AUTOMATED ASSESSMENT OF PHYSICAL-MOTION TASKS FOR MILITARY INTEGRATIVE TRAINING

Neil C. Rowe, Jeff P. Houde, Mathias N. Kolsch, Christian J. Darken, Eric R. Heine, Amela Sadagic

MOVES Institute, U.S. Naval Postgraduate School, 1411 Cunningham Road, Monterey, California, USA

ncrowe@nps.edu, jphoude@nps.edu, mnkolsch@nps.edu, cjdarken@nps.edu, erheine@nps.edu, asadagic@nps.edu

Chumki Basu, Feng Han

Sarnoff Corporation, Box 5300, Princeton, New Jersey, USA

cbasu@sarnoff.com, fhan@sarnoff.com

Abstract: We describe the performance assessment component of the BASE-IT system, a real-time monitoring system of performance of U.S. Marines during training exercises for urban warfare. This automated component measures how well Marines are following procedures and staying safe, by tracking where they are and where they are looking. Such monitoring of physical motion is a relatively new application of computer technology with implications for instruction in physical education, choreography, and police work.

1 INTRODUCTION

Most computer technology supporting education has implemented the electronic equivalent of paper. However, there are important skills that students need to learn that involve different activities like physical motion. Good examples occur in physical education, choreography, industrial training, and military training. Technology now enables us to automatically assess such skills by tracking human motion with wireless communications, computer vision, and sensor analysis. These permit us to measure where people are, how their limbs and torsos are configured, and what gestures they are making. These open new opportunities for automated assistance by computers.

We describe one example, ongoing work for our BASE-IT Project in monitoring U.S. Marine integrative training for urban warfare. We are building a system to noninvasively track the Marines, then analyze what they are doing in real time. While some of this system is specific to Marine needs, many parts of it could be applied to other kinds of education and training.

2 MONITORING PHYSICAL MOTION

In the training of physical motion, video of students is helpful but has drawbacks: Important events can happen too fast to see adequately, they can be occluded by other people or objects, they can be rare within much irrelevant data, and video alone doesn't highlight problems and mistakes. Better results can be obtained with automated video analysis, and this is now being used to aid instruction for such motions as golf and tennis swings (Stepan and Zara, 2002). One technology being explored involves "motion capture" using wearable devices with accelerometers that can measure joint motions precisely (Chen and Hung, 2009; Knight et al, 2007). Also used are special "studio" training environments with multiple cameras. These technologies are starting to be used for choreography (Nakatsu, Tadenuma, and Maekawa, 2001) and other forms of theater, and also in military and police training where motion in crises is important. Putting students in studios for training is not always possible, as in much industrial training. Also, wearable devices are obtrusive as they require special equipment; students are aware of the devices and this affects their behavior. Devices may also be unnecessary for many training tasks for which it suffices to monitor whole-body motions by multi-camera monitoring and data fusion.

3 THE MARINE TRAINING TASK

U.S. Marine soldiers receive extensive training on a wide range of skills. The dangerous nature of their occupation means that improper execution of skills

can be a matter of life or death, so training is important. Urban warfare is particularly difficult because many different skills must be exercised in nonstereotypical ways. Marines have assigned urban-warfare missions such as searching people or vehicles while staying alert to potential dangers from snipers and explosive devices. They must also manage contacts with local civilians who may or may not have friendly intentions.

Urban warfare skills are taught at many times during Marine training programs, but are particularly focused on during the later stages before deployment overseas. This training involves mockups of a town environment in which they must patrol, conduct searches, run checkpoints, respond to unexpected events, and deal with "roleplayers", actors representing local inhabitants. Our BASE-IT Project focuses on these exercises.

Assessment is an important part of training. Instructors watch the Marines during the exercises and provide feedback mostly afterwards during "after-action reviews", a method also used by the U.S. Army (Hixson, 1995). Reviews cover both short-term problems (like weapons safety) and longer-term problems (like the proper sequence for searching a building). Instructors have Training and Readiness Manuals that contain checklists and expected-event sequences that they use to assess performance. This assessment is mostly qualitative, e.g. "Appropriate techniques of movement when crossing danger areas."

Marines carry a good deal of equipment and cannot carry more because of the active nature of their jobs. However, some carry GPS units to report their locations.

4   THE BASE-IT SYSTEM

We describe the performance assessment component of the BASE-IT system covered in broader detail in (Sadagic et al, 2009). It takes inputs from a database of real-time quantitative measurements performed on the trainees during assessment. This data comes from video cameras monitoring the training area from a variety of positions and angles, and from GPS units on the Marines and roleplayers which provide redundant data improving accuracy (Cheng et al, 2009). Camera orientations are automatically controlled by BASE-IT to focus on areas of activity. GPS position data is fused with positions obtained from comparing the camera image to a background model to improve location accuracy.

Computer vision-based analysis of the Marines reveals details about their postures and orientations (Figure 1). Three full-body stances are distinguished (standing, kneeling, and lying down), four torso orientations (towards the camera, away from the camera, left, and right), and four head orientations. This analysis is done on each video frame without a background model. The appearances of small image patches are compared to learned examples of different postures (Wachs, Goshorn, and Kolsch, 2009). Results are improved through temporal post-processing with a hidden Markov model. Information is then correlated with known camera positions and orientations to get real-world orientations.
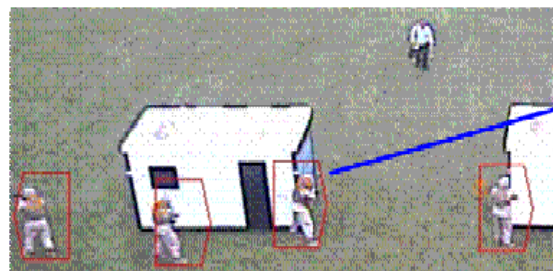

Figure 1: Example of inference of orientations of Marines.

BASE-IT output is visualized in three complementary ways: Moving icons on a "sand table" of three-dimensional white blocks with images projected onto them, video cutouts embedded in a three-dimensional environment, and a "free play game" wherein Marines can be shown from any angle doing what they should have done as well as what they actually did. Performance assessment information can enhance all three of these output options. For instance, we display a timeline with the third option where colored dots code possible mistakes of the Marines (Figure 2).

4.1   Performance Assessment in BASE-IT

Performance assessment attempts to capture things that good human instructors would note during training, but can be missed due to the occlusions by walls, large distances, and the limited number of instructors. We compute both "metrics" and "issues" during training. Metrics are numeric measures of things important to instructors, mostly on a scale of 0 (good) to 1 (bad). Issues are problems that may require comment by instructors and could be potential mistakes, but may have valid excuses in context. Details of the formulas and algorithms we use to compute these are in (Rowe, 2009). Performance assessment is implemented in C++ using a Microsoft Sequel Server database.

The metrics we compute for a team or squad of Marines (4-13 people) are dispersion, collinearity, number of clusters, non-Marine interaction, danger,

awareness, 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 location, and centrality of the leader. The issues we observe automatically are of two kinds, those applying to an individual Marine and those applying to the entire group of Marines being monitored. In the first category are a Marine too close to another, a Marine too close to a window or door, a Marine aiming a weapon at another, a Marine excessively exposed to sniper positions, and a Marine not "pieing" (covering a nearby door or window with a weapon). In the second category are groups too clustered, groups too far from one another, groups too collinear, groups in too few clusters, groups without non-Marine interaction, groups moving too fast, groups too close to windows and doors, groups with poor awareness of potential danger, groups with poor weapons coverage, and groups with poor leader centrality.

Figure 2: Example visualization from above and from ground level, with observed problems color-coded.

Metrics and issues are aggregated to provide statistics on average and maximum metrics and numbers of issues per squad and exercise, per squad over all exercises, per behavior category per squad and exercise, and per behavior category per squad over all exercises. This helps instructors to find squads with particular problems, exercises that are particularly difficult, and trends over time of which instructors should be aware.

4.2 Experimental results: metrics and issues

In our first experiments at Sarnoff Laboratories, we had four Marines and two civilians execute a scenario around two small sheds (Figure 3). The scenario included a civilian being searched (coming from the north) and a sniper that had to be captured (from the south), and took around four minutes to perform. Positions and orientations were recorded at 7 hertz and subsampled to 1 hertz. Figure 3 shows the paths followed for one representative run, experiment 415. Marines started at the southeast, took cover from a sniper on the north side of the sheds (black rectangles), handled the civilian (to north) and the sniper (to south), and exited to the west except for the civilian exiting to the north.
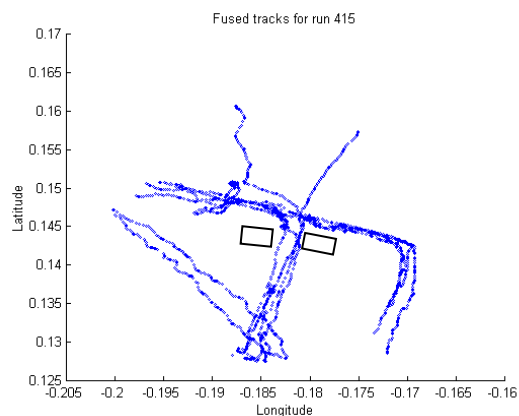

Figure 3: Tracks for experiment 415.

Figures 4-7 show example metrics for experiment 415. Figure 4 shows dispersion. Because they needed to take cover from the sniper, at times they were insufficiently dispersed (values too large). Figure 5 shows roleplayer interactions, which we estimate as times when Marines were facing role players within a minimum distance, a reasonable approximation in the absence of audio (Figure 5). Interactions vary quickly since they are either present or not. Important measurements we can make that are difficult for human instructors to do are the degree of danger to the Marines (Figure 6) and their degree of awareness of it judging by where they are looking (Figure 7). Danger came from potential sniper positions precomputed by analysis of the terrain, including trees and corners of a nearby building. In these preliminary experiments, gaze was estimated by weapon azimuth orientation, which had jitter as Marines moved.
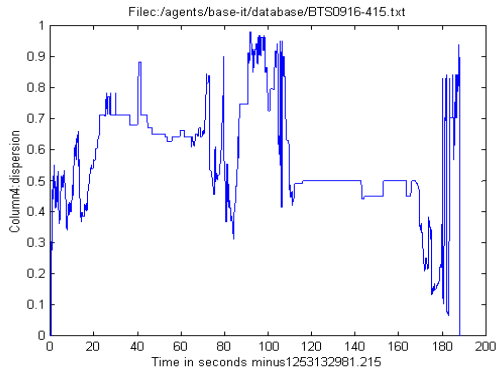
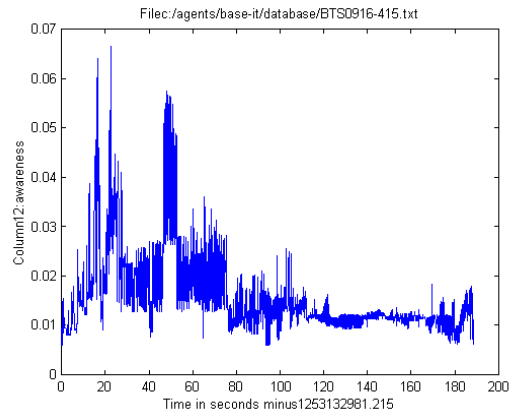Figure 4: Dispersion in experiment 415.
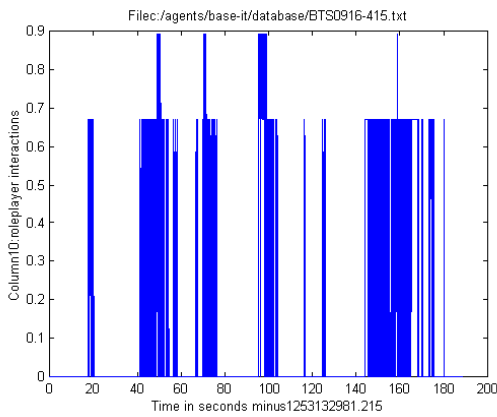

Figure 5: Roleplayer interaction in experiment 415.
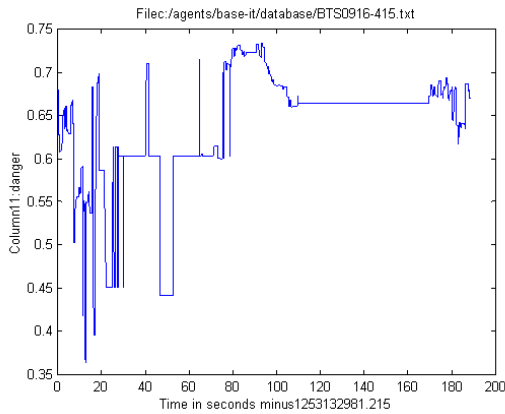

Figure 6: Danger for experiment 415.

Figure 7: Awareness of danger in experiment 415.

These graphs are shown to the instructor after each exercise, making more concrete the evanescent phenomena that occurred. But more important for Marine instructors is the identification of "issues". One way is to display them as dots on a timeline where each row corresponds to a particular issue. Figure 8 shows the plot for experiment 415. Issue 1 is being too close to another Marine, issue 2 is being too close to a window or door, and issue 3 is pointing a weapon at another Marine. It can be seen that the Marines were clustered more than doctrine recommends, in part because of the smallness of the sheds. They came too close to doors when they had to take cover. They also had problems accidentally pointing their weapons at one another when the sniper was arrested. Issues for the group of Marines as a whole were not as important, but a few were noted for this exercise (Figure 9). These graphs can be shown to students, but they are more useful as guides to the instructor before showing video or visualization like Figures 1 and 2 to students.
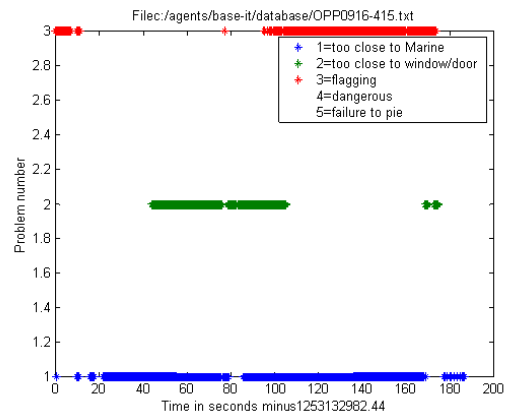

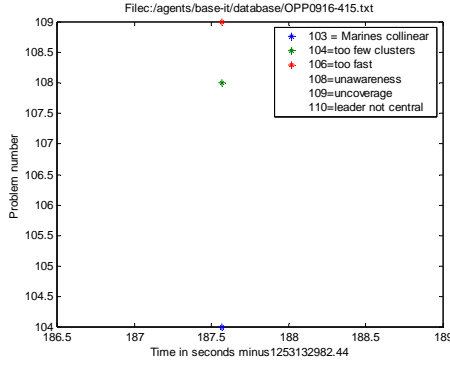Figure 8: Issues for individual Marines in experiment 415.

Figure 9: Issues for the group of Marines.

4.3 Behavioral analysis

A problem with our metrics and issues is that they consider only a narrow context. So we try to infer automatically what state the set of Marines is in, and tabulate metrics and issues separately for each state as in (Minnen et al, 2007).

Figure 10 graphs six inferred behaviors for experiment 415 where height 1 = getting orders, 2 = patrol, 3 = taking cover, 4 = surrounding a target, 5 = roleplayer interaction, and 6 = controlling or directing a roleplayer. We used here a "case-based reasoning" approach where we modeled behaviors by ideal sets of parameters, and found the closest for each time instant. Smoothing was done on both the initial parameters and the inferred behaviors to reduce jitter. Table 1 shows the ideal parameter sets using six metrics (dispersion, clusters at 5m, roleplayer interactions, mobility, speed, and window/door closeness) with weightings of (1, 0.5, 1, 10, 1, 1). Now for Figure 8 we can excuse the first (too close to fellow Marines) and second (too close to windows or doors) issues for the time periods like 43-75 seconds in which the inferred behavior was "taking cover".

For a more general approach, we will be basing behaviors on the Techniques, Tactics and Procedures in Marine manuals and training documents. The set of states, events and properties described there define a vocabulary from which we built an event-detection framework. This framework uses the video and sensor data to classify states and events from a set of known behaviors (Cheng et al, 2009). We will identify about 50 behavioral states using a support-vector machine approach. For each state, we store associated properties including initial classification criteria for the state based on metrics, their triggering events, and their transition states. We also store a "histogram of oriented occurrences"

for each state to aid recognition of complex group activities; it captures the interactions of all entities of interest in terms of configurations over space and time. Taxonomies describe both states and trigger events. For example, patrolling has subtypes of reconnaissance and raids, and involves either single-line, staggered-column, or wedge formations; reacting to a sniper has parts of seeking cover, suppressing the sniper, manuevering, blocking escape routes, and assaulting the sniper. An advantage of such a general-purpose methodology and supporting software is that they can be applied to other types of training situations by using different taxonomies.
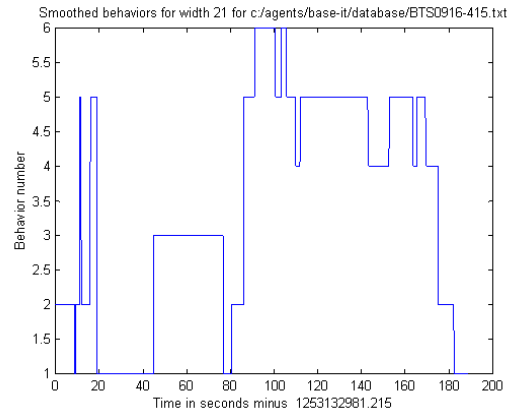

Figure 10: Inferred behaviors for experiment 415.

Table 1: Ideal parameter values for the behavior classes.

	Disp.	Clus.	Inter.	Mob.	Spd.	Wind.
receiving orders	0.5	1	0	0.05	0.1	0.3
patrol	0.2	2	0	0	0.7	0.3
taking cover	0.6	1	0.3	0.1	0	0.6
surround a target	0.8	3	0.2	0.1	0	0.3
roleplayer interaction	0.5	2	0.8	0.05	0.1	0.1
control roleplayer	0.5	2	0.8	0.05	0.7	0.3

Initial assignments of states to times can be improved by using context in the form of hidden Markov models for training activities. We use standard algorithms to make inferences on this model. For instance, if we observe that the Marines make a transition from "patrol" to "take cover", we infer a significant probability that they heard sniper fire. However, we can improve upon hidden Markov models in many cases because some

triggering events may be observed, as when a civilian being searched tries to run away and we see that. A state model permits us to identify new kinds of possible issues for a Marine unit such as forgetting a step in a procedure, performing steps in the wrong order, or repeating steps unnecessarily.

5 EXTENDING THE TECHNOLOGY TO NEW APPLICATIONS

Our technology could make important contributions to physical education and choreography which have previously focused heavily on the performance of the individual in isolation. Just looking at our metrics, instruction in team sports could benefit from measurements of dispersion, clustering, coverage, mobility, speed, and leadership centrality. Instruction in choreography (Smith-Autard, 2004) could benefit from measurements of dispersion, collinearity, lines of sight (from our "danger" calculation), and being too close to objects. Our more global behavior analysis could provide valuable information about pacing for both.

In general, our technology should help quantify a range of physical-motion skills that are historically hard to evaluate fairly (Hay, 2006). (Coker, 2004) provides a taxonomy of errors in motor skills: those due to task constraints, comprehension, perceptions for decisionmaking, decisionmaking itself, recall of previous learning, neuromuscular limits, improper speed-accuracy tradeoffs, visual errors, and proprioceptive errors. Our noninvasive motion-monitoring technology should help particularly with perceptions for decisionmaking, decisionmaking itself, and visual errors, and will indirectly help with recall of previous learning, neuromuscular limits, speed-accuracy tradeoffs, and proprioceptive errors. However, there remain important instructional issues to study in the kind, timeliness, and frequency of the new kinds of feedback from our technology to students, as with any instructional technology.

ACKNOWLEDGEMENTS

REFERENCES

Chen, Y.-J., Hung, Y.-C., 2009. Using real-time acceleration data for exercise movement training with a decision tree approach. *Proc. 8th Intl. Conf. on Machine Learning and Cybernetics*, Baoding, CN, July 2009.

Cheng, H., Kumar, R., Basu, C., Han, F., Khan, S., Sawhney, H., Broaddus, C., Meng, C., Sufi, A., Germano, T., Kolsch, M., and Wachs, J., 2009. An instrumentation and computational framework of automated behavior analysis and performance evaluation for infantry training. *Proc.I/ITSEC*, Orlando, Florida, December.

Coker, C., 2004. *Motor learning and control for practitioners*. McGraw-Hill, New York.

Hay, P., 2006. Assessment for learning in physical education. In Kirk. D., MacDonald, D., O'Sullivan, M., *The handbook of physical education,* Sage, London.

Hixson, J., 1995. Battle command AAR methodology: a paradigm for effective training. *Proc. 27th Winter Simulation Conf.*, Arlington, Virginia, USA, pp. 1274-1279.

Knight, J., Bristow, H., Anastopoulou, S., Baber, C., Schwritz, A., Arvantitis, T., 2007. Uses of accelerometer data collected from a wearable system.. *Personal and Ubiquitous Computing, 11*, pp. 117-132.

Minnen, D., Westeyn, T., Ashbrook, D., Presti, P., Starner, T., 2007. Recognizing soldier activities in the field. *Proc. Conf. on Body Sensor Networks*, Aachen, Germany, March.

Nakatsu, R., Tadenuma, M., Maekawa, T., 2001. Computer technologies that support Kansei expression using the body. *Proc. Intl. Conf. on Multimedia,* September, Ottawa, Canada.

Rowe, N., 2009. Automated instantaneous performance assessment for Marine-squad urban-terrain training. *Proc. Intl. Command and Control Research and Technology Symposium*, Washington, DC, June.

Sadagic, A., Welch, G., Basu, C., Darken, C., Kumar, R., Fuchs, H., Cheng, H., Frahm, J.-M., Kolsch, M., Rowe, N., Towles, H., Wachs, J. and Lastra, A., 2009. New generation of instrumented ranges: enabling automated performance analysis. *Proc.I/ITSEC*, Orlando, Florida, December.

Smith-Autard, J., 2004. *Dance composition, 5th ed.* A&C Black, London.

Stepan, V., Zara, J., 2002. Teaching tennis in virtual environment. *Proc. 18th Spring Conf. on Computer Graphics*, Budmerice, Slovakia, pp. 49-54.

Wachs, J., Goshorn, D., Kolsch, M., 2009. Recognizing human postures and poses in monocular still images. *Proc. Intl. Conf. on Image Processing, Computer Vision, and Pattern Recognition*.