Automatic Performance Evaluation and Lessons Learned (APELL) for MOUT Training

Hui Cheng, Rakesh Kumar, Thomas Germano and Chris Meng
Sarnoff Corporation
Princeton, NJ, USA
{hcheng, rkumar, tgermano, cmeng}@sarnoff.com

ABSTRACT

In order to train war fighters for urban warfare, live exercises are held at various MOUT facilities. Commanders and instructors need to have situation awareness of the entire mock battlefield, and also the individual actions of the various war fighters. The commanders and instructors must be able to provide instant feedback and play through different actions and 'what-if' scenarios with the war fighters. The warfighters in turn should be able to review their actions and rehearse various maneuvers. The system must be able to automatically score the performance of the warfighters and provide them an assessment of their performance.

In this paper, we describe the technologies behind a prototype training system, which tracks and automatically assesses performance of war fighters around an urban site using a combination of ultra-wideband RFID, INS pose, trigger sensor and smart video based tracking. The system is able to:

1. Tag each individual with a unique ID using an RFID system.
2. Track and locate an individual's position, head pose and weapon pose within the domain of interest at all times during an exercise.
3. Associate IDs with visual appearance derived from live videos.
4. Visualize movement and actions of individuals within the context of a 3D model.
5. Store and review activities with (x, y, head pose, weapon pose, gun-trigger, ID) information associated with each individual.
6. Automatically create events and performance metrics for each warfighter. These events are stored in a database. User can click on an event and see the associated video.

An ontology is used to represent the expert knowledge for MOUT training. Using this ontology, the tracks are analyzed, performance metrics and events are automatically created. The metrics of a novice can be compared with the metrics of an expert and overall performance of each soldier can be automatically assessed and measured with each exercise.

Dynamic acquisition and recording of the precise location of individual troops and units during training greatly aids the analysis of the training sessions allowing improved review, critique and instruction.

The prototype training system has been used for simulated Marine Corps exercises and shown improved training efficiency and effectiveness.

ABOUT THE AUTHORS

Dr. Hui Cheng is currently the Technical Manager of the Adaptive and Cognitive Systems Group at Sarnoff Corporation, Princeton, New Jersey, USA. He received his Ph.D. degree in Electrical Engineering from Purdue University in 1999, his MS degree in Applied Mathematics and Statistics from University of Minnesota and his BS degree in both Electrical Engineering and Applied Mathematics from Shanghai Jiaotong University, China. Dr. Cheng was with the Institute of Automation, Chinese Academy of Sciences from 1991 to 1993. Prior to joining

Sarnoff, he was with the Digital Imaging Technology Center, Xerox Corporation. Dr. Cheng's research interests are in the areas of computer vision, image/video processing, pattern recognition, artificial intelligence and statistical image modeling. Since joining Sarnoff, Dr. Cheng has led a number of research and development efforts sponsored by ARDA/VACE program, DARPA/RAID program, NGA Spatial-Intelligence Innovation program, ONR and NIST/ATP program in the areas of UAV and surveillance video understanding, advanced video compression, active imagery acquisition and training and simulation. Dr. Cheng has published more than 30 articles and holds 6 U.S. patents. He is a senior member of IEEE, the Chair of Princeton/Central Jersey Chapter, IEEE Signal Processing Society and a member of IEEE Technical Committee on Multimedia Systems and Application.

Dr. Rakesh "Teddy" Kumar is currently the Senior Technical Director of the Vision and Robotics Laboratory at Sarnoff Corporation, Princeton, New Jersey, USA. Prior to joining Sarnoff, he was employed at IBM. He received his Ph.D. in Computer Science from the University of Massachusetts at Amherst in 1992. He received his MS in ECE from SUNY Buffalo and BTech in EE from IIT-Kanpur, India. His technical interests are in the areas of computer vision, computer graphics, image processing and multimedia. At Sarnoff, he has been directing and performing commercial and government research and development projects in the areas of visual navigation, video surveillance and monitoring, video and 3D exploitation and analysis, object recognition, immersive tele-presence, simulation and training, 3D modeling, medical image analysis and multi-sensor registration. He has been one of the principal founders from Sarnoff for multiple spin-off and spin-in companies: VideoBrush, LifeClips and Pyramid Vision Technologies. He was an Associate Editor for the IEEE Transactions on Pattern Analysis and Machine Intelligence from 1999 to 2003. He has served in various capacities on a number of computer vision conferences and National Science Foundation review panels. He has co-authored one book on Video Registration, more than 50 research publications and has received over 22 patents, with numerous others pending.

Mr. Tom Germano is an Associate Member of Technical Staff in the Vision and Learning Laboratory at Sarnoff Corporation. Prior to joining Sarnoff, he was employed at Vizta3d developing demonstration computer graphics applications. He received his Bachelor of Arts in Computer Science from New York University in 2001. His technical interests are in the areas of real-time computer graphics, computer vision, and multimedia. Since joining Sarnoff, Tom has developed software for the Video Flashlight and APELL systems.

Mr. Chris Meng is currently a Member Technical Staff of the Software Group at Sarnoff Corporation, Princeton, New Jersey, USA. Prior to joining Sarnoff, he was with Video Insight, Inc. He received his M.S. in Computer Science from the University of Houston Clear Lake. He received his M.S. in Chemistry and B.E. in Chemical Engineering from Tsinghua University, Beijing, China. His technical interests are in the areas of digital video surveillance application, SQL database application, GUI, C# and VB.NET. At Sarnoff, he has been a significant contributor to a number of projects in the area of training and simulation and aerial surveillance video processing.

Automatic Performance Evaluation and Lessons Learned (APELL) for MOUT Training

Hui Cheng, Rakesh Kumar, Thomas Germano and Chris Meng
Sarnoff Corporation
Princeton, NJ, USA
{hcheng, rkumar, tgermano, cmeng}@sarnoff.com

INTRODUCTION

The success of urban warfare heavily depends on close-quarter small team operations, such as room clearing. For training of urban warfare, live exercises are held at various MOUT facilities. To measure the performance, give feedback and conduct after-action-review, commanders and instructors need to have situation awareness of the entire mock battlefield, and also the individual actions of the various war fighters. The commanders and instructors must be able to provide instant feedback and play through different actions and 'what-if' scenarios with the war fighters. The war fighters in their turn should be able to review their actions and rehearse different maneuvers. The system must be able to automatically measure the performance of the warfighters and provide feedbacks.

In this paper, we describe the technologies behind a prototype training system (APELL), which tracks and automatically assesses performance of warfighters around an urban site. To track war fighters' actions during an exercise, our system uses a sensor suite including an ultra-wideband Radio Frequency Identification (RFID) and Tracking system (Fontana, 2004), Inertial Navigation System (INS), trigger action capture system and smart video capture system. The prototype training system is able to:

- Tag each individual with a unique ID using RFID .
- Track and locate each individual's position, head pose and weapon pose within the domain of interest at all times during an exercise.
- Associate IDs with visual appearance derived from live videos.
- Visualize movement, actions and events of both individuals and the whole team within the context of a 3D model.
- Store and review activities with (x, y, pose_head, pose_weapon, gun-trigger, ID) information associated with each individual.
- Automatically detect events of interest, such as incorrect procedure or action, for each war fighter. These events are stored in a database. A user can click on an event and see the associated video.
- Automatically compute performance metrics and generate various statistics to measure team performance and training progress.

A training ontology is used to represent the expert knowledge for MOUT training. Using this ontology, the tracks and actions of war fighters are analyzed; events are detected, logged; and performance metrics are automatically created. The metrics of a novice can be compared with the metrics of an expert and overall performance of each soldier can be automatically assessed and measured after each exercise.

Dynamic acquisition and recording of the precise location and action of individual troops and units during training greatly aids the analysis of the training sessions allowing improved review, critique and instruction. The prototype training system has been implemented and used for mock Marine Corps room clearing exercises. The experimental results and participants' feedback has shown improved training efficiency and effectiveness.

APELL SYSTEM

As shown in Figure 1, the prototype training system has six major components: (1) *Sensor System* that captures and computes each warfighter's location, pose and action. (2) *Event Detection* that creates events of interest based on the sensor system's outputs. (3) *Training Ontology* that captures expert knowledge including procedure and strategies of MOUT operation. (4) *Automated Performance Evaluation* that computes performance metrics for each individual warfighter and the entire team according to the training ontology. (5) *Exercise Database* that stores each warfighter's location, pose, action, detected events and performance metrics for After-Action-Review (AAR). (6) *After-Action-Review and Visualization* that provides both an iconic view of movement, actions and events in a 3D environment and a synchronized video display by combining all video feeds onto a 3D model of the MOUT environment.

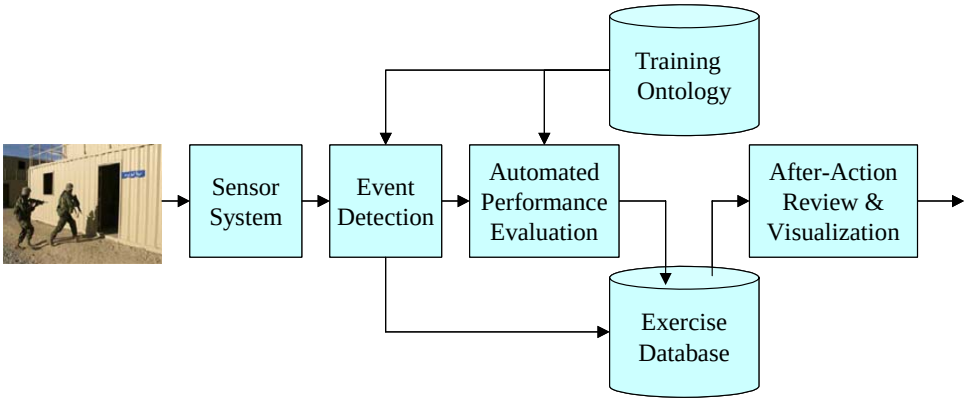


Figure 1. APELL system block diagram.

Sensor System

The sensor system consists of ultra-wideband RFID tags and receivers, INS modules, trigger sensors, microphones[1] and video cameras. Video camera and RFID receivers are mounted in strategic locations around a MOUT facility. Each war fighter carries an RFID tag, INS module, trigger sensor and MIC. The INS, and trigger sensor data are captured by a soldier-worn computer and the data is wirelessly transmitted back to the central processing workstation for tracking, event detection and performance measurement.

Radio Frequency Identification and Tracking

The prototype training system uses a short-pulse ultra-wideband RFID-based location tracking (Fontana, 2004). In the MOUT facility, a number of radio frequency receivers and reference Radio Frequency (RF) tags are installed. Each participant of an exercise carries an RF tag that emits a short-pulse ultra-wideband RF signal at 4-10 Hz. The RF signal emitted by each RF tag has a unique signature that will be used by the receiver to identify the tag. Using time-difference of arrival techniques and triangulation, the location of each participant relative to the reference tag is computed to one-foot accuracy. The location computed by RFID can be further improved by combining video and stereo-video based object detection and tracking (Zhao, etc., 2005).

Inertial Navigation System (INS)

The Inertial Measurement Unit (IMU) is used to measure the gun pose, direction of where the gun is pointing, and the head pose. The output of the IMU is a three dimensional vector consisting of roll, pitch and yaw components, (r, p, y). Given the location of the warfighter from the RFID tracking system and his head and gun pose, the system computes in absolute world coordinates where the warfighter is looking and where his gun is pointing for every instant in the exercise.

Data Processing and Feature Extraction

The data and processing flow is shown in Figure 2. The RFID data together with video data is used to generate persistent track information of each participant. It is represented as the location information (x,y) at time t.

The gun-pose and head-pose captured by the IMU in the form of (r,p,y) at time t and trigger data, g are then synchronized with track data. Finally, a ten-dimension feature vector combining all sensor outputs and the participant ID is stored in the Exercise Database. The feature vector, $F(t)$, is

$F(t)=[x, y, r_{gun}, p_{gun}, y_{gun}, r_{head}, p_{head}, y_{head}, g, ID]$.

AUTOMATED EVENT DETECTION AND PERFORMANCE EVALUATION

In addition to capturing and storing participants' location, pose and action throughout the exercise, our system automatically detects events of interest and computes the performance of both the individual war fighter and the team as a whole. The automated event detection and performance evaluation are guided by the training ontology that defines the procedures and strategies of the MOUT operation. In our prototype system, we focused on one aspect of the MOUT operation: room clearing. However the system and methodology itself is very general and can be applied to other training scenarios by using other ontologies. The following room clearing ontology is extracted from discussions with Marine MOUT training instructors.

[1] Not implemented in the prototype system.

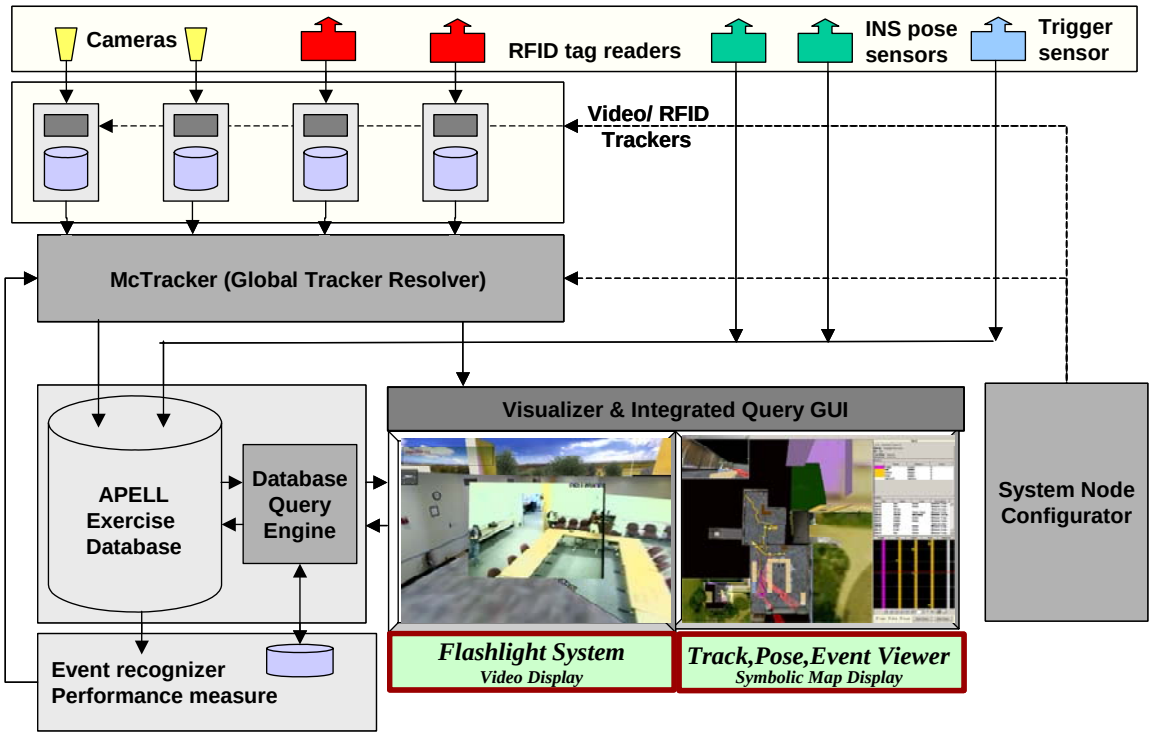


Figure 2. System data and processing flow diagram.

Room Clearing Ontology

Room clearing is a key aspect of urban operation. It is extremely dangerous and needs decisiveness, discipline and coordination among all team members to minimize casualty and maximize effectiveness. Some key stages of room clearing (Figure 3) are:

1. Moving along the hallway approaching a room
2. Staging outside a room
3. Entering room
4. Hasty clear
5. Thorough clear
6. Leaving room

In different stages, different procedures need to be followed and different performance metrics are used. One example is the Hasty Clear.

Hasty Clear is carried out immediately after entering a room. To maximize the firepower and dominate a room, the fire team needs to quickly spread out and occupy strategic locations in a room, such as corners. The strategic locations, referred to as *s-zones* in this paper, are determined by the layout of a room. When moving to a zone, a warfighter should only engage immediate threats and trust team members to engage other threats if needed. After reaching one's zone, the war fighter should scan his/her area of responsibility and call either "Threat" or "Clear".

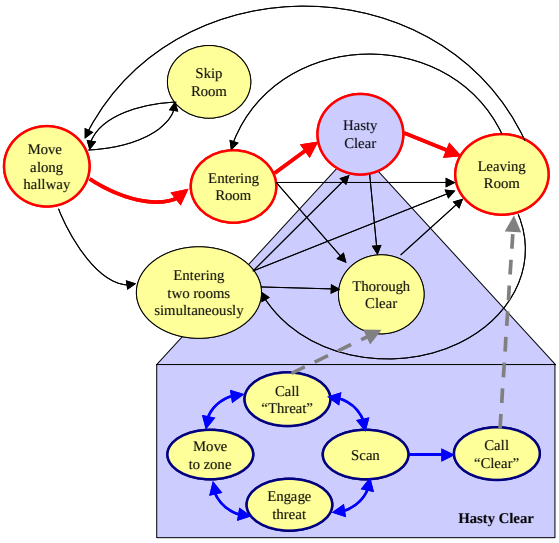


Figure 3. Key stages of Room Clearing and Hasty Clear.

With the help of Marine Corps MOUT instructors, we have developed a room clearing ontology to use for

event detection and performance evaluation. The preliminary ontology is captured using the OWL representation. The key stages of the room clearing operation are coded in the event detection and performance evaluation modules. A GUI is also developed to allow a trainer to pre-specify s-zones as shown in Figure 4.

Event Detection

Directly measuring performance from a time series of ten-dimensional feature vectors as discussed is both difficult and time-consuming. Therefore, we first detect a set of events to partition the time series into segments and measure performance of each warfighter during each segment.

The first category of events detected in our system are zone-based events. These events include entering or leaving a pre-defined zone. Typical zones are hallway, staging area (right outside a door), door, room and s-zones. Figure 4 shows the zone definition for the MOUT room clearing training area. For the prototype system a GUI is developed to define s-zones manually and the main direction of coverage while in an s-zone.

The second category of events is related to the weapons. These include detecting how the weapon is carried and used. For weapon related events, we detect the following events:

- Ready Carry: pointing gun up and ready to shoot, with warfighter gaze aligned with gun-sight.
- Alert Carry: gun pointing 45 degree down to the ground.
- Muzzling or flagging: warfighter points gun at another friendly fellow warfighter or civilian.

We also detect shot related events:

- Firing and hit
- Live fire muzzling: one friendly war fighter shooting another fellow warfighter or civilian.
- Negligent discharge: shot fired at no target.

All detected events including the time they happened, and participant(s) involved are stored in the exercise database.

Performance Evaluation

Based on the detected events, performance metrics are computed to measure both individual performance and team performance. These metrics are also used to identify mistakes and to guide the AAR for lessons learned.

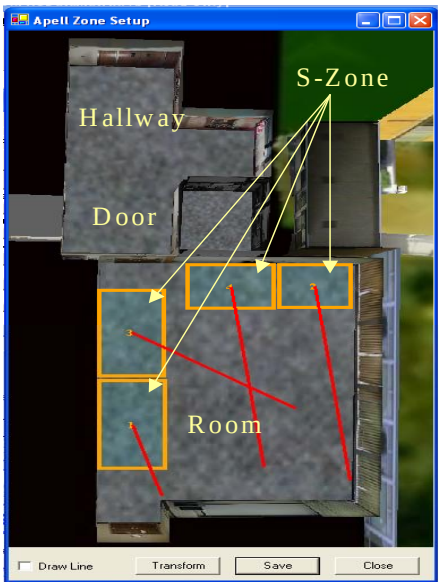


Figure 4. Zone definition for the MOUT room clearing training exercise. Red lines indicate the main direction of coverage while in s-zone.

Different performance metrics may be computed at different stages of the room clearing. For example, in the hallway, the compactness of team formation, whether or not stack order is maintained and the weapon status for each team position is computed and checked for correctness. In hasty clear, the time used to reach all s-zones and the effectiveness of engagement is computed. For the mock Marine Corps exercises, the following metrics were computed:

- Time used to complete the mission
- Number of muzzling events and live fire muzzling events
- Time to reach door
- Time to enter room
- Time to reach s-zone
- Time to clear room
- Compactness of the team during transition
- Maintaining of stack order
- Weapon status

VISUALIZATION AND AFTER-ACTION-REVIEW

To improve training effectiveness, a training system must be able to allow users including both trainees and instructors to quickly access information, such as lessons learned collected during an exercise for After-Action-Review. The APELL system provides a suite of visualization tools that allow a user to view not only

videos captured during an exercise, but also tracks, poses, and events computed by the system in an interactive and easy-to-use manner. Two displays are provided by APELL and they are used simultaneously in a synchronized fashion. They are (1) Symbolic Map Display and (2) Video Flashlight Display.

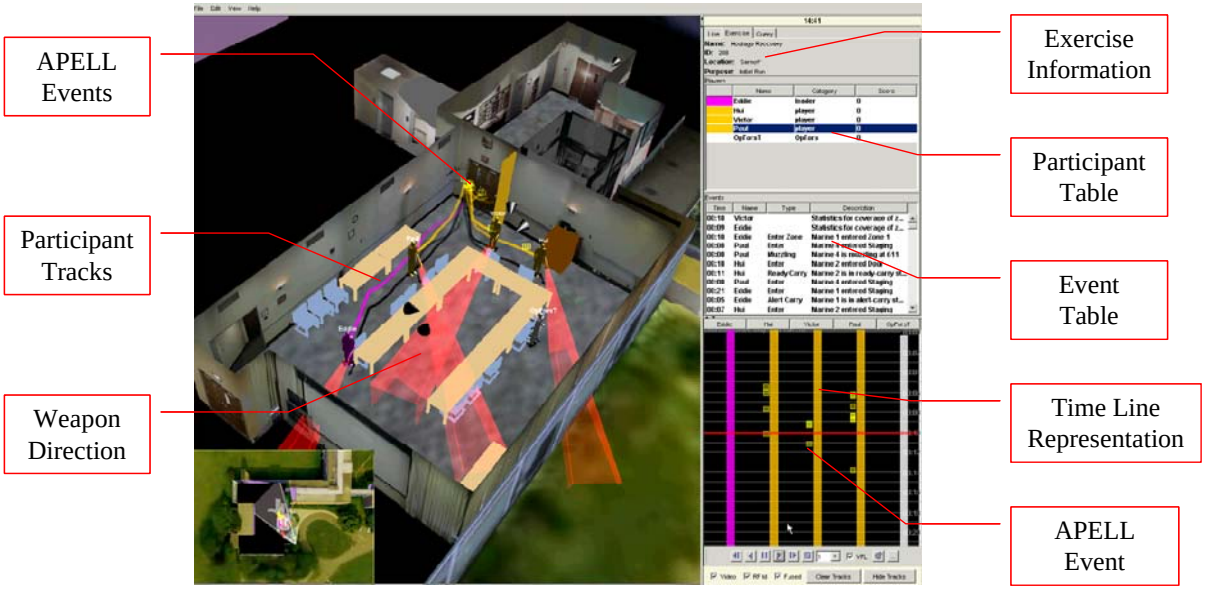


Figure 5. APELL Symbolic Map Display.

APELL Symbolic Map Display

In Figure 5, we show a snapshot of a typical exercise as it is displayed by APELL Symbolic Map Display. The left side of the Symbolic Map Display shows a 3D model of the MOUT environment. Tracks, gun poses, and head poses (not shown) of participants, including war fighters (with ID) and OPFORs are overlaid on the 3D model (Figure 5). Events detected by the system are marked by symbols on the track. A user can view the tracks from different view angles and can click on any event symbol to view the event description. A user can also turn on/off the gun pose or head view of any participants.

The right side of the display (Figure 5) consists of four different parts. General exercise information including exercise ID, starting, ending time and exercise duration are shown at the top. The participant information is shown below the exercise information. The event table consisting of time, participant, the type of the event and description is shown in the middle. A time line representation of the exercise overlaid with events is shown at the lower-right corner.

The Symbolic Map Display allows users to view an exercise at any instant and track movements forward and backward in time. A user can also drag the red time line to any location and play back from there. Additionally, the user can synchronize the Symbolic Map Display with the Video Flashlight Display to view the symbolic representation and video at the same time.

Video Flashlight Display

In the APELL system, multiple video cameras are used to cover a MOUT facility to capture the entire exercise. To better view the videos captured by different cameras, Sarnoff has developed a Video Flashlight Display system (Kumar, etc., 2003) (Hsu, etc., 2000) that can seamlessly integrate multiple video streams into a unified live display. Each of the videos is projected on the 3D model of the MOUT environment and used to update the model textures at 30 Hz. The dynamically textured model is rendered and displayed by the Video Flashlight system. By controlling the viewpoint and the view angle, a user can get a birds-eye view of all the activity covering the area of regard or the user can zoom in and focus on an object or person of interest. An example of the Flashlight Video display is show in Figure 6, where two video streams are projected onto the 3D model of the room being cleared by a four-man Marine Corps team.

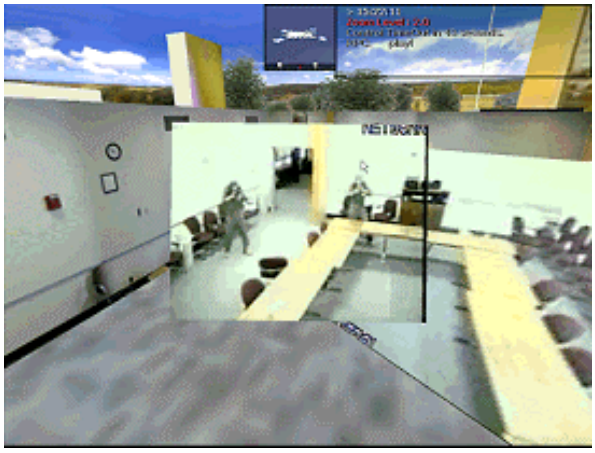


Figure 6. The Flashlight Video Display.

EXPERIMENTS

The prototype APELL system was implemented and used for two mock Marine Corps exercises. Two four-person fire teams (novice and expert) participated in each exercise. The novice team consisted of Sarnoff employees or Marines who have not had training in room clearing. The expert team consisted of Marine Corps MOUT training instructors and people with MOUT experience.

Each team participated in 40 exercises that were divided into a 20 exercises morning session and 20 exercises afternoon session. After-action-reviews were conducted between the morning session and the afternoon session using our system.

The stack order, i.e. the position of each team member was randomly generated for each exercise. Also, in each exercise, the number of OPFORs varied from zero to two. Whether or not there were zero, one or two OPFORs was also decided randomly. The fire team did not know whether or not OPFORs would participate in a particular exercise in advance.

Some of the metrics computed by our system are shown in Figure 7 and Figure 8. Figure 7 plots the number of muzzling events for each exercise. Figure 8 (a) and (b) show time-based performance metrics including time to complete the exercise, time to reach s-zones, time used for room clearing (hasty clear), time in the hallway, entering the room and exiting the room. Both the time-based statistics and the number of muzzling events show that there is a significant difference between the experts' performance and novices' performance. This indicates that the metrics used in the prototype system are meaningful for measuring the performance for training warfighters in room clearing. However, due to the number of Marines who are available to participate in our experiments, we could not establish a control group for performance comparison.

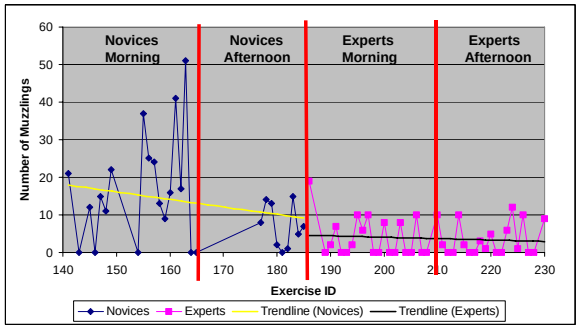


Figure 7. Muzzling statistics of the mock Marine Corps exercises.

Additionally, there was a visible performance improvement after the after-action-review using our system. The improvement is especially significant in the reduction of the number of muzzling events. Muzzling is a big problem for the novice team. In the morning session, the novice team had an average of 14 muzzling events per exercise. After AAR using the system, the novice team achieved an average of 4.3 muzzling events per exercise in the afternoon runs. This is close to the average of 2.4 muzzling events the expert team had before the AAR and is a significant reduction from the average of 14 muzzling events per exercise for the novice team in the morning before the AAR.

The APELL system automatically detected these muzzling events and flagged them for the instructors and trainees. Using the visualization and AAR sub-system, instructors and trainees were quickly able to retrieve and view all muzzling events. The instructor was able to review each of the muzzling events with the Warfighters and point out mistakes made and teach them how the weapon is correctly carried and used during an exercise.

From Figure 8(a) and Figure 8(b), we can also observe significant improvement of time based measures for the novices after the AAR. For example, the average total amount of time used to complete an exercise of the novice team reduced by 15% after AAR. The average time spent by the expert team also reduced by 20% after the AAR review.

The prototype training system also detects events that are difficult to detect by instructors. In Figure 9, we show a muzzling event occurring across the extent of the room and not observable in any one camera. In Figure 9(a), the muzzling is clearly visible in the Symbolic Map Display and is detected by the automated event detection algorithm. In Figure 9(b) and 9(c), we show the video frames containing the two participants involved in the muzzling. The participant marked by the red circle in Figure 9(b) accidentally pointed his weapon at the participant marked by the red circle in Figure 9(c). Since the two participants are far from each other, without the APELL system, it would be very difficult for the trainer to spot these kinds of muzzling events.

CONCLUSION

A prototype training system (APELL) was developed that captures tracks, poses and actions of the participants and automatically assesses the performance of war fighters in the MOUT training environment using a training ontology. This prototype system was used for simulated Marine Corps exercises. The results show that the prototype system can accurately detect mistakes, such as muzzling events, automatically. The experimental results also show the improvement of training and AAR effectiveness through the use of the system and its AAR tools. Using the APELL system, we saw a significant reduction in the number of muzzling events in the novice after one AAR session. The average number of muzzling events dropped from 14 to 4.3 muzzling events per exercise. Compared with 2.4 muzzling events per exercise for the expert team, it shows the significant improvement achieved by using the AAR tools.

In future work, we plan to develop a closed-loop training system where the training scenarios and difficulty levels are automatically adjusted in real-time based on the performance of the war fighters. We plan to enhance the prototype system by using controllable and repeatable stimuli, for example synthetic OPFORs projected on surfaces in the MOUT by a stereo projection system. The behavior and number of these OPFORs would be varied based on the performance of the Warfighters in previous exercises.

We also plan to expand our MOUT training ontology to capture not only the expert knowledge of MOUT operations, including procedure and strategies of Warfighters, but also the behavior and strategies of OPFORs and civilians. The ontology may be used both for performance evaluation and for control of the behaviors of OPFOR and civilians in the exercise. Finally, the entire training system could easily be modified and extended by changing the ontology to cover a larger range of MOUT training exercises.

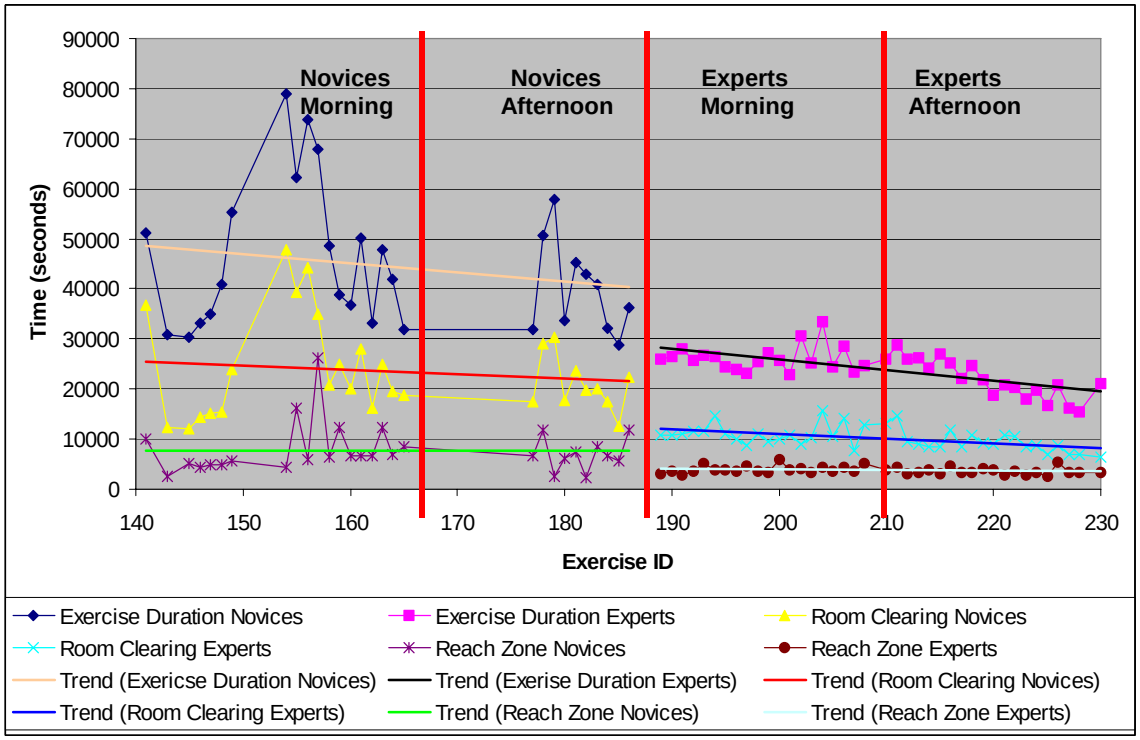
REFERENCES

Hsu, S., Samarasekera, S., Kumar, R., & Sawhney, H.S. (2000), Pose Estimation, Model Refinement, and Enhanced Visualization Using Video, *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, Hilton Head Is., SC, vol. I, 488-495.

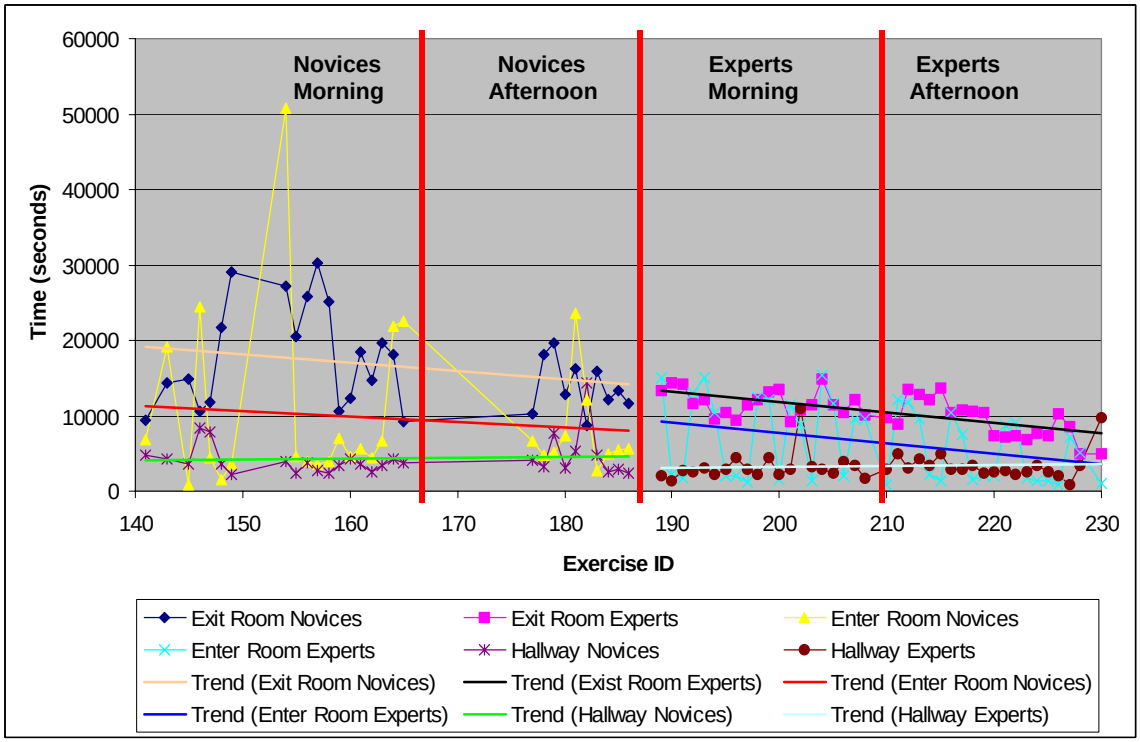
Zhao, T., Aggarwal, M., Kumar, R., & Sawhney, H.S. (2005), Real-time Wide Area Multi-camera Stereo Tracking, Proc. IEEE Conf. on Computer Vision and Pattern Recognition, San Diego, CA.

Kumar, R., Samarasekera, S., Arpa, A., Aggarwal, M., Paragano, V., Hanna, K., Sawhney, H.S., & Sartor, M. (2003), Monitoring Urban Sites using Video Flashlight and Analysis System, GOMAC Proceedings, Tampa Florida.

Fontana, R.J. (2004). Recent System Applications of Short-Pulse Ultra-Wideband (UWB) Technology. IEEE Transaction on Microwave Theory and Techniques, vol. 52 no. 9, September 2004, pp. 2087-2104.

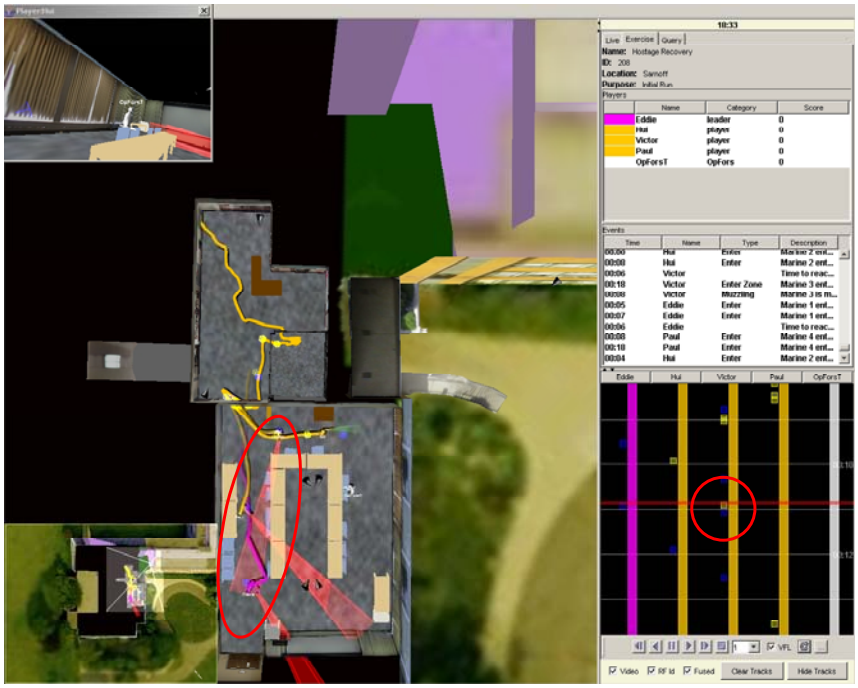


(a)



(b)

Figure 8. Time based metrics computed of the Marine Corps exercises. In order show each curve clearly, we break the six time based metrics into two groups and plot them separately in (a) and (b).



(a)

(b)

(c)

Figure 9. Muzzling event detection by APELL, though the muzzling involves participants across the room from each other. (a) Muzzling is clearly seen in Symbolic Map Display and automatically detected by APELL shown in the time line. (b) Video frame shown the participant marked in the red circle who accidentally pointed his gun at the participant in Figure 9(c) marked in the red circle. Since the two participants are across the room, no a single video frame captures both of them during the muzzling.