less frequent). So even though the visual-inertial odometry component on the binocs is severely challenged while the cameras are facing down and mostly blocked by the user's chest and body, the filter can successfully maintain navigation by fusing IMU readings and position updates from the unaided-eye.

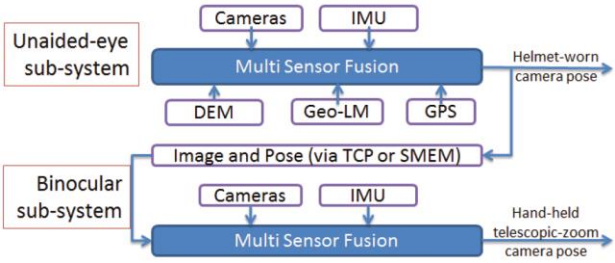

Figure 7. System block diagram for pose tracking

Under this operating condition, the drift on the global position and roll and pitch components of orientation remain very well bounded. However, global heading is prone to drift which needs to be corrected while the user is bringing the binoculars to eye level. For this purpose, the images from the forward looking narrow FoV unaided-eye camera are transmitted once every second and a local database of these images over a sliding window of one minute duration is maintained at all times in the binocs sub-system, including when they are pointed down. This allows for seamless transition between two sub-systems and rapid heading re-initialization of the binocs as soon as they are raised to the eye level by matching the query frames from the binocs to the sliding window unaided-eye image database and obtaining a global yaw measurement after a successful match. Since the helmet and binocs are in very close proximity of each other, rotation-only matching suffices for this purpose. In order to aid the image indexing and matching process between the two systems, we use similar fields of view on the unaided eye's narrow FoV camera and the binoculars' wide FoV camera.

Experimental Results for Navigation and Tracking

In order to assess the quality of our 6 DoF tracking in large open spaces, we moved in a loop on a field while looking at a building and marking its opposite corners. Correct geo-landmark locations were hand-labeled for every fifth frame in the sequence and actual insertions compared to this ground truth. The observer's trajectory, rendering snapshots and quantitative results are depicted in Figure 8. To demonstrate the necessity of continuously running the

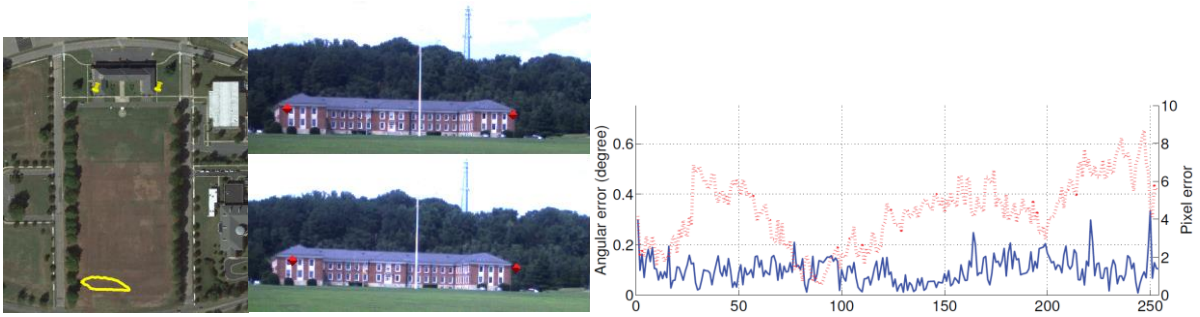

Figure 8. Quantitative evaluation of the pose estimation accuracy. Left: Google Earth view of the field with trajectory and marker geo-locations overlaid in yellow. Middle: Views of a building corner augmented with virtual red markers for a run with the GeoLandmark module (top middle) and without the module (bottom middle) at the end of the run. Right: Plot of the pixel (right axis) and equivalent angular error (left axis) for every hand-labeled marker occurrence. Blue solid and red dotted lines show the tracking error with and without the GeoLandmark module, respectively. Note that the ground truth point numbers shown in the x-axes are not necessarily continuous in time.

3D-2D GeoLandmark matching module, we ran the EKF with and without it. The run without the GeoLandmark module's global orientation correction drifts noticeably especially toward the end of the sequence. In contrast, the full system incorporating the module exhibits a very low root mean-squared error of RMS_{pixel} = 1.33 and RMS_{angle} = 0.10 degrees, which is within the accuracy of the hand labeling process itself.

OPTICAL SEE-THROUGH PROTOTYPE

In an exploratory effort, we have also developed an early-stage augmented reality prototype based on an Optical See-Through (OST) Head-Worn Display (HWD). In an optical see-through HWD, the user views the real world through a combining optics that overlays virtual elements from a display onto a nearly direct real world view. The OST approach offers potential advantages, but comes with a number of challenges. The potential advantages include providing the user a real-time potentially full human field-of-view and resolution and view of the real world. This is particularly valuable for mobility during a live force-on-force exercise. Challenges include providing sufficient bright display for the virtual overlays to be usable in an outdoor environment and the inability to virtual insertions to block the real world view behind them. The HWD used in the prototype is the LARS developed by SA Photonics under an ONR STTR.

Pose Prediction for Optical See-Through Augmented Reality

In general for an immersive augmented reality user experience, accuracy of the pose estimates alone is not sufficient for the user who is wearing the system. The rendered markers also need to appear with very little delay on the display. This is especially demanding for the Optical See Through framework, where the user sees the real world as it is (not an image of it) and there is no-delay in visual perception of the real world. Therefore, the associated rendered markers have to satisfy this highly demanding timing requirement in order for them to appear jitter-free when they are displayed. Otherwise as the user's head is bobbing, the markers will appear to bounce around (also referred to as swim) in the display since they will be lagging in time.

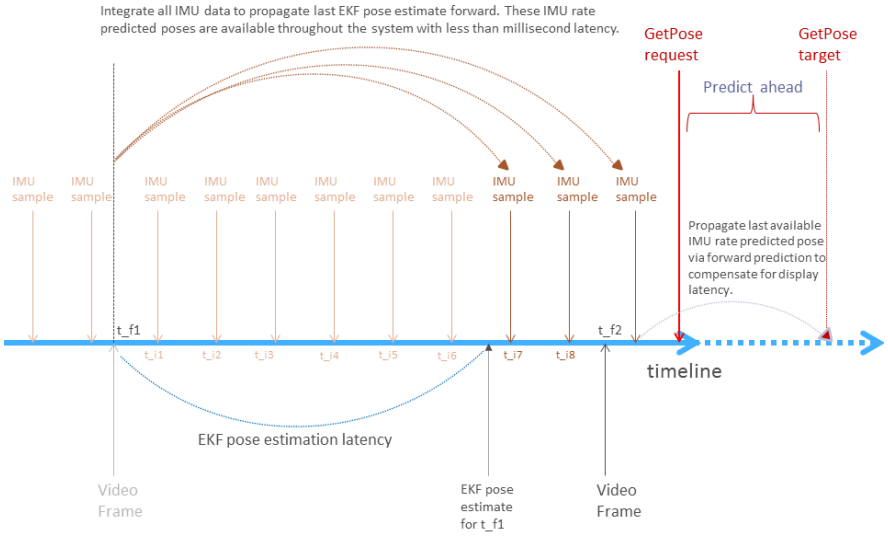

Figure 9. Forward prediction of navigation state estimate for optical and video see through Augmented Reality Displays.

Figure 9 shows the timeline of sensor inputs and algorithm outputs in relation to forward prediction for camera pose estimation. Video frames in general arrive at a much slower rate (e.g., 15 or 30 Hz) than the IMU samples (e.g. 100 Hz.) The pose estimates that incorporate each video frame's information is in general available after 40-50 milliseconds of processing delay. The pose requests from the renderer arrive asynchronously at the highest rate the renderer can accommodate. After the renderer receives a pose, it is displayed on the see-through display after a delay due to both display hardware latency and the video graphic card's rendering pipeline latency. In order to compensate for all the latencies in the system, a forward prediction mechanism is required to estimate the camera pose corresponding to a certain timestamp into the future given all the information that is available until the render request is utilized.

Rendering for Optical See-Through Augmented Reality

Optical See-Through (OST) presents some unique challenges to the rendering system. Since in an OST system your eye sees the real world optically overlaid with computer generated imagery, almost any latency is detectable. At the same time, the rendering requirements for the OST system are greater than for the two-in-one Video See-Through (VST) system. One reason, is that unlike AITT's VST HMD, which has 100 percent overlap between the images displayed to each eye and can be fed a single video stream, the OST HMD has only a 20 degree horizontal overlap between eyes. Consequently, the OST system must perform stereoscopic rendering. This means the rendering system must receive a position and orientation for each eye and render an image for each screen. The Unity 3D game engine only supported a single framebuffer so a large frameless window was created that spanned two display screens. Additionally, the renderer needed to warp each eye to offset the distortion introduced by the OST HMD's optics. This required running a post-processing shader over the frame buffer with input from a warp map texture. This texture encoded two-dimensional displacement values for each pixel along with a mask for areas outside the distorted edge of the frame. This extra processing, along with individual rendering passes for each eye, required utilizing a gaming quality computer for the system. A Gigabyte GB-BXi7G3-760 was used. This computer provides desktop machine CPU and GPU capabilities in a compact 5"x5"x2" footprint. Using this hardware we were able to maintain 100-120Hz update rates for a typical scene.

ACKNOWLEDGEMENTS

This work conducted a part of an Office of Naval Research Capable Manpower Future Naval Capability program. We thank ONR for their support. We also thank PM TRASYS for their assistance with OneTESS and I-TESS II. Finally, we'd like to thank the Marines who have participated in our demonstrations and assessments and whose enthusiasm and feedback have enabled our progress. The views, opinions, or findings contained in this report are those of the authors and should not be construed as an official ONR position, policy, or decision unless so designated by other official documentation.

REFERENCES

Champney, R., Lackey, S., Stanney, K., Quinn, S. (2015). Augmented Reality Training of Military Tasks: Reactions from Subject Matter Experts. In *Virtual, Augmented and Mixed Reality: Systems and Applications.* Berlin: Springer.

Defense Science Board, (2013). *Report on Technology and Innovation Enablers for Superiority in 2030.* Office of the Under Secretary of Defense for Acquisition, Technology, and Logistics, Washington, D.C..

Fischler M. A. and Bolles R. C. (1981). Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. In Communications of the ACM, vol 25(5) pp 381-395.

Harris C. and Stephens M. (1988). A combined corner and edge detector. Proceedings of the 4th Alvey Vision Conference pp 147-151.

Hartley R., and Zisserman A. (2004). *Multiple View Geometry in Computer Vision*, Second edition, Cambridge University Press, Cambridge, UK.

Kumar, R., Samarasekera,S., Sizintsev, M., Oskiper,T., Branzoi,V., Schaffer, R., Cullen, S. , Krishnaswamy, N. (2013). Augmented Reality Training for Forward Observers, in Proceeding of the Interservice/Industry Training, Simulation & Education Conference 2013. Arlington, VA: National Training and Simulation Association.

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

Oskiper T., Sizintsev M., Branzoi V., Samarasekera S., Kumar R. (2013). Augmented reality binoculars. In *2013 IEEE International Symposium on Mixed and Augmented Reality*.

Oskiper T., Sizintsev M., Branzoi V., Samarasekera S., Kumar R. (2014). Augmented reality binoculars on the move. In *2014 IEEE International Symposium on Mixed and Augmented Reality*.

Schaffer, R., Cullen, S., Meas, P., & Dill, K. (2013). Mixed and Augmented Reality for Marine Corps Training. In R. Shumaker (Ed.) *Virtual, Augmented and Mixed Reality: Systems and Applications* (pp 310-319). Berlin: Springer.

U.S. Army PEO STRI. (2015). Live Training Transformation (LT2) Live Player Area Network Standard (Document Number PRF-PT-00549, Revision D). Orlando, FL. PEO STRI.