Visual Analysis and Filtering to Augment Cognition

Mathias Kölsch, Juan Wachs, and Amela Sadagic

jpwachs@purdue.edu, asadagic@nps.edu
http://movesinstitute.org/~kolsch

Abstract. We built and demonstrated a system that augments instructors' sensing abilities and augments their cognition through analysis and filtering of visual information. Called BASE-IT, our system helps US Marine instructors provide excellent training despite the challenging environment, hundreds of trainees and high trainee-to-instructor ratios, non-stop action, and diverse training objectives. To accomplish these objectives, BASE-IT widens the sensory input in multiple dimensions and filters relevant information: BASE-IT a) establishes omnipresence in a large training area, b) supplies continuous evaluation during multi-day training, c) pays specific attention to every individual, d) is specially equipped to identify dangerous situations, and e) maintains virtual vantage points for improved situational awareness. BASE-IT also augments and personalizes the after-action review information available to trainees.

This paper focuses on the automated data analysis component, how it supplements the information available to instructors, and how it facilitates understanding of individual and team performances on the training range.

Keywords: Augmented cognition, information analysis, training range instrumentation.

1 Introduction

The keystone in US Marine training is conducted at a purpose-built training range at the Marine Corps Air Ground Combat Center in Twentynine Palms, CA, which provides a realistic environment that immerses hundreds of trainees in a small town with houses, markets, road traffic and human role players. Instructors observe the non-stop 72 hour urban operation and provide performance feedback at regular intervals. This is a challenging situation for a number of reasons.

First, this training is the final and most realistic training provided to US Marines immediately before their deployment into theatre. There is immense pressure on both trainees and trainers to achieve the training objectives as it will have a tremendous impact on the performance in theatre. Techniques not taught or not taught well, as well as mistakes not caught in training might have severe and far-reaching consequences later.

D.D. Schmorrow and C.M. Fidopiastis (Eds.): AC/HCII 2013, LNAI 8027, pp. 695–702, 2013.

Second, the training facility is very expensive due to many factors, necessitating efficient and effective training. Due to these pressures on expedience and performance, as well as due to accepted best training practices, training is rarely stopped to provide feedback. In fact, the more advanced scenarios train and evaluate multiple skills simultaneously, for many individuals, with little to no room to pause and discuss or correct mistakes until the after-action review.

Third, the training range is almost one square mile large, with many buildings, roads, foot and vehicle traffic, geographic features and other realistic aspects of a small town. It is impossible for instructors to always keep an eye on and to give feedback to every trainee on individual or group behavior. Instead, observations are made at crucial times and locations, and feedback is provided in short debriefing sessions.

Fourth, some aspects of individual and group behavior are very difficult to observe from a single vantage point. For example, the precise position of an individual and his head location behind cover cannot be determined accurately from just one point of view. Also, the formation of a squad that is on foot patrol in between buildings is hard to observe from just one location due to occlusions. Viewpoint limitations might cause actions to pass unobserved and evading evaluation and feedback.

Fifth, while video streams from cameras on the range are available and are being recorded, the sheer amount of data relegates their use to isolated review questions. Rather than augmenting the instructor's *cognition*, this additional information requires *additional attention*. Also, while pole-mounted pan-tilt-zoom (PTZ) cameras are available, aerial cameras are not. Hence, occlusions from buildings in a single view are common.

Sixth, due to the aforementioned constraints, some behavioral mistakes cannot be focused on during this training, as it would distract from the main training objectives. One such mistake is unintentionally pointing a weapon system towards a fellow Marine, also called flagging. Additionally, flagging is difficult to determine unless an instructor happens to be very near the occurrence and paying attention to the swift movements of trainees.

BASE-IT, short for Behavioral Analysis and Synthesis for Intelligent Training [1], was developed at the MOVES Institute at the Naval Postgraduate School, the University of North Carolina, Chapel Hill, and the Sarnoff Corp. (now part of SRI). The goals of BASE-IT are to address some of these difficulties and:

- to improve the preparation of trainees before their arrival,
- to supplement the information available to instructors, both in real-time during the exercise and for after-action review (AAR), and
- to automatically generate AAR resources for individual feedback.

The main components of BASE-IT include an automated camera management and sensing system, automated individual performance evaluation, automated analysis of unit behaviors, 3D visualization of recorded data sets with the ability to search for significant events, and automated behavior synthesis for exploration of 'what-if' scenarios.

Fig. 1. Observing in a "prone" posture, crossing a danger area, and results from our posture recognition method

This paper focuses on the extensive video analysis component, how it supplements the information available to instructors, and how it facilitates understanding of individual and team performances exhibited on the training range.

2 Related Work

Military training and performance measurement has long received tremendous attention. BASE-IT was built together with the Marine Air Ground Task Force Training Command at the Marine Corps Air Ground Combat Center at Twentynine Palms, California, which has one of the most advanced training facilities of the nation. Similar in instrumentation but without the analysis component is the Future Immersive Training Environment (FITE) at, Camp Pendleton's I Marine Expeditionary Force in California and at Camp Lejeune, NC (see, for example, [2, 3]). FITE's focus is on providing a training experience through augmentation, whereas BASE-IT as discussed here focuses on providing augmentations to the instructors and, particularly, to help analyze training. The US Army has similar training range instrumentations, for example, the Combat Training Center Military Operations on Urban Terrain Instrumentation System (CTC MOUT-IS) video system [4].

3 Solutions for Overcoming Cognition Limitations

Here we discuss the six limitations of unaided human cognition and what solutions we have applied to augment instructor cognition.

3.1 Omnipresence

The simulated town at the US Marine base at Twentynine Palms is nearly one square mile large and has many roads, houses, creeks, and other typical elements of any inhabited location. This prohibits instructors to have good visibility of all locations. We installed pole mounted fixed and PTZ cameras – a "sea of cameras" – to achieve omnipresence even in otherwise view-obstructed locations. Omnipresence through cameras provides augmentation of the spatial field of information. However, while this enables an instructor to virtually be at any one

of multiple locations, paying attention to multiple data streams simultaneously is difficult at best. Hence, we also require some degree of automated analysis of these additional data streams.

3.2 Continuous, Always-On Coverage

Human observation of dozens of live video feeds is impracticable if not infeasible due to the number of cameras and the continuous, always-on, non-stop 72 hours training scenario. Instead, we trained computer vision methods on the specific clothing, backpacks and helmets to detect US Marines and to estimate various body posture and weapon parameters, thereby filtering out empty scenes and scenes without any trainees. Night-time operations were observed to the degree possible with visible-spectrum cameras, plus the GPS and accelerometers on trainees and weapons. This presumably improves the instructor's ability to absorb information, essentially expanding the temporal horizon ("always-on"), the temporal resolution (several measurements per second), and the spatial extent (omnipresence). Despite the increased spatio-temporal field of view, information processing (see subsequent subsections) keeps the data volume manageable, and cognition augmented.

3.3 Posture Recognition and Head Localization

Fast action, multiple trainees, and the instructor's vantage point often prohibit precise estimates for the trainees' body and head positions. Yet these are important, for example, in order to determine whether a Marine has sought sufficient cover in case of enemy fire. We built posture recognition methods that can determine whether a Marine is standing or taking a knee, and we custom-trained head detection methods on the specific helmets to precisely locate them in the 3D environment [5, 6]. This offloads the spatial reconstruction task from the instructor to the computer and permits eyes-on *more* trainees at any time.

3.4 Monitoring Security

It is vital for US Marines to maintain "360° security" at all times, requiring coordination between individuals to visually scan in all directions as a team. Again, it is difficult if not impossible for instructors to assess this continuously, particularly if part of a team is hidden from view. Our computer vision methods automatically estimate the torso (shoulder) orientation and the head orientation of each trainee. A subsequent performance analysis module [7] monitors this information for the entire team and flags incidents of likely incomplete situational awareness. Cognition is augmented spatially again, around corners and through occlusions. It is also augmented through simultaneous assessment of head orientations of *all* squad members and automated calculation of the 360° coverage.

3.5 Identifying Weapon Flagging

One of the performance traits that is continuously observed and evaluated is flagging – unintentionally pointing a weapon system towards a fellow US Marine. Our system continuously determines the orientation of the weapon system with acceleration sensors and vision-based processing. It then checks against known positions of nearby US Marines, and identifies the times and places where incidents of flagging happened, including the identification of the individual who caused each flagging incident. Such a list of incidents speeds the instructor's comprehension of the trainee's performance. Further, it provides a second, unbiased look at trainees through the eyes of other modalities.

3.6 Foot Patrol Analysis

Another important team behavior concerns patrol formations and their dispersion across the terrain, that is, the distance between individual trainees and their spatial configuration. Depending on the situation, it is more or less dangerous to be close to each other or further apart, to walk in single file or offset, and so on. Similarly, foot patrols need to "cross danger areas" in a particular fashion: running, not walking, and not all at once (see Fig. 1). The BASE-IT performance analysis module utilizes the precise position estimates from our visual analysis to measure distances and velocities and to provide pre-analyzed results to the instructors. Again, these objective measurements supplement the subjective and often incomplete instructor's observations.

4 Results

Does BASE-IT indeed augment the instructors' cognition? This hypothesis can ultimately only be answered by directly measuring the cognition, either through objective means or through a questionnaire that assesses cognition. Neither of these options was viable for BASE-IT due to time and financial constraints, as well as due to the difficulty of constructing a control group of instructors for these one-time training actions. The approach taken here determines whether the instructors were given information that conceivably would *result* in augmented cognition.

Figure 2 depicts the various augmentations to the information available to instructors: additional spatial and temporal information, information in additional sensing modalities, and, last but not least, pre-processing and filtering of information. But let us take a closer look.

Spatial Augmentation. Merely providing information about previously inaccessible areas can suffice to make instructors cognizant of a situation they had no previous information on. For example, a foot patrol formation that was previously out of view comes into view with the help of our cameras. Provided the instructor looks at the imagery, he will become cognizant of this information. He will be able to have eyes on more trainees and avoid occlusions.

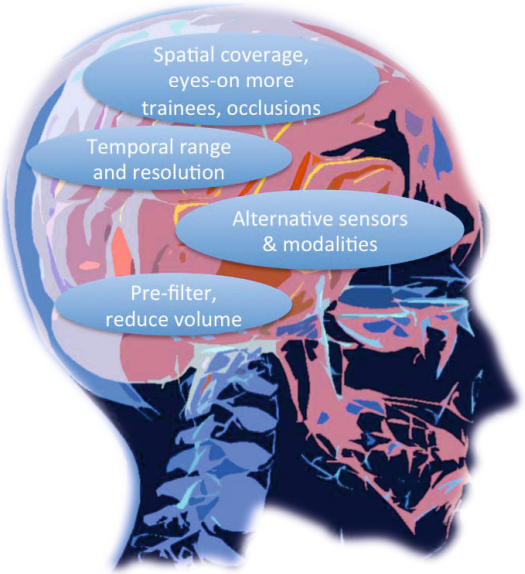

Fig. 2. Augmentations to human cognition

Temporal Augmentation. The cameras provide continuous coverage in lit areas, not taking a break and not getting tired. They also capture events at a frame rate that permits analysis of short-lived actions such as a brief weapon flagging event. As before, this increased temporal duration and resolution of data makes information accessible to the instructors, but they still have to actively seek out this information.

Sensory Augmentation. Since the trainees are tracked by visual information and GPS sensors, and their weapons' orientations are tracked with specific "Inertial Navigation System" (INS) sensors mounted on the weapons, "more than what meets the eye" helps determine locations and orientations of US Marines and weapons. These additional sensors and measurement modalities again increase the *amount* of information available to the instructors.

Pre-cognitive Filtering. Naturally, more information does not directly result in better understanding or cognition, just like the wealth of information available on the internet does not immediately translate into smarter surfers. However, filtering the information to only the most relevant aspects and thereby reducing its amount increases the chances that an instructor will find time to inspect it, especially if this information cannot easily be gleaned from other sources. Similarly, a site of distinct and mostly relevant information is more likely to be visited. BASE-IT provides pre-processed information that obviates the need for tedious video review and instead makes the most pertinent information available immediately. For example, presenting instances of weapon flagging or cases of "bunching up" is clearly much more useful than requiring review of hours of mostly uneventful video.

Note that the computer does not produce final results or even make the decisions. Instead, automated visual analysis and filtering "merely" improves the scene that is presented to the human. This is an important consideration for applied computer vision systems since it is unrealistic to expect perfect performance when translating recent research into practical application. Note also that BASE-IT distinguishes the two phases of data acquisition and processing in the terms "sensing" and "sense making." Sensing includes sensor management system, tracking of individuals (including pose and posture), and sense making includes automated behavior analysis and performance evaluation. Omnipresence and continuous coverage fall under the "sensing" aspect and the remaining solutions are mostly "sense making," albeit they use the additional sensor data from video or accelerometers, for example.

To illustrate the capability of pre-processing, we repeat here the results of the BASE-IT automated video analysis [6]. Using the automated video analysis, Marines were detected successfully (in uncluttered conditions) in 98.73% of the tested instances. When the subjects were partially occluded, the recognition was negatively impacted and only 53% of the torso orientations were correctly identified. The number of correctly classified instances (per marine and per frame) was determined to be 76% and 72% for the torso and head, respectively (see the confusion matrices in [6]). Speed performance tests showed that the detection task was accomplished in 1.9 seconds and that it scaled sub-logarithmically with an increase in image size. The combination of per-frame detection and posture recognition with semantic consistency checking and temporal smoothing [5] provides sufficient accuracy for determining tracks. These tracks can then be analyzed further for troop formation [7]. This is a task that is difficult to perform for human instructors, as discussed in Sec. 3.6, hence we consider BASE-IT augmenting the instructor's cognition.

By stressing salient activities and filtering out unimportant aspects we reached our objective of radically improving the control of and insight into the training exercise, enabling detailed after action review within minutes of completion of the exercise, and further enhancing and supplementing an already invaluable training experience.

5 Conclusions

Training US Marines for complex situations requires training in a complex environment, which poses a great challenge to instructors and their ability to assess the trainees accurately. In this paper, we described how BASE-IT attempts to improve upon the information available to instructors in the hope that it improves their understanding and analysis of the trainees.

Our experience shows that only in tandem does more information and its pre-processing truly augment cognition. BASE-IT pays specific, uninterrupted attention to individuals, anywhere on the range, with help of a multi-modal sensor suite, and through multi-stage analysis modules. BASE-IT provides value as a tool for both instructors and trainees, both for training preparation and for personalized review and analysis (AAR). In the near future, we expect many more

tools that pre-process the "big data" from training observations and, together with a human in the loop, permit semi-automatic analysis and much-improved feedback to the trainees.

Acknowledgements. This work was funded under ONR-BAA-05-023. We thank the volunteer contributors and excellent collaborators in the US Marine Corps without whom this project would not have been possible.

The views expressed in this document are those of the authors and do not reflect the official policy or position of the Department of Defense or the U.S. Government.

References

[1] Sadagic, A., Kölsch, M., Welch, G., Basu, C., Darken, C., Wachs, J.P., Fuchs, H., Towles, H., Rowe, N., Frahm, J.M., Guan, L., Kumar, R., Cheng, H.: Smart Instrumented Training Ranges: Bringing Automated System Solutions to Support Critical Domain Needs. Journal of Defense Modeling and Simulation (2012) (accepted for publication)

[2] Livingston, M.A., Rosenblum, L.J., Brown, D.G., Schmidt, G.S., Julier, S.J., Baillot, Y., Swan II, J.E., Ai, Z., Maassel, P.: Military Applications of Augmented Reality. Springer, Heidelberg (2011)

[3] Muller, P., Schmorrow, D., Buscemi, T.: The Infantry Immersion Trainer – Today's Holodeck (September 2008)

[4] Blake, J.T.: Products and Services Catalog (2010), `http://www.peostri.army.mil/`

[5] Wachs, J.P., Kölsch, M., Goshorn, D.: Human Posture Detection for Intelligent Vehicles. Journal of Real-Time Image Processing (2010)

[6] Wachs, J.P., Goshorn, D., Kölsch, M.: Recognizing human postures and poses in monocular still images. In: Proc. Intl. Conf. on Image Processing, Computer Vision, and Signal Processing, IPCV 2009 (2009)

[7] Rowe, N., Houde, J., Kölsch, M., Darken, C., Heine, E., Sadagic, A., Basu, A., Han, F.: Automated assessment of physical-motion tasks for military integrative training. In: Second International Conference on Computer Supported Education, Valencia, Spain (2010)