Original article

Journal of Defense Modeling and Simulation: Applications, Methodology, Technology
10(3) 327–342

DOI: 10.1177/1548512912472942
dms.sagepub.com

Smart instrumented training ranges: bringing automated system solutions to support critical domain needs

Amela Sadagic[1], Mathias Kölsch[1], Greg Welch[2], Chumki Basu[3], Chris Darken[1], Juan P. Wachs[4], Henry Fuchs[2], Herman Towles[2], Neil Rowe[1], Jan-Michael Frahm[2], Li Guan[5], Rakesh Kumar[6], and Hui Cheng[6]

Abstract

The training objective for urban warfare includes acquisition and perfection of a set of diverse skills in support of kinetic and non-kinetic operations. The US Marines (USMC) employ long-duration acted scenarios with verbal training feedback provided sporadically throughout the training session and at the end in a form of an after-action review (AAR). The inherent characteristic of training ranges for urban warfare is that they are the environments with a high level of physical occlusion, which causes many performances not to be seen by a group of instructors who oversee the training. We describe BASE-IT (Behavioral Analysis and Synthesis for Intelligent Training), a system in development that aims to automate capture of training data and their analysis, performance evaluation, and AAR report generation. The goal of this effort is to greatly increase the amount of observed behavior and improve the quality of the AAR. The system observes training with stationary cameras and personal tracking devices. It then analyzes movement and body postures, measures individual and squad-level performance, and compares it to standards and levels of performance expected in given situations. An interactive visualization component delivers live views augmented with real-time analytics and alerts; it also generates a personalized AAR review in a three-dimensional virtual or mixed reality environment, indexed by automatically extracted salient events and accompanied by summary statistics of unit performance. The approaches presented in the system have the potential to radically change the analysis and performance assessment on physical training ranges and ultimately this type of training itself.

Keywords

Instrumented training ranges, multi-sensor systems, automated behavior analysis, behavior synthesis, computer vision, after-action-review, simulations

1. Introduction

Training realism and ability to provide a comprehensive feedback about the performances observed in training, and advise on skill remediation, are the imperatives to training effectiveness. Yet, to fully achieve all those elements costs significantly more time and money. We describe our efforts to increase training effectiveness at one of the most heralded yet most expensive training installations, through system-supported better preparation, better trainee observation, and better post-exercise review capabilities.

Undeniably, conducting a unit's training in conditions similar to in-theater operating conditions is of utmost

[1]Naval Postgraduate School, Monterey, CA, USA
[2]University of North Carolina at Chapel Hill, Chapel Hill, NC, USA
[3]IBM India Research Labs, Bangalore, India
[4]School of Industrial Engineering, Purdue University, West Lafayette, IN, USA
[5]GE Research
[6]SRI International Sarnoff

Corresponding author:
Amela Sadagic, Naval Postgraduate School, 700 Dyer Rd, WA-283, Monterey, CA 93943, USA.
Email: asadagic@nps.edu

importance for any unit's military training. This is especially true for urban warfare operations where a number of individual and team skills need to be integrated and practiced. A unit's commanding cadre needs to perfect their tactical decision-making and leadership skills when commanding their unit, skills focused on integrating other military assets in ground operations (e.g. air), and coordinate the operations of their unit with other units. Marines themselves need to integrate their skills and knowledge of military tactics and procedures, apply unit-specific Standard Operating Procedures (SOP) while being physically exerted when moving within environment with full combat gear and extreme weather conditions. No virtual simulation even comes close to providing and enabling such training conditions, and so the last stages of pre-deployment training do rely on extensive use of physical training ranges for their 'run' training phase ('crawl-walk-run' staged training regimen).

Another important characteristic of military training is that it is happening in ever-changing landscape of contemporary warfare, making it an increasingly complex job and forcing all services to continuously analyze and update their training procedures. One of the most critical steps in preparing today's ground troops is their preparation for a highly demanding and complex spectrum of operations in urban warfare. The Behavioral Analysis and Synthesis for Intelligent Training (BASE-IT) project funded by the Office of Naval Research (ONR) represents one of the efforts that the US Navy has established with the goal to acquire advanced technologies for the 21st century and in support of the US Military Corps' (USMC) current and future training needs. The focus of this project is to design and prototype the elements of an intelligent instrumentation system for physical training ranges that maximize the value of training hours on physical ranges and provide a set of tools for all participants for before, during and after such training. The three institutions that have been involved in this research effort are the Naval Postgraduate School, SRI International Sarnoff Corporation and the University of North Carolina at Chapel Hill.

Military operations on urban terrain (MOUT) have become one of the main focuses of today's fragmented warfare, presenting great difficulty, high complexity, and high demands on communications, and the need to acquire and maintain a holistic operating picture.[1] A good portion of a unit's pre-deployment training is dedicated to MOUT operations conducted on specially designed physical training ranges with simulated buildings, staged scenarios with varying levels of difficulty, human role players, a broad range of military assets and multi-unit interactions. This type of training is highly effective, however, the challenges on its organization and execution are numerous: (1) organizing and conducting the training on physical ranges requires complex and significant logistical and material resources, (2) training ranges are heavily used and shared by a number of other units, and so the time one unit can spend on the range is limited, (3) the sheer scale of the training environment consisting of hundreds of buildings, many individuals and vehicles that military instructors need to observe, as well as plenty of visual occlusion as is typical for urban environments, are the main reasons why some performances and events may be missed and not evaluated (reported), (4) after-action review (AAR) usually consists of the instructor's brief account and a critique of unit's performance on the range, supplemented with a small number of quantitative data the instructors were able to record during the exercise; while the unit is still on the range there is no time for a detailed AAR. Additionally, after the training is completed, the unit receives a take-away package (Take Home Package (THP)) in form of video footage,[2] yet this information is either too long (demanding several hours of watching and evaluation by instructors and trainees; this time is rarely if ever available to the unit and as a result the take-away package is rendered unusable) or it has only a small selection of scenes recorded by cameras, as they were selected by the operator who may have missed to see all events of great significance during that training session.

The BASE-IT approach uses a multi-sensor automated data collection, analysis and visualization of collected data sets to improve the units' preparation *before* arrival for training, to support instructors' decision-making process *during* training exercises, and to improve the unit's ability to conduct a detailed AAR and analysis of their performances *after* leaving the training facilities.[3] Early in our BASE-IT project, we invested a considerable amount of time analyzing a wide spectrum of training needs in this domain; the ultimate goal was to make sure that (a) project solutions are highly relevant to identified domain needs, (b) they represent a significant qualitative advance when compared with approaches and systems currently used in the same training environments, and (c) they provide a solid basis for the future upgrades and changes that may need to be introduced in response to the evolving training needs in this domain.

2. The problem space

Physical training ranges for urban warfare operations are characterized by a large number of simulated multi-story buildings, wide and narrow streets, town squares, open spaces, with the typical urban and warfare clutter distributed within the range. The active participants in any training exercise are multiple units that coexist with a number of constantly moving role-players who simulate a typical town life that any unit may encounter in an active military zone.

Typical scenes captured on training ranges during our field visits are shown in Figure 1.

Figure 1. Typical scenes on physical training ranges for urban warfare: (a) urban patrol; (b) providing security in cordon and search; (c) indoor site exploitation; (d) vehicle check point (VCP).

Figure 2. Optical sensor systems on training ranges.

2.1. Current instrumentation on training ranges

A usual approach present in many training situations involving human performance, in both military and civilian domains, is to use a camera system and make a full recording of training session. The goal of that approach is to have a visual illustration of exhibited performance, which later on can be reviewed at will by anyone who needs to analyze it. For that purpose a set of videos are played back in their entirety, and if they were properly marked and referenced in time one can acquire fairly good understanding of what happened during the training session in a fairly small training environment (assuming that important human performances were captured by the cameras). By its nature, the review procedure will consist of multiple hours of watching video recordings, which is a burden that cannot be avoided with this type of media.

There are many training ranges for urban warfare that use the very same approach; Figure 2 shows the examples of optical sensor system installed on one such environment. Since the nature of this environment is that it is highly occluded and very difficult to 'cover' with relatively few cameras, the number of cameras and the amount of material recorded for even fairly short training session grows dramatically. That fact alone is a big obstacle in making a take-away material that will not impose on time the unit needs to review it.

2.2. Identified domain needs

As a part of the project a thorough domain analysis has been conducted; we identified several major areas where the new approaches with recording, analyzing and reviewing human performance were the most critical.

Table 1 lists the most important issues identified in this training domain; we specify current approach and its deficiencies, and contrast it with BASE-IT approach and the critical new features it enables.

3. BASE-IT concepts

A major feature that BASE-IT approach provides is the ability to acquire a data set that reflects unit's performance and derive selected quantitative measures of unit's performance, allowing the instructors to augment their predominantly qualitative AAR and provide a more comprehensive critique of unit performances. Specially dedicated smart, automatically controlled sensor network of fixed and pan, tilt, and zoom (PTZ) cameras installed throughout the range, Marine-worn GPS-unit (ID of each Marine) and

Table 1. Review of current approaches and solutions implemented in BASE-IT project.

Issue, domain need, and current approach	BASE-IT approach (new approach)
Recording of human performances: Need: ● Provide a unified spatial representation of performances of all individuals across entire large terrain ● Enable a compact form of data representation and be able to view it even on a laptop platform ● Data to be viewed (explored) from an arbitrary viewpoint Current solution: A set of disjoint 2D video streams (one video stream per camera) that need to be watched and inspected by humans in their entirety.	A set of terrain-registered 3D tracks with behavior-specific information associated with each individual (personal ID, pose and posture). All performance data saved in the data base - a 'perfect memory' that extends beyond human capabilities.
Camera calibration and camera manipulation: Need: ● Reduce the cost of camera operation ● Have a solution viable for an environment with a large number of cameras Current solution: Semi-automated camera systems.	Automated camera calibration and automated camera management.
Identification of each Marine: Need: Identify 'who did what' – the mistakes and successes need to be attributed to individuals who perform them (goal: support effective AAR and skill remediation). Current solution: None – Marines cannot be uniquely identified in the videos.	Enabled. It is possible to know 'who did what'.
Individual and team performance assessment during training: Need: ● Generate alarms - indicate the moments and places where some critical events happened (events may or may have not been seen by the instructors) ● Provide instructors with the information that will help them fine-tune unit's performance on fly Current solution: Only if seen and commented by instructors.	System aims to continuously provide a number of performance metrics; to generate automated alarms for the main instructors (cameras can be omnipresent – instructors cannot).
AAR: Need: More effective, comprehensive AAR that allows identification of individual and team trends. Provide both qualitative and quantitative data. Current solution: Done by instructors.	Done by instructors, and supplemented with the results of a quantitative performance evaluation done by the system.
Individual and team performance assessment after the training: Need: ● Enable automated recognition and quantitative evaluation of human performances (individual and team) ● Provide instructors with the quantitative pointers about the team and individual performances even if the instructors did not see them (e.g. identify all individuals who were repeatedly bunching up) **Currently**: Assessed and reported only if instructors saw particular performance, wrote a note about it or they remembered it.	Instructors provided with the quantitative data related to unit's performance (team and individual performance data).
Exploration of 'what-if' scenarios ('play forward'): Need: Allow hypothesis testing (mission planning). Enable team discussion. Current solution: None.	Enabled (Behavior Synthesis module).
Pre-Mojave Viper Info package given to the units: Need: Enable more effective familiarization of the unit with the terrain and performances the unit is going to experience on the ranges. Current solution: Text information and slides. No opportunity to visually experience a set of courses the unit will do on ranges.	Library of performances recorded on the same ranges. Unit can view and query (search) 3D data at will.

(continued)

Table 1. (continued)

Issue, domain need, and current approach	BASE-IT approach (new approach)
Take-away package for the units: Need: Take-away package that requires minimal time to review. Allow selective viewing – enable smart searches of performance data and bring the viewer quickly to the points of interest. Current solution: Multiple DVDs with video streams showing the unit's performance. No guarantees that important performances were recorded. Hours needed to review it.	Viewable 3D data set that can be queried and selectively visualized from arbitrary viewpoint.
Analysis of historical trends: Need: Enable effective historical trends analysis per unit or per selected performance (provide service-wide benefit). Current solution: Comprehensive analysis is nonexistent. Instructors may keep internal documentation that lists only the major trends during the most recent period.	Historical trends analysis enabled.

inertial measurement units (IMUs) are used to transcode an analog event like a training exercise on the range, into a digital form – a set of time-stamped pose and posture data with corresponding ID stored in a database. As a part of that process each Marine's position, body posture, and even the torso, head and weapon orientation are continuously estimated and recorded. This data set is then analyzed by BASE-IT behavior analysis algorithm that recognizes the events like combat patrol, personnel check point, cordon and search, and evaluates different aspects of individual and team performances (examples: dispersion within the unit, weapon flagging, spatial arrangement of all Marines in a check point situation).

BASE-IT effort is constituted of multiple segments, where each segment addressees the training need expressed by the unit in different stages of their training regiment. One segment addresses units' needs in getting acquainted with the training and exercises they will encounter on the ranges – an interactive three-dimensional (3D) game-based system that visualizes training sessions done by other units, has a much better chance of being engaging and illustrative when compared with traditional text and slide presentations. The same module, enriched with the smart searches of recorded performance data, can be used to review a unit's own performances. This form now represents a unique take-away package that is easy to browse; it helps replace the task that lasted multiple hours if not days with several minutes long searches of performance data. Another segment in the system is focused on the needs of instructors while on the training range; the system features that could be available to them consist of alerts and other useful information being presented to the instructors in a visual or auditory form during the training run, ultimately assisting them in fine-tuning of the exercise and maximizing a potential of the training range and a limited time the unit and its instructors can spend on it.

The data collected on the training range is also the basis for automated behavior analysis at the individual and unit-level, measured with performance metrics against known military techniques and procedures. The works presented in Knerr et al., Hone et al., and LaVoie et al. suggest the ways AAR can be automated.[4–6] In particular, Knerr et al. timestamp user-selected events during the exercise and cross-link to the respective video frames during AAR.[4] A key novelty in BASE-IT system is the availability of information about the individual Marine. This will enable the personalized AAR reports in form of a custom sequence of videos in which the individual can be seen, custom statistics about the proximity to other Marines and civilians, and even a first-person viewpoint flythrough inside a 3D virtual environment augmented with the actual video streams, or watch the recorded activity from any vantage point.

Finally, the data collected during the training runs could be packaged and made available to the units and individuals in an easily accessible format. It could provide much needed statistics about the unit performances, like a frequency of combat patrolling and a number of Marines involved, how many 'cordon and search' operations were executed, as well as a measure of the unit dispersion across the terrain. 'Smart' searches can yield a set of pointers to noteworthy events like 'bunching up', 'weapon flagging' or sniper rifle fire incidents. The events selected this way can then be used as a basis for unit's discussion on improving its tactics and planning of future missions.

While there is great value in analytical nature of BASE-IT effort, the project introduces an entirely new capability: the generation of 'what-if' or 'play-forward' scenarios that are synthesized (created) 'on fly'. Replay of actual events can be paused and an agent-based play-forward engine turned on, displaying a fictional variation of the actual events. The goal of this capability is to demonstrate an alternative, possibly better or worse behavior than what the

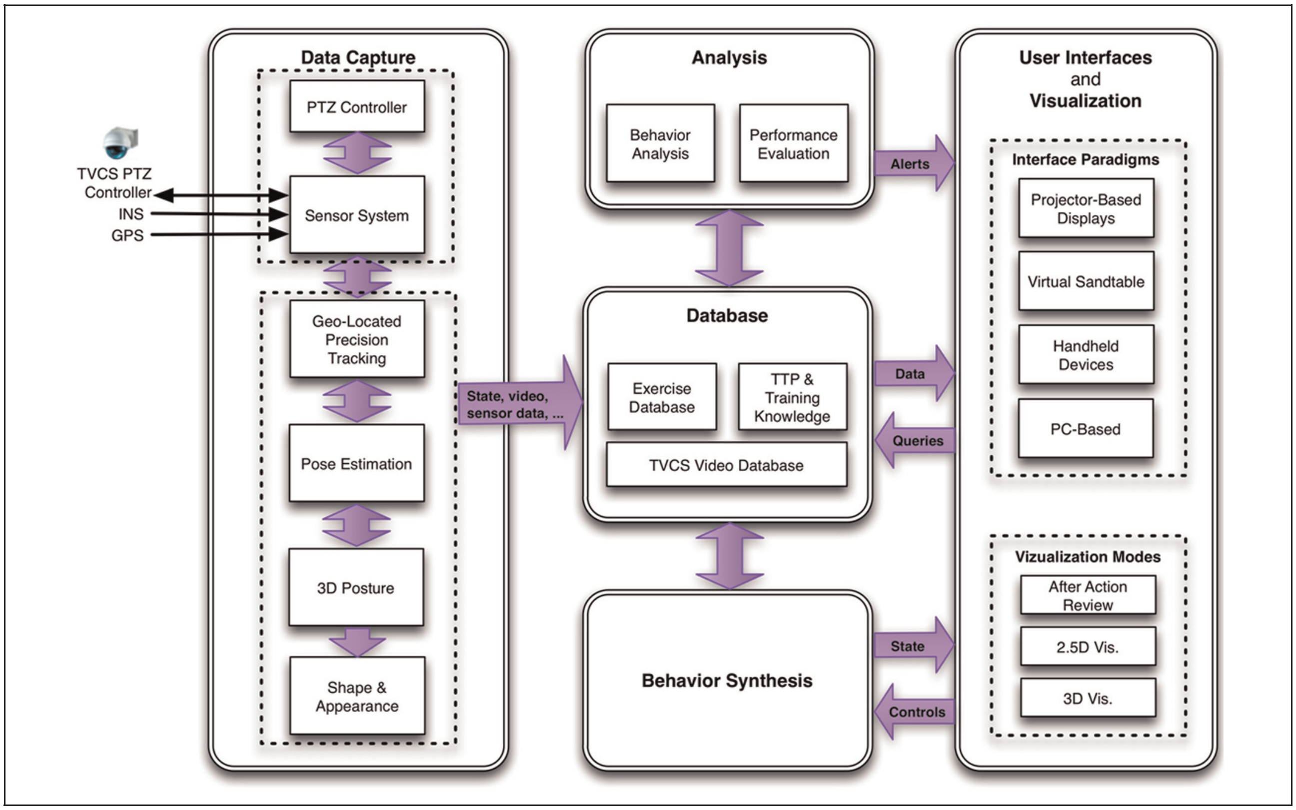

Figure 3. A block diagram of BASE-IT system architecture.

unit decided to do, and use that simulation as an additional resource for unit discussion and critique.

4. System overview

The BASE-IT system architecture is organized around a central database that connects all other components. 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 (Figure 3). The data capture block is responsible for producing time-stamped data on exercise participant position, head and weapon orientation, and body posture. This block includes real-time control of PTZ and static cameras on the range so as to optimally track participants, thus providing the imagery needed for accurate geo-referenced position estimation (by fusing video and GPS estimates) and 2D/3D posture estimation. Output of all capture-module processing is stored in the exercise database. The behavior analysis block reads participant data from the database, and then uses it to generate performance assessment data, which itself is then stored in the database. The behavior synthesis (play-forward) block supports the user with the exploration of alternative courses of action. It can read an initial situation from the database, and then simulate the rest of an exercise under the interactive control of the user. The visualization block reads exercise data and analysis results from the database and renders it on conventional display devices (PC, laptop, handheld mobile) or immersive projective displays that are more appropriate for team collaboration. The handheld mobile device is also designed as a capture device, and enables instructors during at exercise to add time-stamped voice, text, or video annotation to the exercise database.

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 a 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 SQL database reads and writes.

5. Range instrumentation: transparency of sensor system

An important goal for instrumented ranges is to make sure that the instrumentation (e.g. sensors) is inconspicuous to the trainees and doesn't get in the way of training. In addition to the cameras positioned on the poles in 'outside-looking-in' fashion, the sensors used to instrument the trainees are located unobtrusively in the head area (on the helmet) and on the weapon. Each trainee is equipped with a small set of sensors including GPS for outdoor training and Inertial Navigation Sensors (INS). Data collected from these sensors are wirelessly transmitted to a central Device Server, which is part of our data capture system.

5.1. GPS-assisted visual tracking

For outdoor training, in order to cover a large area using minimal number of cameras, we often use PTZ cameras actively following the trainees during a training exercise. Although GPS is effective for outdoor tracking, its accuracy is not sufficient for behavior analysis and performance evaluation. Visual tracking using videos on the other hand can provide accurate localization but are not robust against occlusion, view changes and lighting etc. We decided to combine the strengths of each approach and develop a GPS-assisted visual tracking algorithm.

Our visual tracking algorithm consists of a person detector that uses a set of templates to create person detection likelihoods in a video frame. A set of the person's silhouette and motion templates are produced using a variety of human walking postures. The mean templates from different walking postures at varying scales are saved a-priori. For each video frame, motion blobs are obtained using frame-to-frame registration. These, together with gradient images, are used for correlation-based human template matching to produce the person detection likelihoods. The likelihoods are projected onto a geo-spatial reference plane also used by the GPS data. A local maxima search is used to obtain person localization hypotheses reflecting a high degree of consensus between both the visual detection and GPS localizations for all participants. Tracking over time is then formulated as a correspondence problem between previous track locations and current detection hypotheses, which is solved using bi-partite graph matching. We model GPS error as a Gaussian process and estimate the GPS drift; this improves detection accuracy and provides better track prediction in the case of catastrophic failure of visual tracking.

6. Robust volumetric body reconstruction and tracking

Real-time 3D shape estimation and motion tracking are important for Marine training observation and evaluation. However the major focus of the current computer vision technology is on good reconstruction quality in a well-controlled indoor environment.[7,8] The outdoor Marine training site has additional challenges, which all pose significant challenges to traditional 3D reconstruction and tracking: the camouflaged uniforms, shadows, lighting variation, and complex background motions, e.g. flickering tree leaves.

In order to achieve robust estimation of the shape in the presence of the challenges mentioned above, we represent the 3D scene as a probabilistic volume.[9] For every voxel, we check if it is projected into the silhouette of the shape in all views. In other words, we compute the posterior probability of every voxel being ''occupied'' by the shape, given the silhouette observations from all camera views. In the camera views, the silhouette is also computed probabilistically by maintaining an RGB Gaussian background model at every pixel. As shown in Figure 4(e) and (f), the final probability volume shape provides stable information for tracking.

The Marine posture tracking starts from fitting a standard 3D human model of 3D skeleton and surface triangle mesh to the first frame of the recovered volume.[10] For the following frames, the model is warped within the degrees of freedom of the skeleton joints and mesh, so that the projected shape in all cameras best match the observed silhouettes, and best explain the observed image motion from the previous frame to the current.[11] The current performance reaches 8–9 seconds per frame using eight cameras, on an 8-core PC machine with GPU acceleration on the probabilistic volume and image flow computation.

7. System automation: behavior analysis

Automated behavior analysis and event detection is a key component for an intelligent sensor system on instrumented training ranges. Detection and recognition of arbitrary human behavior is an extremely challenging problem because of the vast number of possibilities in an unconstrained environment. However, for training applications, behavior analysis is greatly simplified due to the controlled environment, the staged stimuli and the set of expected behaviors already known to the system. Taking advantage of the training domain, our behavior analysis framework uses a finite state machine (FSM) model where trainees' behavior are the *states* and the transitions of states are caused by stimuli that we refer to as *trigger events*. The goal of behavior analysis is to estimate the states that the trainees are *actually* in and the states that the trainees *should* be in at any given time. The former are used for exercise and scenario control and the later are used for performance evaluation. To robustly detect each state, we build classifiers for not only for each state, but also for each trigger event. At a given time, based on the state estimation, a set of related classifiers are activated for detecting trigger events and states that can be transitioned to and from the current states.

7.1. Behavior analysis framework

We model a training exercise as a finite state machine (FSM), a quintuple $(\Sigma, S, s_0, \delta, F)$, where:

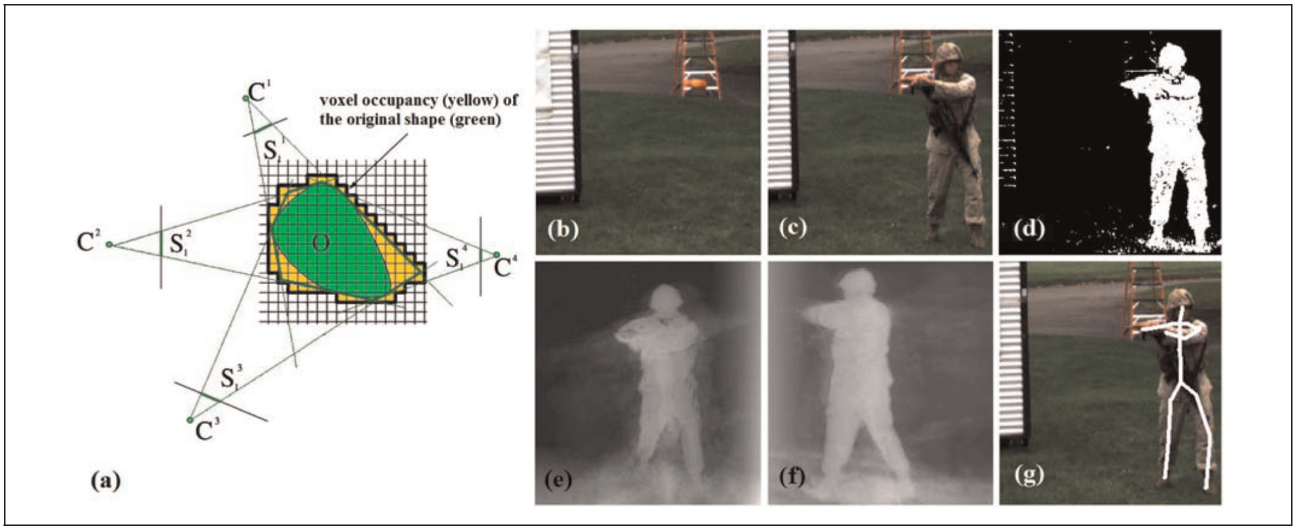

Figure 4. (a) A 2D illustration of the occupancy voxel computation with multiple cameras; (b) an empty background frame; (c) the same camera view with the Marine; (d) silhouette probability from (b) and (c); (e) 3D probability occupancy volume from 8 views; (f) another angle of the volume; (g) fitted and tracked skeleton model.

- Σ is the input alphabet (finite and non-empty) representing the set of stimuli or trigger events
- S is a finite, non-empty set of states representing the set of possible behaviors (i.e. states of the participants)
- s_0 is an initial state, an element of S.
- δ is the state-transition function that returns a set of transition probabilities: $\delta : S \times \Sigma \rightarrow P(S)$. This is the reaction to a stimulus (both correct and incorrect reactions).
- F is the set of final states, a subset of S.

For a training system, states S can only be perceived through sensor observations, O. Then, behavior analysis is to estimate states $S = \{s_0, s_1, \ldots, s_n\}$ given sensor observation $O = \{o_0, o_1, \ldots, o_n\}$. In our system, the sensor inputs include positions of all participants, their head, body and gun poses and shot/hit data.

Initially, it would appear that behavior analysis can be solved using the standard Hidden Markov Model (HMM). However, given the special circumstance of the training application, behavior analysis for training can be modeled using an Augmented HMM (AHMM).[12] We describe the details of our augmented HMMs in Cheng et al.[13]

Based on the AHMM, to estimate the state that trainees are in, one needs to define and compute the conditional probability $P(O_i|S_i)$ for all possible states and the prior probability $P(S_i|S_{i-1},T_i)$ for all possible transitions. For simple states, such as walking, running, shooting, the conditional probability $P(O_i|S_i)$ can be manually defined. However, complex activities, such as group formations and group interaction, are difficult to detect using a rule based approach.

We propose a novel feature, Histogram of Oriented Occurrences (HO2),[14] to model and recognize complex group activities. HO2 captures the interactions of all entities of interests in terms of configurations over space and time and can be used with standard classifiers, such as SVM (Support Vector Machine) for complex activity detection and classification. The output the these classifiers will be normalized as the conditional probability $P(O_i|S_i)$. To build one specific event detector, annotated samples of the event are used as the training set. The HO2 features are computed for all entities. Then, an SVM classier is built using this training set.

7.2. 2D posture recognition

Currently, there is no automated tool to analyze the performance of individual Marines during their training. Instructors visually observe the unit's behavior and note noteworthy events on paper notepads, which they will discuss during the AAR debrief. Many behaviors such as weapon flagging or a lack in appropriate scanning the environment will go unnoticed merely due to the many tasks an instructor has to take care of. BASE-IT's automated tools perform continuous, best-effort surveillance at the individual Marine level, providing an entirely new capability and data. Human posture recognition has in the past been an offline process, constrained to controlled environments, or had to rely on several cameras on the same individual. BASE-IT demonstrates a scalable approach that may yield useful results in real-time.

After individuals have been identified and their exact position has been located in the image plane, the 2D posture recognition module analyzes the region of interest (ROI) in the image where the person has been located. BASE-IT employs real-time capable algorithms for posture estimation, based on full-body appearance and on the configuration and orientation of body parts and the weapon. After hand-annotating a Marine-specific training set using a slightly modified version of the LabelMe tool,[15] various classifiers were trained on Marine appearances in general and on specific postures, in particular. The best performance was achieved with a combination of a Viola–Jones style detector and a parts-based multi-class classifier similar to Torralba's method.[16,17] Torralba's method selects small rectangular areas predominantly on the contour of the Marine, convolves those with several filters, and employs a boosting scheme to simultaneously create classifiers for several postures. During testing, the same filters are applied and strong matches 'vote' for a posture-location combination in the image. The votes from all filters are combined and thresholded to determine the occurrence of a particular posture at a particular position in the image. Spatio-temporal non-maxima suppression with an adaptive mean-shift filter combines several nearby detections and picks the most likely posture. This post-processing step is necessary as the detector is scanned across the ROI to account for ROI inaccuracies, slight offsets and scale differences, resulting in multiple detections nearby.[18]

The head and weapon are found with two additional detectors that are applied to the ROI. The eventual goal is to extract head and weapon position and their orientation with another multi-class classifier. The output of all three detectors is syntactically verified in yet another post-processing step to avoid false results such as torso-facing-left and head-facing-right. Figure 5 shows the posture detector and correct identifications of standing Marines (red) and kneeling Marines (yellow), and the respective direction of the torso (toward or away from the camera, left or right).

7.3. Individual performance metrics

Semantic analysis infers what the Marines are doing and checks whether it obeys Marine doctrine, using their calculated positions and orientations. After discussion with domain experts [6 USMC, 7 USMC] and preliminary experiments, we identified fourteen performance metrics for a small group (4–13 Marines) every second: dispersion, collinearity, number of clusters, non-Marine interaction, danger, awareness of danger, mobility, speed, 'flagging' (pointing weapons at one another), weapons coverage, being too close to a window or door, being too far from a window or door, surrounding of a target, and centrality of the leader.[19] We also automatically identify boolean 'issues' of interest and potential comment by an instructor, both on individual Marines (e.g. a Marine failing to sweep ('pie') a nearby window or door with his weapon) and on the entire group of Marines (e.g. forming a single column causing them to be an easier sniper target). The metrics are on a scale of 0 (good) to 1 (bad), and require nonlinear mappings of the form $g(x) = 0.5 + (1/\pi)\arctan((x-\mu)/\sigma)$ on measured quantities, like average distance between Marines, to ensure this. The issues require thresholds on measured quantities.

The most difficult metrics to calculate concern a danger, which is approximated by finding cumulative danger from a finite set of dangerous points including both good sniper positions (windows, doors, and occluding corners of buildings) and centers of representative areas of terrain, and weighting them by the inverse of their distance to model sniper accuracy. Danger also decreases when the Marines are near good cover (a 'blur' in the spatial dimension), decreases quickly as the Marines visually scan a

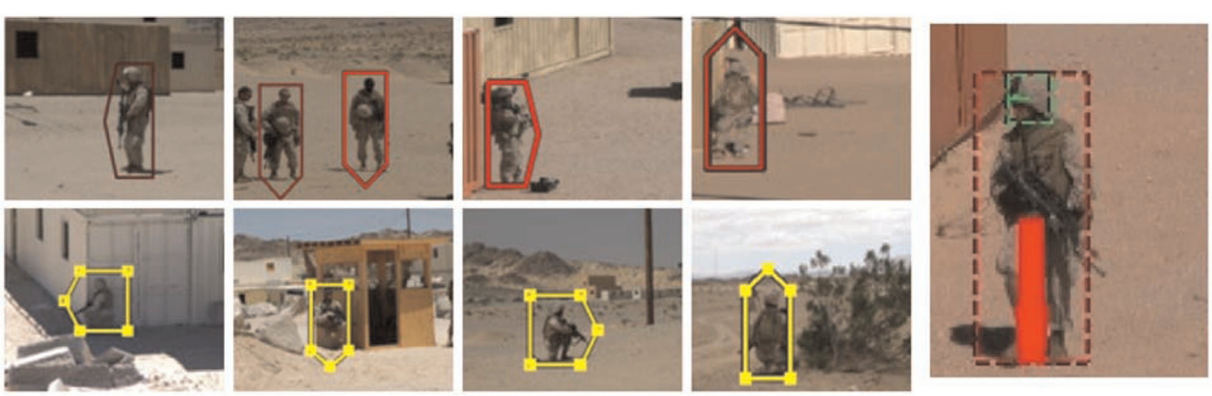

Figure 5. Marines standing and kneeling, their torso and (in the right image) their head orientation indicated with arrows or pointed bounding boxes.

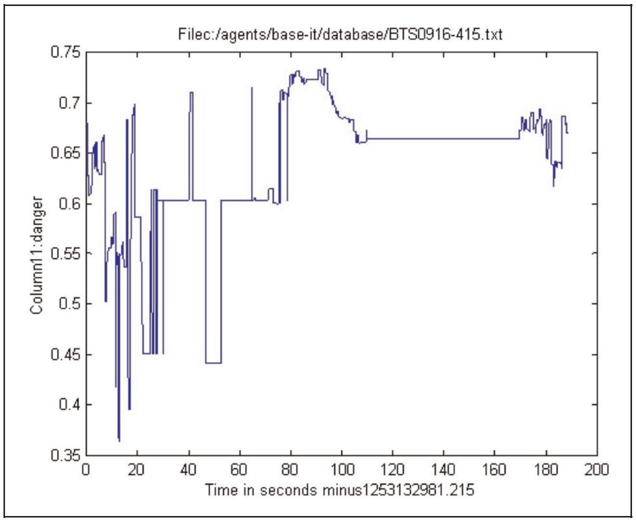

Figure 6. Danger calculated during an exercise.

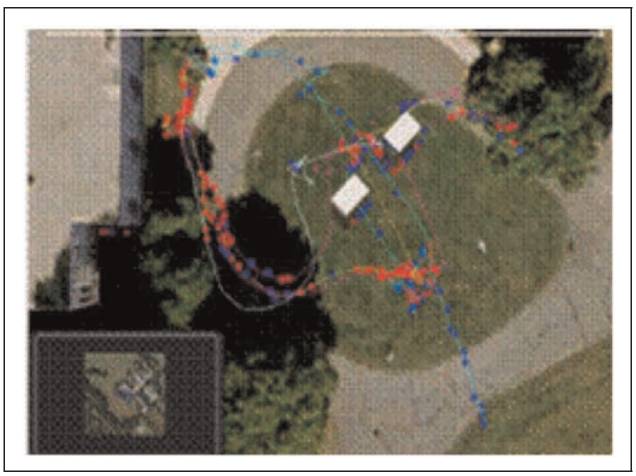

Figure 7. Example visualization with colored dots representing issues.

safe area, and then increases slowly when the Marines are not scanning an area (a 'blur' in the time dimension). Figure 6 plots the calculated danger for the group of Marines over an example exercise, and Figure 7 shows our visualization viewed from above with dots on the tracks indicating issues of being too close to one another (blue), being too close to a window or door (green), and 'flagging' (red).

The performance metrics are aggregated over each exercise to measure overall performance, in the form of means, minima, maxima, and counts of low, medium, or high values in their sub-range. These provide a basis for the after-action review.[20] Particularly anomalous aggregated values are noted as additional issues for instructors to discuss. Aggregation is also done over all squads on the same exercise (to compare exercises), over all exercises for a squad (to compare squads), and over each possible type of 'behavior'. Behaviors are defined as 30 important categories of Marine activity such as receiving instructions, patrolling, running a checkpoint, surrounding a building, and searching a building. We infer behaviors at each second initially from a support-vector model based on metrics plus additional information such as posture. We refine the inferences with hidden-Markov inference, using as a Markov model the known sequences within Marine exercises based on doctrine.

8. Visualization: live, play-back, and play-forward

8.1. 3D visualization and play forward

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.

A 3D rendering is accomplished using the open-source Delta3D simulation/game engine;[21] Figure 8(a) shows one such scene. 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 camera motion models including one inspired by real-time strategy games. Additionally, Delta3D should make providing DIS or HLA networking relatively straightforward 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. We developed a novel technique to control gaze and aim via a constant-time algorithm for generating blend weights, which are then fed to a conventional blending capability in Delta3D's character animation subsystem.

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 position or route of the simulated Marines), and generate 'what-if' scenarios on the fly, i.e. to support a 'free play' mode.

This functionality is provided by the BASE-IT behavior synthesis module. 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

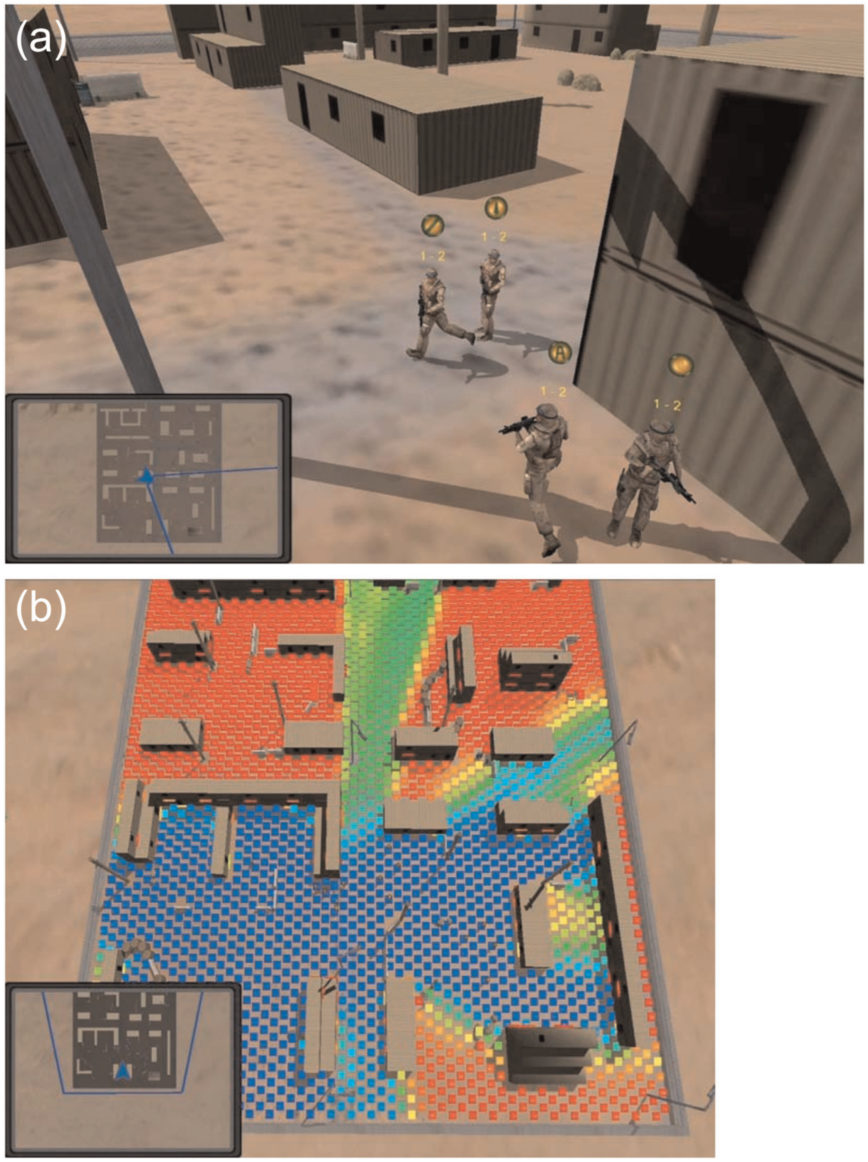

Figure 8. (a) 3D visualization window and (b) possible threat location heatmap.

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 to be at the present time. We explicitly model the subjective likelihood of a threat being at a given position by a statistical model (see heatmap visualization in Figure 8(b)). The model drives where the simulated Marines look at any given time, and the threat likelihoods in turn are updated based on where they look, i.e. unless a target is actually spotted, the threat probability in the viewed region is reduced.

Pathfinding has received a lot of attention in recent years, particularly in the video game industry.[22] We take an approach that is novel in that, while a conventional A* path is planned for a fireteam as a whole, individual greedy searches are then carried out for each Marine. While the

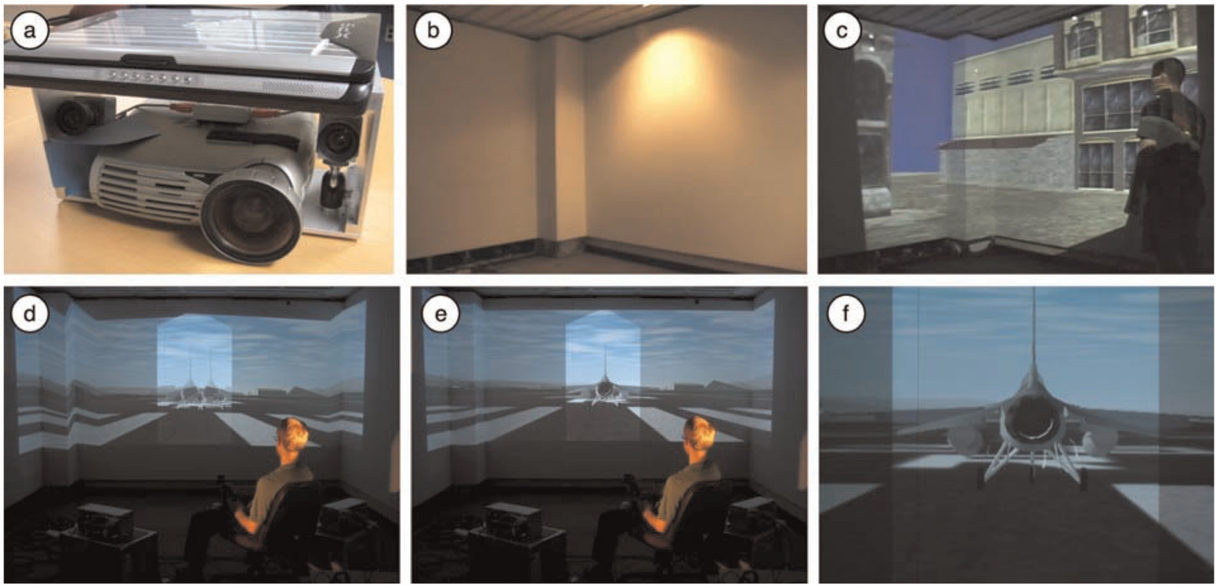

Figure 9. Multi-projector display systems with continuous automatic calibration. (a) A projector, two cameras, and a laptop computer combined into an IPU. (b) The corner of a room that includes a support pillar. (c) An IPU used to render imagery that is corrected for the unusually shaped display surface. (d)–(f) Two IPUs used with a flight simulator application. There is no projector blending so that the overlap region is visible (brighter region in center). (d) Both IPUs were moved, resulting (temporarily) in distorted imagery. (e) After a few seconds the imagery has been automatically corrected. (f) A close-up of the undistorted overlapping projector regions showing.

fireteam path minimizes distance and maximizes cover, the individual Marine paths optimize additional factors such as fireteam dispersion and distance from fireteam leader.

The 3D visualization and play forward capability combine to make a system that we hope will be highly useful for a Marine trainee. Not only can he readily see the mistakes his unit has made in a previous exercise, but he can also experiment with various corrections using ‘‘free play’’ mode.

9. Display solutions

We have developed 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 to provide displays that are simultaneously large and high-resolution compared to traditional 2D displays. For pre-operation instruction and AAR we have also developed a projector-based virtual-physical sand table display that supports computer generated appearances and markings, while preserving the true 3D of physical tabletop models. Finally, we have identified and will be developing handheld interfaces for instructors during field exercises to support geo-located and pose-aware voice, image, and video annotations.

Traditionally training plans and results are displayed using flat panel displays or single projectors that do not provide the simultaneous large size and high resolution needed to accommodate a large group of viewers. 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.[23,24] We have developed new automatic methods for camera-based calibration and associated rendering to support the creation of ad hoc multi-projector display systems that support large, high resolution imagery. To make such ad hoc displays more operationally deployable and robust, we have developed a the concept of ‘intelligent projector units’ (IPUs) and a distributed framework for continuous calibration that runs concurrently with the application and adapts to small perturbations in projector pose,[25] thus reducing the need to stop and re-calibrate. An example of an IPU and some calibration results are shown in Figure 9.

Physical ‘sand tables’ have been used in the military for many years.[26] Their advantages include simplicity and a natural means for conveying tactical plans and reviews related to the positions and movements of assets over time.

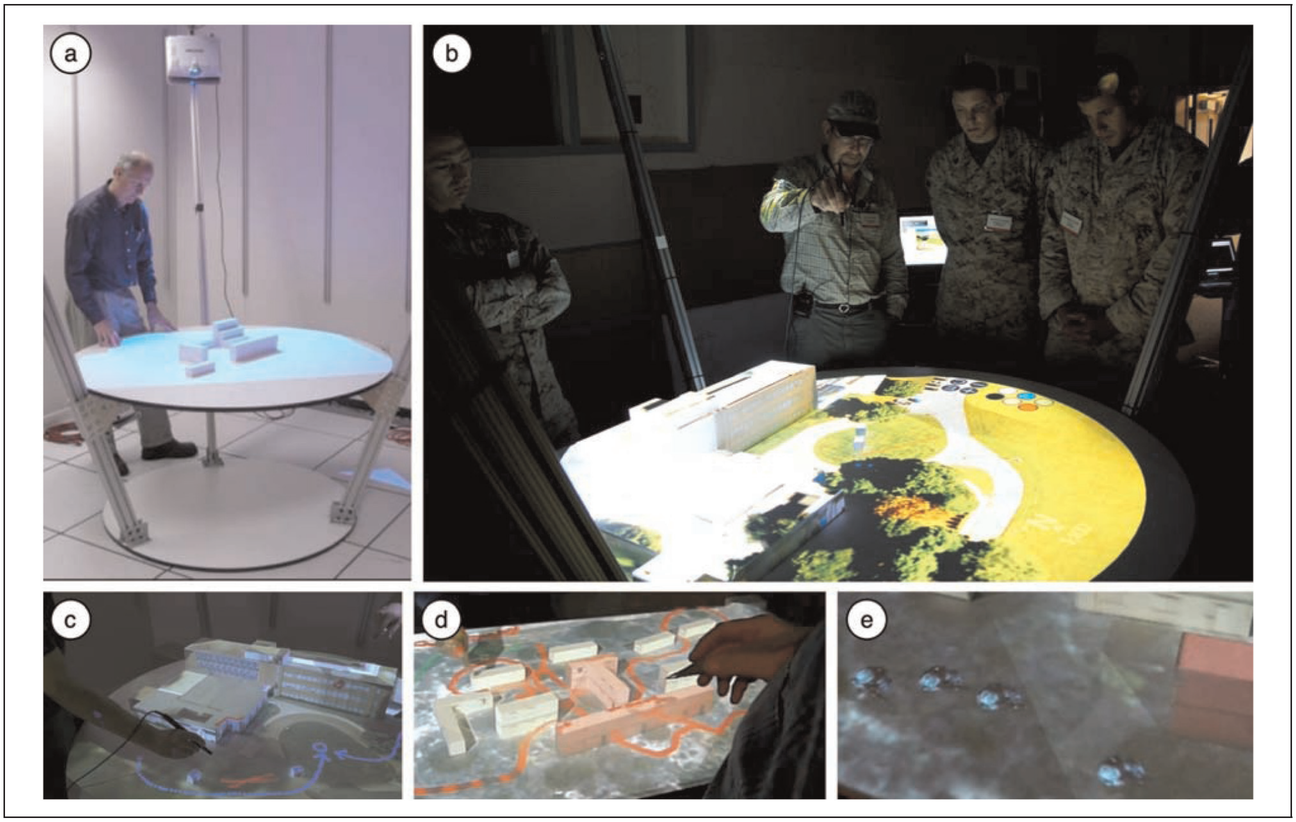

Figure 10. Our Projector-Based Physical Sand Table prototype. (a) The overall physical structure of the system with projectors and physical blocks. (b) The system was demonstrated to some Marines during an annual project demonstration meeting (September of 2009). Their motion – captured during a previous exercise – is being discussed by retired USMC Maj. Matthew Denney (center, holding the stylus). (c) A close up view of an annotation in progress. (d) A demonstration with a model of the Kilo 2 MOUT site at Camp Pendleton. (e) Some simulated Marine movements depicted by top-down Marine icons.

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. We have developed a modern version that combines the simplicity and intuitiveness of physical sand tables with dynamic digital training content. To accomplish this we employ digital projectors, Shader Lamps techniques,[27,28] and specialized projector blending techniques to render textures, tracks, moving icons, and training-related information onto a physical model of the site. We have developed a transportable prototype system we call a Projector-Based Physical Sand Table.[25] The prototype, shown in Figure 10, also includes a magnetic tracking system to estimate the 6D pose of a moving stylus, and associated techniques that support free-form annotation of the table in the style of a 'telestrator'. We have built two copies, one at UNC and another that is being used at NPS.

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. Traditionally they have used clipboards and radios to manage the training events and record anomalies. While they have recently begun to use conventional handheld devices to communicate with each other and take picture, we envision networked handheld devices that are aware of their pose (position and orientation), and can record pose-tagged voice, image, and video notes. While we believe that such pose-aware notes will be valuable, conventional handheld user interfaces are arguably inferior to the traditional clipboard in terms of complexity and associated lower speed and robustness. In an attempt to overcome these issues we are working on new 'point and shoot' methods that support touch tagging voice, images, and video to individuals in the imagery. The user can pan and zoom to change the view, while seeing actual motion tracks (pulled from a database) and spatially-located icons indicating images or audio recordings that the user can touch to see/play.

10. Future work

The BASE-IT approach focuses on collecting and analyzing one type of data: positional data. Such data set provides a solid level of understanding about different aspects about unit achievements on the training range, however it is important to note that this does not offer a complete picture about the unit performance in given training session. One would be closer to that goal only if all other types of data pertinent to unit performance are collected and analyzed: audio recordings consisted of radio communication and shouts used by all members of the unit, their use of hand signals, the information about weapons deployment and the effects of their deployment, use of smoke signals, and air assets. The system architecture offered in the BASE-IT system allows for a gradual expansion and augmentation by adding new types of data and new types of data analysis; in the future we see the research community extending and leveraging this work with the ultimate goal of acquiring a comprehensive understanding about unit performance regardless of the environment in which the training is organized.

11. Conclusion

Paramount to military training is skill perfection in conditions that resemble actual conditions the trainees might experience in the theatre. The physical training ranges that simulate those conditions remain to play a significant role in units' preparations and skill integration leading most directly to the actual deployment. In order to derive maximum benefit from such critical resources, a new approach has been designed and prototyped in a system called BASE-IT with the complementary assistive technologies. A primary goal was to provide all participants in training sessions, with very much needed products currently not available with traditional infrastructure and traditional approaches. Introduction of a multi-sensor system capable of collecting and analyzing the performance data represents a significant departure and improvement in recording and understanding the unit's performance during training sessions. The instructors are provided with quantitative pointers relevant to unit's performance, and the units themselves have tools that allow them quick examination and review of their training run.

BASE-IT rests on the foundation of a thorough, long-term task analysis study. It brings recent technological advances in automatic data and video analysis, and artificial intelligence directly into the hands of the trainers and trainees, providing customized support aimed at enhancing training effectiveness and resource utilization. We believe this approach will allow the training force to make a significant step towards the warfighters' better preparedness for today's complex MOUT situations.

Acknowledgements

The authors would like to acknowledge the following agencies and individuals for their continuing support: the Program Manager for Training Systems (PM TRASYS) and Training and Education Command (TECOM), leadership and instructors of Tactical Training Exercise Control Group (TTECG), Twentynine Palms and Kilo Two, Camp Pendleton, as well as many USMC units and individuals who supported different elements of our effort.

Funding

This work was supported by the Office of Naval Research (ONR).

References

1. US Marine Corps. Infantry training and readiness manual. Directive NAVMC DIR 3500.87, September 2005.
2. Morrison JE and Meliza LL. Foundations of the after-action review process. Special Report 42, US Army Research Institute for the Behavioral and Social Sciences, July 1999.
3. Sadagic A, Welch G, Basu C, et al. New generation of instrumented ranges: enabling automated performance analysis. In: *Proceedings of I/ITSEC 2009.*
4. Knerr BW, Lampton DR, Martin GA, et al. Developing an after-action review system for virtual dismounted infantry simulations. In: *Proceedings of I/ITSEC 2002*.
5. Hone G, Swift D, Whitworth I, et al. The case for coarse-grained after-action review in computer aided exercises. In: *Proceedings of the 13th International Command and Control Research and Technology Symposium*, Bellevue, WA, June 2008.
6. LaVoie N, Foltz P, Rosenstein M, et al. Automated support for AARs: exploiting communication to assess team performance. In: *Proceedings of I/ITSEC 2008*.
7. Franco J-S and Boyer E. Exact polyhedral visual hulls. In: *British machine vision conference 2003*, pp.329–338.
8. de Aguiar E, Theobalt C, Stoll C, et al. Marker-less deformable mesh tracking for human shape and motion capture. In: *CVPR 2007*.
9. Franco J-S and Boyer E. Fusion of multi-view silhouette cues using a space occupancy grid. In: *ICCV, II*, pp.1747–1753.
10. Ballan L and Cortelazzo GM. Marker-less motion capture of skinned models in a four camera set-up using optical flow and silhouettes. In: *Proceedings of 3DPV'08 – the fourth international symposium on 3D data processing, visualization and transmission*, 2008.
11. Zach C, Pock T and Bischof H. A duality based approach for realtime TV-L1 optical flow. In: *Pattern Recognition (Proceedings of DAGM)*, 2007.
12. Rimey RD and Brown CM. Sequences, structure, and active vision. In: *Computer vision and pattern recognition, 1991. Proceedings CVPR'91*, June, pp.706–709.
13. Cheng H, Kumar R, Basu C, et al. An instrumentation and computational framework of automated behavior analysis and performance evaluation for infantry training. In: *Proceedings of I/ITSEC 2009*.
14. Cheng H, Yang C, Han F, et al. HO2: A new feature for multi-agent event detection and recognition. In: *Computer vision pattern recognition workshop*, 2008.

15. Russell BC, Torralba A, Murphy KP, et al. LabelMe: a database and web-based tool for image annotation. *Int J Comput Vision* 2008; 77: 157–173.
16. Viola P and Jones M. Robust real-time object detection. In: *Second international workshop on statistical and computer theories of vision*, Vancouver, July 2001.
17. Torralba A, Murphy KP and Freeman WT. Sharing visual features for multiclass and multiview object detection. *IEEE Trans Pattern Anal Mach Intell* 2007; 29: 854–869.
18. Wachs JP, Goshorn D and Kölsch M. Recognizing human postures and poses in monocular still images. In: *International conference on image processing, computer vision, and pattern recognition (IPCV)*, 2009.
19. Rowe NC, Houde JP, Kolsch MN, et al. Automated assessment of physical-motion tasks for military integrative training. In: *Second international conference on computer supported education*, Valencia, Spain, April 2010.
20. Hixson JA. Battle command AAR methodology: a paradigm for effective training. In: *Proceedings of the winter simulation conference*, Year?, pp.1274–1279.
21. Darken CJ, Anderegg BG and McDowell PL. Game AI in Delta3D. In: *Proceedings of computational intelligence in games (CIG)*, 2007.
22. Guerrero M and Darken C. Blended inverse kinematics: Delta3D system utilization. In: *Proceedings of I/ITSEC 2008*.
23. Raskar R, Welch G and Fuchs H. Seamless projection overlaps using image warping and intensity blending. In: *Fourth international conference on virtual systems and multimedia*, Gifu, Japan, November 1998.
24. Surati RJ. *Scalable self-calibrating display technology for seamless large-scale displays*. PhD Thesis, Massachusetts Institute of Technology, 1999.
25. Johnson T, Towles H, State A, et al. A projector-based physical sand table for tactical planning and review. Technical Report TR09–017, Department of Computer Science, The University of North Carolina at Chapel Hill, USA, December 2009.
26. Bond PS and Crouch EH. *Tactics: the practical art of leading troops in war.* The American Army and Navy Journal, Inc., 1922.
27. Raskar R, Welch G and Chen W-C. Table-top spatially-augmented reality: bringing physical models to life with projected imagery. In: *Second international workshop on augmented reality (IWAR'99)*, 20–21 October 1999.
28. Raskar R, Welch G, Low K-L, et al. Shader lamps: animating real objects with image-based illumination. In: *Eurographics workshop on rendering*, June 2001.

Author biographies

Dr. Amela Sadagic is a Research Associate Professor at the Naval Postgraduate School, Modeling Virtual Environments and Simulations (MOVES) Institute, Monterey, CA. In the past she was a Director of Programs at Advanced Network and Services Inc. where she coordinated National Tele-Immersion Initiative. Her research interests include Virtual Environments, training systems, human factors, learning and training effectiveness, and 3D telepresence. Dr Sadagic received her PhD degree in Computer Science from the University College London, UK, in 1999.

Dr. Mathias Kölsch is Associate Professor of Computer Science at the Naval Postgraduate School in Monterey, California. He is also affiliated with the MOVES Institute, the Space Systems Academic Group, and the Remote Sensing Center at NPS. His research interests include computer vision, hand gesture recognition, augmented reality, and embedded computing. Mathias received his PhD in Computer Science from the University of California Santa Barbara in 2004.

Dr. Greg Welch is a Research Professor in the Institute for Simulation & Training and the Department of Electrical Engineering & Computer Science at The University of Central Florida, and The University of North Carolina at Chapel Hill. Previously he has worked at NASA's Jet Propulsion Laboratory and Northrop-Grumman's Defense Systems Division. His research interests include human tracking systems, 3D telepresence, projector-based graphics, and medical applications of computers. Dr. Welch received his PhD in Computer Science from UNC-Chapel Hill in 1997.

Dr. Chumki Basu is a Program Manager and Research Scientist at IBM India Research Labs, Bangalore, India. Previously, she was at Sarnoff Corporation, Princeton, NJ, and prior to that, at Bell Communications Research/Telcordia Technologies, Morristown, NJ and Piscataway, NJ. Dr. Basu received her PhD from Rutgers University in 2002.

Dr. Christian Darken is an Associate Professor of Computer Science at the Naval Postgraduate School, and faculty affiliate of the MOVES Institute. Previously he was Project Manager and Senior Member of Technical Staff at Siemens Corporate Research in Princeton, New Jersey. He was also the AI programmer on the team that built Meridian 59, the first 3D massively-multiplayer game. His research focus is on human behavior simulations. Dr. Darken received his PhD in Electrical Engineering from Yale University in 1993.

Dr. Juan Wachs is an Assistant Professor in the School of Industrial Engineering at Purdue University. He is the director of the Intelligent Systems and Assistive Technologies Lab (ISAT) and he is affiliated with the Regenstrief Center for Healthcare Engineering. In the past Dr. Wachs held an NRC Research Postdoc position at the Naval Postgraduate School, MOVES Institute, Monterey, CA. Dr. Wachs received his PhD in Industrial

Engineering and Management from Ben Gurion University of the Negev, Israel, in 2007.

Dr. Henry Fuchs is the Federico Gil Distinguished Professor of Computer Science and Adjunct Professor of Biomedical Engineering at UNC Chapel Hill. In the past he was a member of the faculty of the University of Texas at Dallas. He received his PhD from the University of Utah in 1975.

Herman Towles, formerly a research engineer in the Department of Computer Science at UNC-Chapel Hill, is currently a senior R&D systems engineer with Applied Research Associates in Raleigh, NC. He received his MS in Electrical Engineering from University of Memphis in 1972.

Dr. Neil C. Rowe is a Professor of Computer Science at the U.S. Naval Postgraduate School where he has been since 1983. Dr. Rowe received his PhD in Computer Science from Stanford University in 1983.

Dr. Jan-Michael Frahm is an Assistant Professor at the University of North Carolina at Chapel Hill, director of Computer Vision at the Renaissance Computing Institute at UNC. Dr. Frahm received his PhD in Engineering at Christian-Albrechts University of Kiel, Germany, in 2005.

Dr. Li Guan is a researcher at GE Research. He earned his PhD from the University of North Carolina at Chapel Hill in 2010.

Dr. Rakesh "Teddy" Kumar is a Sr. Technical Director of the Vision and Robotics Lab at SRI International Sarnoff. At SRI Sarnoff, he performs and direct research in computer vision, robotics, 3D modeling, video surveillance and simulation and training. Dr. Kumar received his PhD in Computer Vision from University of Massachusetts at Amherst in 1992.

Dr. Hui Cheng is the Sr. Technical Manager of Cognitive and Adaptive Vision Systems Group at SRI International. His research focuses in the area of image/video understanding, data mining, machine learning, pattern analysis, cognitive systems and informatics. He received his PhD in Electrical Engineering from Purdue University in 1999.