New Generation of Instrumented Ranges: Enabling Automated Performance Analysis

**Amela Sadagic[1], Greg Welch[2], Chumki Basu[3], Chris Darken[1], Rakesh Kumar[3],
Henry Fuchs[2], Hui Cheng[3], Jan-Michael Frahm[2], Mathias Kolsch[1], Neil Rowe[1],
Herman Towles[2], Juan Wachs[1], Anselmo Lastra[2]**

(1) Naval Postgraduate School (NPS)	(2) University of North Carolina (UNC)	(3) Sarnoff Corporation
Monterey, CA	Chapel Hill, NC	Princeton, NJ
asadagic@nps.edu, cjdarken@nps.edu, kolsch@nps.edu, ncrowe@nps.edu, jpwachs@nps.edu	welch@cs.unc.edu, fuchs@cs.unc.edu, jmf@cs.unc.edu, herman@cs.unc.edu, lastra@cs.unc.edu	cbasu@sarnoff.com, rkumar@sarnoff.com, hcheng@sarnoff.com

ABSTRACT

Military training conducted on physical ranges that match a unit's future operational environment provides an invaluable experience. Today, to conduct a training exercise while ensuring a unit's performance is closely observed, evaluated, and reported on in an After Action Review, the unit requires a number of instructors to accompany the different elements. Training organized on ranges for urban warfighting brings an additional level of complexity—the high level of occlusion typical for these environments multiplies the number of evaluators needed. While the units have great need for such training opportunities, they may not have the necessary human resources to conduct them successfully. In this paper we report on our US Navy/ONR-sponsored project aimed at a new generation of instrumented ranges, and the early results we have achieved. We suggest a radically different concept: instead of recording multiple video streams that need to be reviewed and evaluated by a number of instructors, our system will focus on capturing dynamic individual warfighter pose data and performing automated performance evaluation. We will use an in situ network of automatically-controlled pan-tilt-zoom video cameras and personal position and orientation sensing devices. Our system will record video, reconstruct dynamic 3D individual poses, analyze, recognize events, evaluate performances, generate reports, provide real-time free exploration of recorded data, and even allow the user to generate 'what-if' scenarios that were never recorded. The most direct benefit for an individual unit will be the ability to conduct training with fewer human resources, while having a more quantitative account of their performance (dispersion across the terrain, 'weapon flagging' incidents, number of patrols conducted). The instructors will have immediate feedback on some elements of the unit's performance. Having data sets for multiple units will enable historical trend analysis, thus providing new insights and benefits for the entire service.

ABOUT THE AUTHORS

Amela Sadagic (Ph.D., University College London, UK, 1999) is a Research Associate Professor of Modeling Virtual Environments and Simulations (MOVES) at the NPS. In the past she was a Director of Programs at Advanced Network and Services Inc. where she led programs focused on the use of emerging technologies in learning, and coordinated National Tele-Immersion Initiative. Her interests include VR, human factors, multiuser collaborative environments, game-based systems, use of emerging technologies for training and learning, and diffusion of innovation.

Greg Welch (Ph.D., UNC-Chapel Hill, 1997) is a Research Associate Professor of Computer Science and Adjunct Associate Professor of Applied & Materials Sciences at the UNC at Chapel Hill. Before UNC, he worked at the NASA Jet Propulsion Laboratory, and the Northrop-Grumman Defense Systems Division. His research interests include human motion tracking systems, 3D telepresence, projector-based graphics, computer vision and view synthesis, and medical applications of computers.

Chumki Basu (Ph.D., Rutgers University, 2002) is a Member of Technical Staff at Sarnoff Co. She is a Program Manager in the areas of simulation and training, Internet-based computer vision applications, and

artificial intelligence. She is also a technical lead in the areas of knowledge representation and language analysis. From 1994-2001, she was at Bell Communications Research as a Research Scientist. Her interests include ontologies, multi-media information retrieval, text mining, machine learning, recommender systems, and user modeling.

Christian J. Darken (Ph.D., Yale University, 1993) is an Associate Professor of Computer Science at the NPS, and an academic faculty member of the MOVES Institute. His research focus is on cognitive and behavioral modeling for simulations. Previously he was Project Manager at Siemens Corporate Research. He was also a programmer on the team that built Meridian 59, the first 3D massively-multiplayer game.

Rakesh "Teddy" Kumar (Ph.D., University of Massachusetts at Amherst, 1992) is the Senior Technical Director of the Vision and Robotics Laboratory at Sarnoff Co. His technical interests are in the areas of computer vision, computer graphics, and robotics. At Sarnoff, he has been directing and performing on R&D projects in the areas of visual navigation and perception for robotics, video surveillance and monitoring, immersive tele-presence, simulation and training, 3D modeling, and medical image analysis.

Henry Fuchs (Ph.D., University of Utah, 1975) is the Federico Gil Professor of Computer Science and Adjunct Professor of Biomedical Engineering at UNC Chapel Hill. He's been active in computer graphics, rendering algorithms, hardware, virtual environments, tele-immersion systems and medical applications. He is a member of the National Academy of Engineering, a fellow of the American Academy of Arts and Sciences, the recipient of the 1992 ACM-SIGGRAPH Achievement Award, and the 1997 Satava Award.

Hui Cheng (Ph.D., Purdue University, 1999) is a Technical Manager of the Adaptive and Cognitive Systems Group at Sarnoff Co. Prior to joining Sarnoff, he was with the Digital Imaging Technology Center, Xerox Co. His research interests include image/video analysis and event detection, machine learning and statistical modeling.

Jan-Michael Frahm (Ph.D., Christian-Albrechts University of Kiel, Germany, 2005) is a Research Assistant Professor of Computer Science at the UNC at Chapel Hill. He worked on single/multi-camera systems for static and dynamic scenes (creation of 3D models of the scene); real-time multi-view; robust and fast estimation methods from noisy data; high performance feature tracking; and the development of data-parallel algorithms for commodity graphics hardware for efficient 3D reconstruction.

Mathias Kölsch (Ph.D., University of California, Santa Barbara, 2004) is an Assistant Professor of Computer Science at the NPS in Monterey, CA. He is also affiliated with the MOVES Institute, Chair of the MOVES Academic Committee and the Academic Associate for the MOVES curriculum. His research interests include computer vision, hand gesture recognition, augmented and virtual environments, sensor networks and mobile/embedded computing.

Neil C. Rowe (Ph.D., Stanford University, 1983) is a Professor of Computer Science at NPS. His main research interest is the role of deception in information processing, and he has also worked on surveillance systems, image processing, multimedia databases, robotic path planning, and intelligent tutoring systems. He is the author of over 140 technical papers and a book.

Herman Towles (M.S., University of Memphis, 1972) is a Senior Research Engineer in the Department of Computer Science at UNC-Chapel Hill. His research interests include multi-projector display, telepresence, and graphics/vision processing on GPUs.

Juan Wachs (Ph.D., Ben Gurion University of the Negev, Israel, 2007) is an Assistant Professor at the Industrial Engineering Department at Purdue University. Prior to taking this appointment in 2009, Dr. Wachs was a postdoctoral fellow at the NPS. His research interests include computer vision, hand gesture recognition, robotics control, healthcare and assistive technologies.

Anselmo Lastra (Ph.D., Duke University, 1988) is a Professor and Chair of the Department of Computer Science at the UNC at Chapel Hill. In the past he worked on the Pixel-Planes and PixelFlow machines. His research interests are in the areas of image-based modeling and rendering, and graphics hardware architectures. Dr. Lastra is currently tutorials editor of IEEE Computer Graphics and Applications.

New Generation of Instrumented Ranges: Enabling Automated Performance Analysis

Amela Sadagic[1], Greg Welch[2], Chumki Basu[3], Chris Darken[1], Rakesh Kumar[3], Henry Fuchs[2], Hui Cheng[3], Jan-Michael Frahm[2], Mathias Kolsch[1], Neil Rowe[1], Herman Towles[2], Juan Wachs[1], Anselmo Lastra[2]

(1) Naval Postgraduate School (NPS) Monterey, CA	(2) University of North Carolina (UNC) Chapel Hill, NC	(3) Sarnoff Corporation Princeton, NJ
asadagic@nps.edu, cjdarken@nps.edu, kolsch@nps.edu, ncrowe@nps.edu, jpwachs@nps.edu	welch@cs.unc.edu, fuchs@cs.unc.edu, jmf@cs.unc.edu, herman@cs.unc.edu, lastra@cs.unc.edu	cbasu@sarnoff.com, rkumar@sarnoff.com, hcheng@sarnoff.com

INTRODUCTION

This paper describes our US Navy/Office of Naval Research (ONR)-sponsored project on a new generation of instrumented ranges and the early results we have achieved; the project is titled "Behavioral Analysis and Synthesis for Intelligent Training" (BASE-IT). The approaches and solutions suggested in this research effort represent a paradigm shift in terms of the type of the data collected on the physical training ranges, as well as the way those data are processed and used both on the ranges and outside of that realm.

Problem Description and Motivation

Military training conducted on physical ranges that matches a unit's future operational environment as closely as possible, provides an invaluable experience. Today, to conduct a training exercise in an urban environment while ensuring a unit's performance is closely observed, evaluated, and reported on in an After Action Review (AAR), the unit requires a number of highly experienced instructors to accompany different segments of the unit throughout their training run. Additionally, as a form of better familiarization with the training environment and the complexity of actions they will encounter, the unit receives an introductory package for the training they are just about to conduct. After the completed training, the unit receives AAR and a take-away package as an illustration of their performance on the same range.

The current approach, which is practiced on a number of military training ranges, is to use a large number of skilled instructors who accompany the unit in its training. This poses a problem, however: the number of specialized instructors available to assume this role is always limited. For example, a unit may want to conduct the training where no special instructor—apart from its own staff—may be available. As an introductory package the units receive textual information, maps, charts and images, and may also get videos of another units' training runs. Such information has limited power and is unlikely to be an optimal illustration of all topics on which the unit needs to be informed. This type of information is hard to browse (it requires multiple hours of careful scrutiny), and the review is inherently a rather passive activity, involving minimal human interaction with the material. If the training range is equipped with a camera system, as a take-away package the unit will receive multiple videos showing the training run it just completed. Each video may be several hours long. This again poses a problem, as no unit can dedicate the necessary number of hours to review this information. In fact, it may also turn out that the same videos did not actually contain information of interest to the unit. Reviewing the videos (essentially a linear type of data representation) has another inherent problem: it is nearly impossible to resolve which recorded events happened simultaneously, which is essential in order to understand the coordination of actions done by different segments of the unit.

To address all those issues we suggest a radically different concept: instead of a traditional approach of recording multiple video streams that need to be reviewed in their entirety by a number of instructors and then be evaluated, our system will focus on technology-supported capturing of non-linear information about exhibited performances based on dynamic individual warfighter poses (3D position, and torso, head and weapon orientation) and posture (standing, kneeling, prone) data. This type of data enables us to apply an automated performance analysis and evaluation that requires minimal to no human

intervention; as a result the automated system provides a set of quantitative metrics about the recorded performances and skills that were trained.

Our team looked for a technical solution that:
  (1) is omnipresent—it should be 'everywhere' and see everything/everyone,
  (2) has a perfect memory—it should record everything, and
  (3) produces and delivers relevant information in a timely fashion and in a format that is useful to both the instructors and the warfighters.

Additionally, we are interested in extending the value of the physical training range outside of that realm, and using the collected data to affect the training that comes before and after the range-based training. This approach ensures that a full training circle is addressed with the family of related products. In this context, we included the ability to explore the dynamic 3D data in an interactive fashion, allowing the warfighters to visualize the data sets as well as to conduct fast smart searches and look for events of special interest. For example, Which squad was 'bunching-up' while moving through the range? Did Marines cross a danger zone in 'ones' and 'twos', and how fast were they?

TRAINING RANGES FOR URBAN WARFARE

Special attention was paid to domain analysis. Because our goal is to make a solid connection between the needs of operational forces and the scientific work that our team would propose, it was an absolute prerequisite to acquire a thorough understanding of the domain. We conducted several field visits to different training sites where we observed the training 'up-and-close' and collected assorted sets of data. A number of discussions with the Subject Matter Experts (SMEs) were also

Modeling Virtual Environments and Simulations (MOVES) Institute, NPS, also proved to be invaluable. The environment where urban warfare training is conducted should provide a realistic, accurate and relevant environment. If the environment is designed in this way, it provides the troops with optimal conditions for highly effective training of their warfighting skills. The basic characteristics of this environment are: (1) it is a large environment to enable joint training of multiple agencies and the engagement of large units; (2) it is a highly occluded environment—the streets and passages can be very narrow; (3) there is a large number of buildings, some of them representing multistory houses; (4) there is a large number of individuals (unit members and role players) and a certain number of vehicles, all of which need to be observed and have their performance evaluated; and (5) in addition to buildings, the environment also has the urban and battlefield clutter, both important for the level of realism of such a site. All of this suggests that a large number of human instructors is usually needed to observe all warfighters' actions and provide expert opinions on their training.

The buildings are usually simulated by metal cargo containers, some of them stacked on top of each other to represent multistory buildings. Figure 1 illustrates several typical views inside such a training range. The spaces between the containers simulate the streets, junctions, courtyards, markets and village squares. Inside the containers are the rooms and facilities that represent typical house environments to enable segments of units' training focused on indoor urban warfighting skills. The same training range is used to conduct different training courses that each may last from 30 minutes up to several days, and include both kinetic and non-kinetic warfighting skills.


Figure 1. Typical Views Inside the Training Range for Urban Warfare

organized, and we had a continuous dialog with both active and retired USMC officers, some of them directly provided to us by our transition customer, The Program Manager for Training Systems (PM TRASYS). The experiences and insights provided by the USMC officers who are currently students at the

SYSTEM ARCHITECTURE

The BASE-IT system architecture is organized around a central database that connects all other components (see Figure 2).
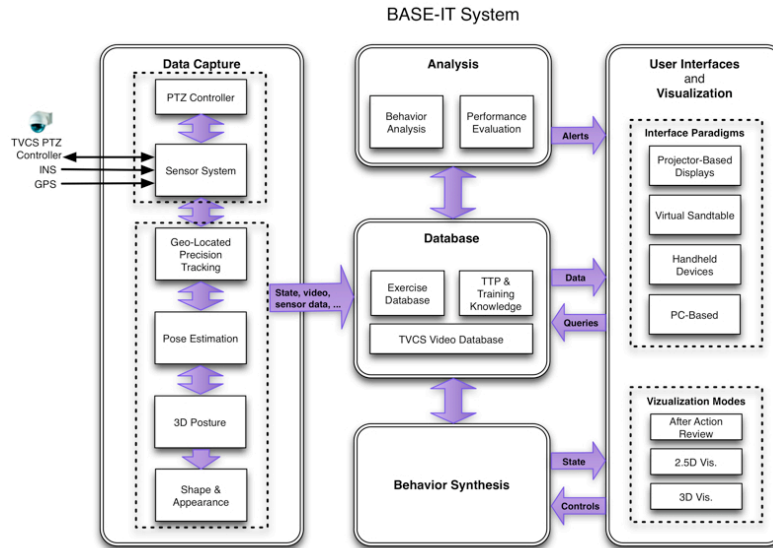
BASE-IT System

Figure 2. BASE-IT Architecture

In addition to the database, there are four other major blocks: data capture, behavior analysis, behavior synthesis, and visualization.

Each block of the system can be easily understood in terms of the data that it consumes or produces for the central database. The data capture block is responsible for producing data on exercise participant position, orientation, and posture. This includes camera control and fusion of Global Positioning System (GPS) and camera data. The behavior analysis block reads participant data from the database, and then uses it to generate performance assessment data, which is then returned to the database. The behavior synthesis block supports the user with the exploration of alternative courses of action. It can read an initial situation from the database, and then simulates the rest of an exercise under the interactive control of the user. The visualization block reads exercise data from the database and renders it in an accessible form on various target devices, including fixed and handheld PCs.

A database-centric architecture has both pros and cons. Its attractive features are several, and for us, ultimately compelling. It provides enhanced modularity of the system as a whole, yielding dividends in ease of development (particularly over multiple physical development sites), ease of managing (browsing, storing) system data, flexibility in adding new sensors and support for exercise playback. This helps us fulfill our goal of realizing a prototype that is extensible, for example to allow additional or new/unforeseen sensors, new analysis methods, and new interfaces. The primary downside of the architecture is the (small) added latency associated with database reads and writes.

DATA ACQUISITION FRAMEWORK

An essential component of the overall BASE-IT system is a data acquisition framework consisting of an in situ network of automatically controlled pan-tilt-zoom (PTZ) video cameras, and personal position and orientation sensing devices—GPS and Inertial Navigation Systems (INS) devices. In this section, we describe data inputs to this system, sensor devices, and data capture and processing.

Data Inputs

The data needed for our system includes (1) the positions of the Marines; (2) the poses of the Marines; (3) the poses of the Marines' weapons or equipment; (4) the skeletal postures of the Marines; (5) video coverage for position, pose tracking and situational awareness; and (6) time stamps for each of the aforementioned items. It is important to note that our system does not record auditory information (e.g., voice commands), nor does it record weapon deployment data or use of other military assets (e.g., air).

Sensors

Our sensor system for data capture includes (1) GPS and video for positions; (2) IMU and video for poses of Marines; and (3) PTZ video cameras for skeletal posture estimation. In our framework, to capture video, we use the Bosch G4 series of PTZ cameras. The GPS unit is a low cost and wearable off-the-shelf GPS system. The GPS units also serve to provide an ID for

each Marine wearing them. In our experiments, we use the GlobalSat BU-353 GPS unit, which has a WAAS enabled 2dRMS of 5 meters. We are also experimenting with the Intersense Intertial Cube2 INS units for orientation calculations.

Data Capture and Processing

In Figure 3 we show a diagram of our data acquisition framework (outlined box) and data processing components.
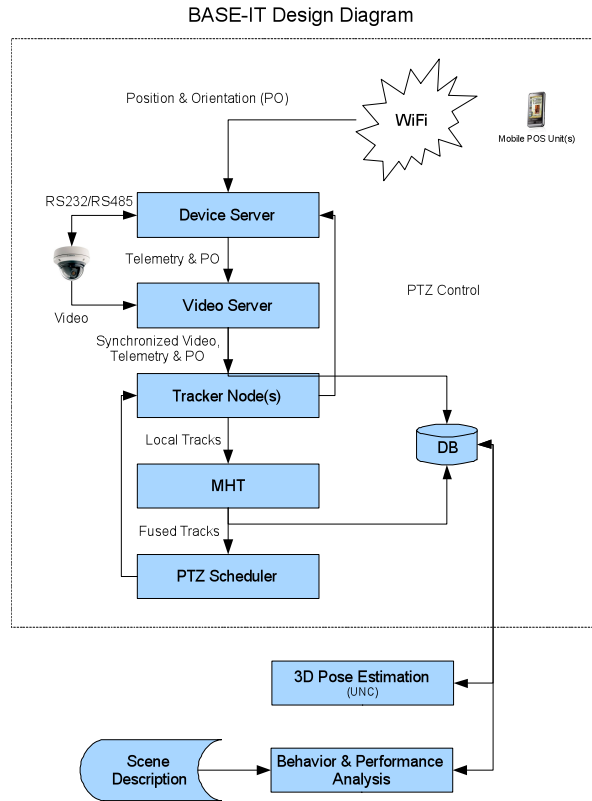
Figure 3. BASE-IT Data Capture and Processing Diagram

Data Flow

The central Device Server collects telemetry data from the PTZ cameras, and position and orientation data from GPS & INS units. The Device and Video Servers output data to the Tracker Nodes, or the 3D posture estimation module, and the raw data is stored in the database. A Tracker Node performs GPS assisted visual tracking and pose estimation of Marines in a designated camera/view. Each Tracker Node, outputs local tracks to the Multiple Hypothesis Tracker (MHT) module. Here local tracks are fused to obtain global tracks consistent with all available views using MHT. The global tracks are stored in the database and sent to the PTZ Scheduler. The PTZ Scheduler is responsible

for deciding when cameras should capture images and when they should maneuver to new pan, tilt and zoom settings. The Device Server receives PTZ control commands and sends these to the cameras. For areas observed by multiple cameras simultaneously the data are fused into a 3D body posture of the observed Marines (see Figure 4).
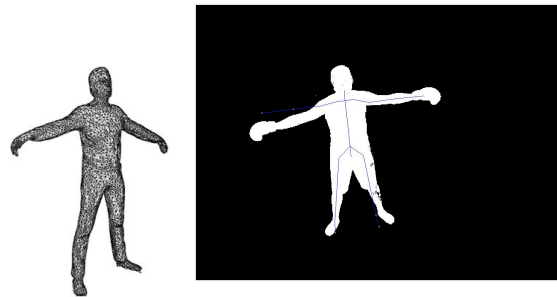
Figure 4. 3D Model and Associated Skeleton (Posture) Obtained From Multiple PTZ cameras

PTZ Camera Control

We examined existing approaches used in surveillance, extracted useful elements, and developed a new scheduling approach that is better suited for 3D reconstruction. Application-specific requirements include higher image resolution, longer capture intervals, more cameras capturing the same target, and geometric calibration. To satisfy these requirements, we are developing an approach consisting of a *global optimization method* which assigns cameras to target and a *local scheduling method* which schedules pre-assigned cameras for optimal coverage. Both components make use of a quality metric that characterizes and predicts the performance and usefulness of possible camera configurations. We have implemented the local scheduling method and tested it in the simulated environment shown in Figure 5; a top-down view shows six color-coded cameras and a visibility map that takes into account static occluders represented by three vertical columns.
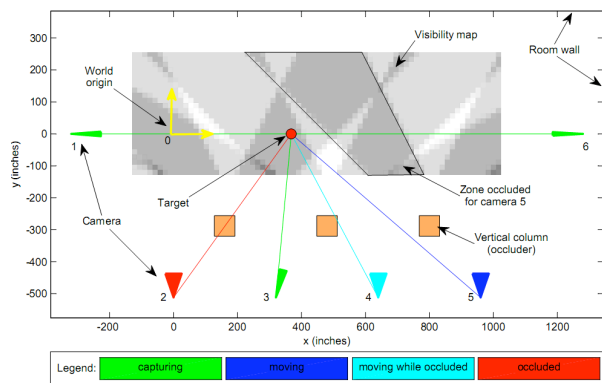
Figure 5. Top-Down View of the Simulated Environment

Our scheduler uses a pre-computed visibility map and the predicted target trajectory and schedules camera maneuvering and capturing intervals such that a quality metric is maximized. Figure 6 shows an example schedule obtained using our local scheduling method. Green indicates cameras that are capturing (i.e. green is good).
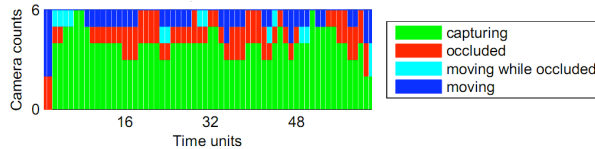
Figure 6. Example Schedule for 6 Cameras Following a Single Target

BEHAVIOR ANALYSIS

Introduction

The overarching goal of behavior analysis is to provide a fine-grained understanding of movements and actions undertaken by individuals and groups of warfighters. Behavior analysis can be decomposed into the following main components: (1) maintaining a library of known behaviors in an ontology; (2) detecting and grouping behaviors observed by the system into known classes (e.g., states) and scoring the accuracy of detection or classification; (3) understanding the body posture of the participants to improve classification; and (4) assessing and evaluating performance for after-action review (AAR).

Behavior Ontology

Our aim is to develop an extensible ontology for simulation and training focused on urban warfare scenarios, yet built on general-purpose principles. The ontology is a knowledgebase that captures USMC Tactics, Techniques and Procedures—TTP (including definitions and measurements) for dismounted small team behaviors, maneuvers, procedures, and rules of engagement. The ontology is not restricted to a single representational formalism.

The ontology also encodes the decomposition of states and events that are necessary for detection and recognition. For example, a higher-level state such as "Patrolling" can be decomposed into a number of sub-states that define procedures for "Crossing danger areas" such as intersections. We use a finite state machine (FSM) representation of states as part of the behavior analysis architecture and show a top-level FSM in Figure 7.
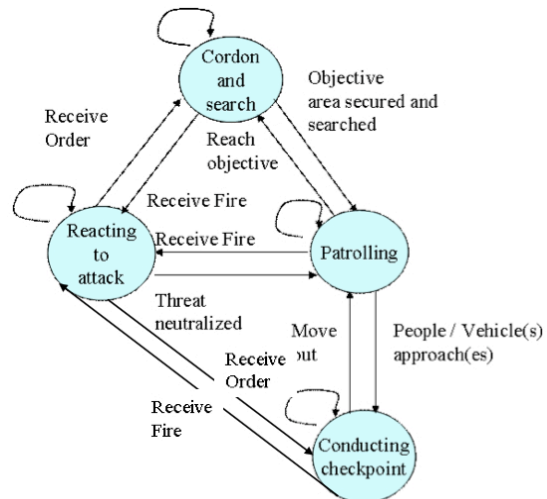
Figure 7. Top-level Finite State Machine

Detection of States and Events

One of the challenges of state or activity detection is modeling inter-object relationships along with the participants involved. We train Support Vector Machine (SVM) classifiers to detect activities or states. Example activities include military patrols moving in formation. For the patrol formation, we can recognize one of n classes such as single-column formation, staggered column formation, wedge, etc. An example of a staggered column formation is shown in Figure 8.
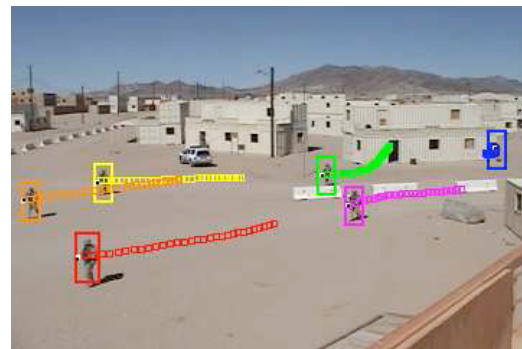
**Figure 8. Example of a Patrol Formation**

Activities are detected or classified at different levels in the system. For example, to recognize low-level activities, the system combines information from video with INS (e.g., weapon) with low-level pose information to classify activities performed by warfighters. Context is also important for behavior analysis, e.g., the position of the warfighter with respect to features of the environment such as buildings. By composing low-level actions, we are able to recognize complex activities.

Posture Recognition

Behavior analysis of the human almost always requires knowledge of the body posture, including overall stance, orientation (Figure 9), head gaze direction (Figure 10) and hand/arm poses. We built a real-time system using various features and classifiers that analyze the live video streams with known position estimates for the Marines (Wachs et al., 2009). Stance and torso direction is estimated both with a Viola-Jones style multi-class detector (Viola et al., 2001) and with a parts-based detector using full-body appearance models (Torralba et al., 2007). These estimates act as priors for detecting heads, weapons, and their respective orientations with similar algorithms. Syntactic and statistical post-processing improves the accuracy and reliability of the posture recognition.

Figure 9. The Three Distinguished Stances in Two Orientations Each
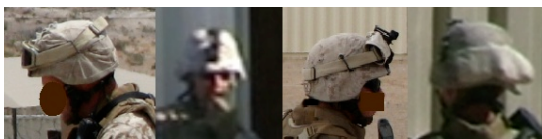
Figure 10. The Four Detected Head Orientations

Performance Evaluation

Automatically tracking warfighters using camera feeds during training enables the automatic assessment of performance; even the best instructors cannot see everything that happens. Important mistakes may be visible only briefly, may be occluded from the instructors' view by buildings, or several things may happen simultaneously, ultimately making it difficult or impossible to notice everything.

In our system, there are two categories of performance evaluation. The first is the evaluation of state estimation (see "Detection of States and Event"). The second is the evaluation of performance measures computed across states during an exercise. In this paper, we focus on the latter.

Multiple performance measurements are recorded by our system. For example, there are certain measures that are associated with all states that the Marines can be in during an exercise. These would include "360-security," "weapon flagging," etc. Every state also has a set of performance measures that provide finer levels of detail on how well the actions were performed. For example, did the Marines "pie" off a window or door as they passed it? For states that involve a combination of actions, some combination of performance measures may be necessary to generate an overall assessment. In this section, we present details of some of these performance measures and trends.

Our approach is to extract positions and orientations of the Marines by correlating data from the cameras and GPS units. We then automatically calculate key measures of performance at regular time intervals. The measures are based on Marine documents (U.S. Marine Corps, 1998; U.S. Marine Corps, 2005), discussions with SMEs, and preliminary experiments. Some of these measures include: (1) dispersion—whether the Marines are sufficiently far apart; (2) collinearity—whether the Marines form a single column; (3) the number of clusters; (4) the number of conversation events with non-Marines; (5) the dangerousness of the position; (6) sniper awareness—Marine ability to see potential sniper position from his location; (7) mobility—ability to escape if necessary; (8) weapons danger (weapon flagging); (9) weapons coverage of threats; and (10) being too close to windows and doors.

We automatically aggregate these measures over time to get overall performance trends and note problems for after-action review (Hixson, 1995). Aggregation is done separately for units, the type of exercise, and the type of behavior involved in the exercise (receiving orders, patrol, cordon search, manning a checkpoint, etc.). Aggregation calculates the average, the minimum, the maximum, and counts in three different ranges for each of the above performance measures. Values sufficiently far from norms will be flagged for review. For instance, we will flag all times that situation awareness was in the 10th percentile for a unit during an exercise, all units whose average situation awareness was in the 10th percentile, and all exercises for which the average unit was in the 10th percentile. This will enable instructors to quickly find significant problems in training and identify their likely causes.

BEHAVIOR SYNTHESIS

The ability to play back the performance that was recorded, and to comment on it, is very valuable in the

training of every Marine. In order to provide new training opportunities we introduced an additional feature—the ability to pause a playback, change the course of action (e.g., the direction in which simulated Marines would be moving, the number of Marines), and generate "what-if" scenarios on the fly, i.e. to support a "free play" mode.

This functionality is provided by a module called *behavior synthesis*. The goal of this module is to provide sufficiently realistic behavior to enable the user to substitute new behavior for what was actually captured in a particular exercise, in the interest of determining how the outcome would have changed. The user commands the Marines using a Graphical User Interface (GUI) intentionally similar to that used in real-time strategy games. The details of the behavior synthesis module would be familiar to those aware of best practices in the military simulation and video game industries, with the exception of innovations in the modeling of threat and the pathfinding algorithm.

Threats that are not yet confirmed are usually not modeled in simulations. However, most of the time, in even the hottest battles, Marines act in response to where the enemy *might* be, rather than where they can actually be observed at the present time. We explicitly model the likelihood of a threat being at a given position by a statistical model—the Heatmap codes illustrate subjective probability that a threat occupies a position (see Figure 11). The model drives where the simulated Marines look at any given time, and in turn is updated based on where they look. In other words, unless a target is actually spotted, the threat probability in the viewed region is reduced.
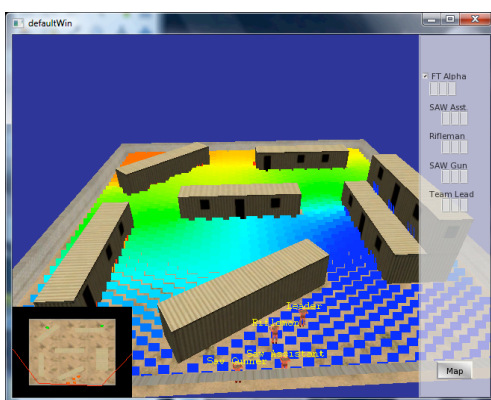

Figure 11. The Heatmap Codes

Pathfinding has received a lot of attention in recent years, particularly in the video game industry. We take the novel approach of using a conventional A* path for a fireteam as a whole (see the white line in Figure 12; the colored lines are paths of individual Marines), and

then carrying out individual greedy searches for each Marine. While the fireteam path minimizes distance and maximizes cover, the individual Marine paths optimize additional factors such as fireteam dispersion and distance from fireteam leader.
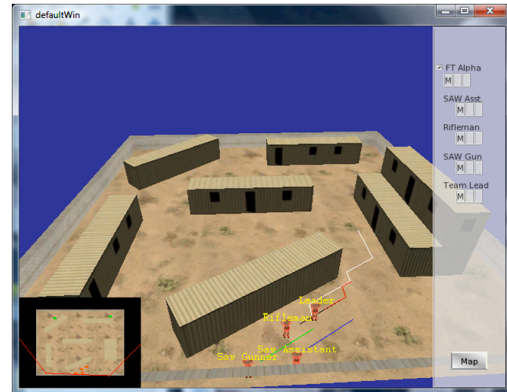

Figure 12. Fireteam and Individual Marine Paths

VISUALIZATION

The rich data in the BASE-IT database only becomes useful once it is visualized. In addition to the positions, orientations and pose of the exercise participants, the results of behavior analysis must be visualized.

3D rendering is accomplished using the open-source Delta3D simulation/game engine. Delta3D provides several essential visualization capabilities such as a scene graph complete with shader capabilities including support for shadows; a human character animation capability including hardware skinning; and standard motion models including one inspired by real-time strategy games. Additionally, providing DIS or HLA networking should be possible if required in the future.

Most of the technology in the PC visualization component is familiar to those acquainted with the state of the art in the development of 3D military simulations or video games. However, it was necessary to innovate to produce an inexpensive method for driving the gaze and aim of the simulated characters (see Figure 13). We developed a novel technique to control gaze and aim via a constant-time algorithm for generating blend weights (Figure 14), which are then fed to a conventional blending capability in Delta3D's character animation subsystem.

Because we are using cameras to observe individuals as they actually look, using that video information is sometimes preferable to a 3D reconstruction (simulated Marine). To enable this we provided 2.5D visualization of the data. The name 2.5D is appropriate, because

while the environment geometry in the visualization is true 3D, the characters are video, i.e. 2D.

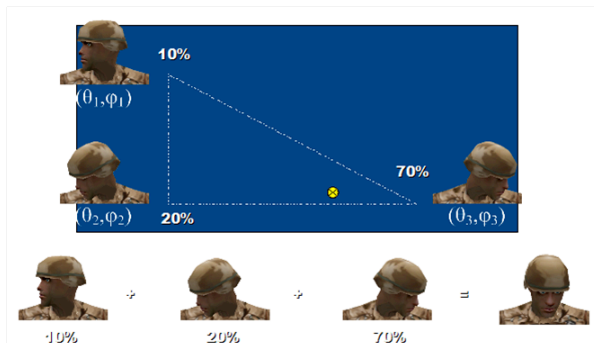

Figure 13. Marine Gaze is Modeled in Detail

Figure 14. Constant-Time Gaze or Aim Direction

The 2.5D system, called MapView, provides for immersive coverage and intelligible embedded display of video and auxiliary data annotations over a wide area of both urban and open terrain during AAR, including events, and performance metrics. The MapView approach is to project multiple videos onto a site model, enabling interpretation of visual action in the global context of the model. We also display symbolic activity and an event view, showing tracks and events for all objects on line in real time. A graphic of our MapView is presented in Figure 15.

DISPLAYS AND USER INTERFACES

Our BASE-IT system includes several display platforms to facilitate individuals and group tasks during the various phases of training. For pre-operation briefings or AAR done indoors, we have developed *seamless multi-projector displays* that can be rapidly set up in a variety of indoor spaces. For pre-operation instruction and AAR we have also developed a projector-based *virtual-physical sand table display*. Finally, we have identified and will be developing

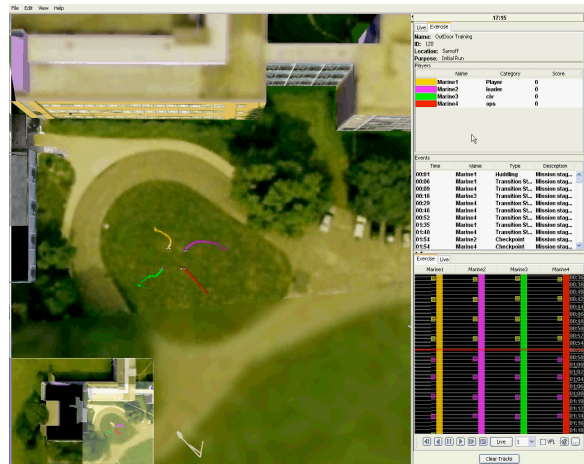*hand-held interfaces* for instructors during field exercises.

Figure 15. MapView

Seamless Multi-Projector Displays

While projection-based displays have long been used to create large-format displays, only in the last decade have researchers developed camera-based calibration techniques that can turn the arduous task of display setup and maintenance into the mundane (Raskar et al., 1998; Surati, 1999). During this period, projective display research at UNC has focused on improving both calibration and rendering techniques for creating ad hoc multi-projector display systems that are compact, rapidly deployable, and robust.

We feel two keys in achieving these goals are the ability to display on the surfaces of ordinary room walls (no need for special display screens) and the concept of an *intelligent projector unit* (IPU) that combines the projector, calibration cameras, and the image generator into a single module. We have demonstrated the use of prototype IPUs (Figure 16 (a)) to create seamless immersive display on walls and across the corners of an ordinary room as shown in Figure 16 (b)–(c).

To make such ad hoc displays more operationally robust, we have developed a distributed framework for continuous calibration (Johnson et al., 2009) that runs concurrently with the application to make the display robust to small perturbations in projector pose, thus reducing the need to stop and re-calibrate.

A Virtual-Physical Sand Table

Physical "sand tables" have been used in the military for many years (Bond and Crouch, 1922). Their advantages include simplicity and a natural means for

conveying tactical plans and reviews related to the positions and movements of assets over time. Examples in use today for training include setups that replicate towns using fixed wooden blocks for buildings, and paint or markers to indicate roads and walls.

Our BASE-IT system inherently includes the digital training content to facilitate a modern version of the traditional table. To accomplish this we employ digital projectors and *Shader Lamps* techniques (Raskar et al., 1999; Raskar et al., 2001) to render textures, tracks, moving icons, and training-related information onto a physical model of the site. We have developed a transportable prototype at UNC, shown in Figure 16 (d)–(g), and a duplicate to be used at NPS. The design involved managing constraints such as the number of projectors, pixel densities on the various surfaces, accessibility to multiple users, transportability and size.

Handheld Interfaces (Future Work)

In addition to a display system for groups, we have identified a need for a small, portable device for use by the instructors during training. We envision a networked handheld device capable of displaying video and 3D graphics—perhaps an Ultra-Mobile PC

(UMPC) or even an iPhone or Android phone. Uses for the device include live annotations by the trainers, communication with other trainers (with the ability to send annotated images), and virtual views (either live video from cameras or synthesized views using 3D graphics). Adding pose estimation capability to the handheld will enable augmented-reality overlays so the trainer can see information about the locale, for example that there is a squad of Marines behind a building or that a pile of roadside rubbish is booby trapped.

CONCLUSIONS

We are now almost two-thirds of the way through a three-year effort aimed at realizing the demonstration of a vision for an integrated and automated system. The effort involves experts in training practices, human factors, behavioral analysis, computer graphics, and computer vision; along with collaborators in the USMC and support/leadership from the Office of Naval Research. We recognize that the realization of such a long-term vision is a challenging goal, both in terms of the new algorithms that need to be developed as well as numerous technical issues that need to be resolved in such a complex system. However we are optimistic about the results so far, and we look forward to our
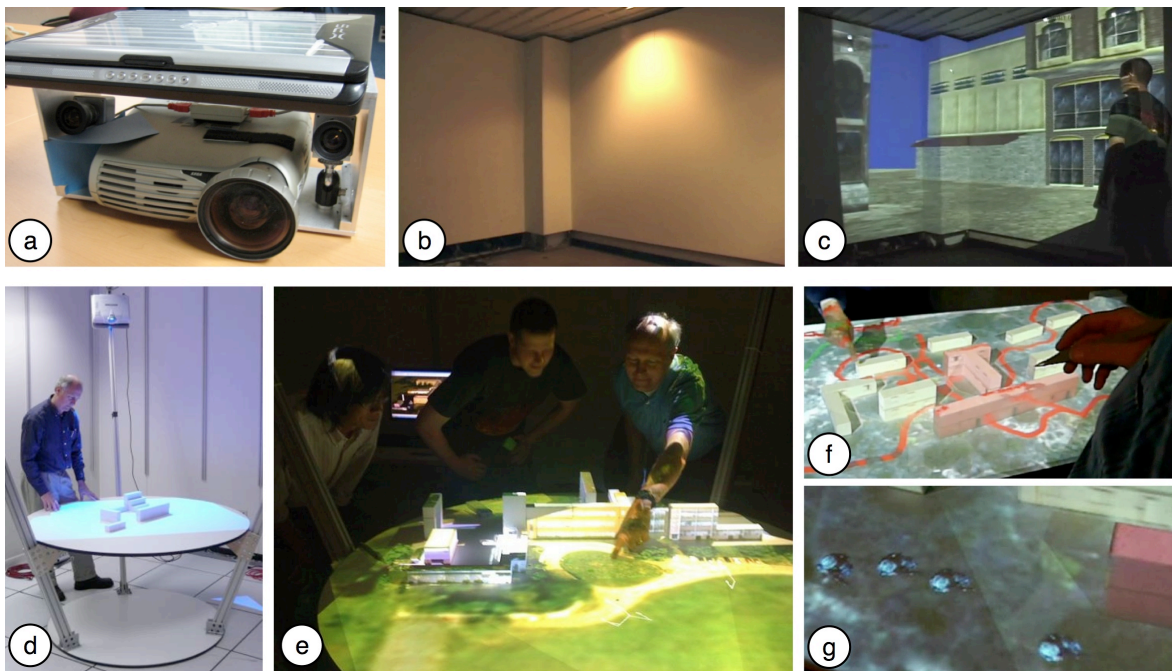
Figure 16: Current display system research-related prototypes. (a) A projector, two cameras, and a laptop computer combined into an "intelligent projector unit" or IPU. (b) The corner of a room that includes a support pillar. (c) We use an IPU to render imagery that is corrected for the unusually shaped display surface. (d) Our projector-based sand table prototype, showing only the white surfaces of the physical model of a training site. (e) The sand table system being used to visualize the Sarnoff facility in Princeton, NJ. (f) The demonstration of an interactive annotation/marker capability. (g) A demonstration showing Marine icons.

eventual data capture and system demonstrations involving actual training exercises.

Today, the dual goals of maximizing the potential that physical training ranges have, and making sure the associated training is the most effective, are of paramount importance to the military in its determination to train to fight and fight to win. We are confident that striving to bring different sensor systems and increasing the level of automation of the sensor manipulation and performance analysis in such environments is the best way to ensure that the future training ranges respond to such a high goal.

While in this system we incorporated only three types of sensors (cameras, GPS and INS), in the future the systems for instrumented ranges will also incorporate sensors that capture audio (commands and shouts), weapon deployment and other information pertinent to unit activity.

We also believe that the fundamental methods suggested in this project could be applied to other types of training ranges, including the portions of full combined arms exercises, for example. And beyond training, we believe long-term potential applications include port and airport security, surveillance, homeland security, and intelligence operations.

ACKNOWLEDGEMENTS

REFERENCES

Bond, Paul S. and Crouch, Edwin H. (1922). Tactics: The Practical Art of Leading Troops in War, *The American Army and Navy Journal, Inc., 1922.*

Darken, C., Anderegg, B. & McDowell, P. (2007). Game AI in Delta3D, *Proceedings of Computational Intelligence in Games (CIG) 2007.*

Guerrero, M. & Darken, C. (2008). Blended Inverse Kinematics: Delta3D System Utilization, *Proceedings of the Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC) 2008.*

Hixson, J. (1995). Battle Command AAR Methodology: A Paradigm for Effective Training, *Proc. Winter Simulation Conference*, pp. 1274-1279.

Johnson T., Welch G., Fuchs H., Laforce E., and Towles H. (2009). A Distributed Cooperative Framework for Continuous Multi-Projector Pose Estimation. *proceedings of IEEE Virtual Reality 2009*, Mar 14–18, 2009.

LaVoie, N., Foltz, P., Rosenstein, M., Oberbreckling, R., Chatham, R., and Psotka, J. (2008). Automated Support for AARs: Exploiting Communication to Assess Team Performance, *The Interservice/ Industry Training, Simulation, and Education Conference*, December 2008.

Raskar, R., Welch, G., Low, K.-L., & Bandyopadhyay, D. (2001). Shader Lamps: Animating Real Objects With Image-Based Illumination, Rendering Techniques 2001, *Proceedings of the Eurographics Workshop in London,* United Kingdom. S. J. Gortler and K. Myszkowski. University College London (UCL), London, England, Springer, NewYork: 89-102

Raskar, R., Welch, G., and Chen, W.-C. (1999). Table-Top Spatially-Augmented Reality: Bringing Physical Models to Life with Projected Imagery, *Second International Workshop on Augmented Reality (IWAR'99),* October 20-21, 1999 San Francisco.

Torralba, S.A., Murphy K.P., and Freeman W.T. (2007). Sharing visual features for multiclass and multiview object detection, IEEE PAMI, 29:5, pp. 854-869.

U.S. Marine Corps (1998). *Military Operations on Urbanized Terrain* (MOUT, Marine Corps Warfighting Publication MCWP 3-35.3.

U.S. Marine Corps (2005). *Infantry Training and Readiness Manual*, Directive NAVMC DIR 3500.87, September 2005.

Viola, P. & Jones, M. (2001). Robust Real-time Object Detection, *Second Intl Workshop on Statistical and Comp. theories of Vision*, Vancouver, 2001.

Wachs, J.P., Goshorn, D., & Kölsch, M. (2009). Recognizing Human Postures and Poses in Monocular Still Images. *Intl. Conf. on Image Processing, Computer Vision, and Pattern Recognition (IPCV)*, Las Vegas, NV 2009.