Optically-Based Small Arms Targeting

John R. "Buck" Surdu, Joshua Crow, Chris Ferrer, and Rick Noriega
Cole Engineering Services, Inc.
Orlando, FL
Buck.Surdu@cesicorp.com, Joshua.Crow@cesicorp.com, Christopher.Ferrer@cesicorp.com Rick.Noriega@cesicorp.com

Pat Garrity
US Army Research Laboratory
Human Research and Engineering Directorate
Advanced Training and Simulation Division
Orlando, FL
patrick.j.garrity4.civ@mail.mil

ABSTRACT

Current tracking technologies used to estimate Soldier and weapon location and orientation are insufficient to support long-range direct fire engagements in live force-on-force exercises and mixed and augmented reality (AR) training applications. Tracking solutions developed for augmented reality only support engagements at ranges just over 50 meters, but Soldiers and Marines are trained to fire at targets at 375 meters or more. Laser-based systems suffer from a number of drawbacks: (1) they do not require shooters to lead moving targets, (2) they do not require soldiers to properly adjust weapon elevation based on target range, (3) they cannot represent grenade launchers, and (4) they are blocked by foliage. Each of these drawbacks results in negative training. We present a novel approach to determining hits and misses with small arms fire that enable longer-range engagements without the use of lasers for live, force-on-force training applications. The basic intuition to this novel approach is that instead of using high-resolution accuracy of soldier and weapon location (to within four centimeters) and orientation (to within 100 micro radians or better), the starting point for a calculation is the reported locations of the shooter and target and the shooter's sight picture when the round is fired. No harnesses, markers, reflectors, or indicia are used in this approach. With a sight picture as the starting point of the computation we know the aim point with respect to the target precisely at the time the trigger is pulled. Was the shooter leading the target properly? Did the shooter aim above center of mass of the target because he was firing at long range? This information is used to compute whether a shooter should accurately get credit for a hit or miss at long range. The system, Optically-Based Small Arms Targeting (OBSAT) successfully demonstrated accurate recording of hits and misses at distances of up to 375 meters against stationary, moving, and partially occluded targets. Results of this work are presented as well as areas for future research required to make this technology fieldable.

ABOUT THE AUTHORS

Dr. John R. "Buck" Surdu retired from the US Army after almost 29 years of service. After working in the research directorate of the National Security Agency, he began work for Cole Engineering as a senior scientist. He was a successful project manager for the Army's OneSAF program as well as a PM at DARPA for the Deep Green program, among other projects. He is a Ph.D. in computer science with background and refereed publications in artificial intelligence, modeling and simulation, and software engineering. He is a charter member of the Modeling and Simulation Professional Certification program. His current research interests are diverse and include developing technologies to improve live, force-on-force training.

Joshua Crow is a Senior Software Engineer with 15 years of full-lifecycle software development experience, much of it spent developing data conditioning pipelines across a wide range of consumer and military applications. He holds a B.Sc. in Computer Science, an M.S. in Informatics, and an M.B.A.

Chris Ferrer is a Software Engineer for Cole Engineering Services, Inc. (CESI). He has a B.Sc. in Computer Engineering from the University of Central Florida. Mr. Ferrer has nine years of experience in the military simulation and training industry, including work on programs such as One Semi-Automated Forces (OneSAF), Live Virtual Constructive-Integrating Architecture (LVC-IA) and Combined Arms Command & Control Trainer Upgrade System (CACCTUS).

Rick Noriega is a Senior Software Engineer with 20 years of full-lifecycle software development experience, much of it spent developing software for communications and mobile devices including the AN/CYZ-10, AN/PRC-117F, KOV-21, KIK-20, KIV-19M, as well as, commercial Windows CE and android devices. He holds a B.Sc. in Electrical Engineering and an M.S. in Electrical Engineering.

Pat Garrity is a Chief Engineer at the Army Research Laboratory (ARL) - Human Research and Engineering Directorate (HRED) Advanced Training and Simulation Division (ATSD). He currently works in Dismounted Soldier Simulation Technologies conducting research and development in the area of dismounted soldier training and simulation where he was the Army's Science and Technology Manager for the Embedded Training for Dismounted Soldiers program. His current interests include Human-In-The-Loop (HITL) networked simulators, virtual and augmented reality, and immersive dismounted training applications. He earned his B.S. in Computer Engineering from the University of South Florida in 1985 and his M.S. in Simulation Systems from the University of Central Florida in 1994.

Optically Based Small Arms Targeting

John R. "Buck" Surdu, Joshua Crow, Chris Ferrer, and Rick Noriega
Cole Engineering Services, Inc.
Orlando, FL
Buck.Surdu@cesicorp.com, Joshua.Crow@cesicorp.com, Christopher.Ferrer@cesicorp.com Rick.Noriega@cesicorp.com

Pat Garrity
US Army Research Laboratory
Human Research and Engineering Directorate
Advanced Training and Simulation Division
Orlando, FL
patrick.j.garrity4.civ@mail.mil

INTRODUCTION

Current tracking technologies used to estimate Soldier and weapon location and orientation are insufficient to support long-range direct fire engagements in live force-on-force exercises and mixed and augmented reality (AR) training applications. Tracking solutions developed for augmented reality only support engagements at ranges just over 50 meters, but Soldiers and Marines are trained to fire at targets at 375 meters or more. Laser-based systems suffer from a number of drawbacks: (1) they do not require shooters to lead moving targets, (2) they do not require soldiers to properly adjust weapon elevation based on target range, (3) they cannot represent grenade launchers, and (4) they are blocked by foliage. Each of these drawbacks results in negative training. We present a novel approach to determining hits and misses with small arms fire that enable longer-range engagements without the use of lasers for live, force-on-force training applications. The basic intuition to this novel approach is that instead of using high-resolution accuracy of soldier and weapon location (to within four centimeters) and orientation (to within 100 micro radians or better), the starting point for a calculation is the reported locations of the shooter and target and the shooter's sight picture when the round is fired. No harnesses, markers, reflectors, or indicia are used in this approach. With a sight picture as the starting point of the computation we know the aim point with respect to the target precisely at the time the trigger is pulled. Was the shooter leading the target properly? Did the shooter aim above center of mass of the target because he was firing at long range? This information is used to compute whether a shooter should accurately get credit for a hit or miss at long range. The system, Optically-Based Small Arms Targeting (OBSAT) successfully demonstrated accurate recording of hits and misses at distances of up to 375 meters against stationary, moving, and partially occluded targets.

The desire of target-weapon pairing or firing electronic bullets has been recognized by PEO STRI (PM TRADE, 2012). The challenge has been that the sensor accuracy needed to accurately engage targets at ranges out to 375 meters and beyond required sensor accuracy not affordably achievable today. For instance, to accurately predict the trajectory of a rifle bullet to get a correct assessment of a hit or miss requires a location accuracy of approximately four centimeters and orientation accuracy of approximately 100 micro-radians. Soldier-worn GPS is generally considered to have an accuracy of approximately a meter or 1.5 meters, and current, affordable orientation sensors are accurate to approximately a degree and a half, or 26180 micro radians. The question for this research effort then, was if sensors of desired accuracy do not exist, what can be done with the sensors that are available?

The purpose of this research was to investigate a novel approach to compensate for poor sensor fidelity. Digital weapon sights, such as the next generation Thermal Weapon Sight (TWS) (Anonymous, 2016) offer an opportunity to take advantage of information not previously available, specifically the soldier's sight picture when the trigger is pulled. Optically-Based Small Arms Targeting (OBSAT) uses the sight picture as the starting point to determine soldier's aim point with respect to the target. Was the shooter leading the target properly? Did the shooter aim above center of mass of the target because he was firing at long range? This information is incorporated with AMSAA-validated direct fire models used in OneSAF to compute whether a shooter should accurately get credit for a hit or miss at long range.

Additional maturation is needed before this technology will replace laser engagement systems for force-on-force training or testing. Once sufficiently mature, Marines and Soldiers will train with their go-to-war equipment with little or no additional appended equipment. It is important to note that this approach requires no laser transmitters, no “halos,” and no sensor harnesses or vests. Instead, the design uses technology that has been fielded or will soon be fielded, such as a Marine- or Soldier-worn smart phone and a digital thermal weapon sight. In addition to reducing the encumbrance and distraction of appended training equipment, this approach also addresses other deficiencies of laser engagement systems, including the inability to fire through foliage, and negative training, such as not reinforcing basic rifle marksmanship skills like leading a moving target or adjusting the elevation of the weapon based on target range. This research points to technology that, once mature, can be cheaper and better than current laser-based systems.

THE OBSAT APPROACH

The design for a laser-less system to test the theory that accurate engagements are possible at realistic ranges (out to 375 meters) involves a cloud-based server architecture and a minimalist soldier-worn subsystem. It captures the sight picture seen by the shooting soldier when the trigger is pulled to compensate for poor sensor accuracy of position and orientation. The reported position and orientation along with the captured sight picture are then used to accurately assess a target hit or miss to ranges at least as great as 375 meters. Throughout development the team’s philosophy has been to use as much fielded or soon-to-be fielded equipment as possible. In this way the OBSAT technology points to a fieldable live, force-on-force training or testing capability that requires minimal or no appended hardware.

Server-Based Architecture

In Figure 1 the shooter is indicated as (1), and the target is (2); however, in this approach the roles may be reversed. Both the shooter (1) and target (2) periodically report (3) their estimated location, orientation, and speed to a wireless communication relay (4). These location updates are transmitted (5) to a Remote Sever (6), where they are

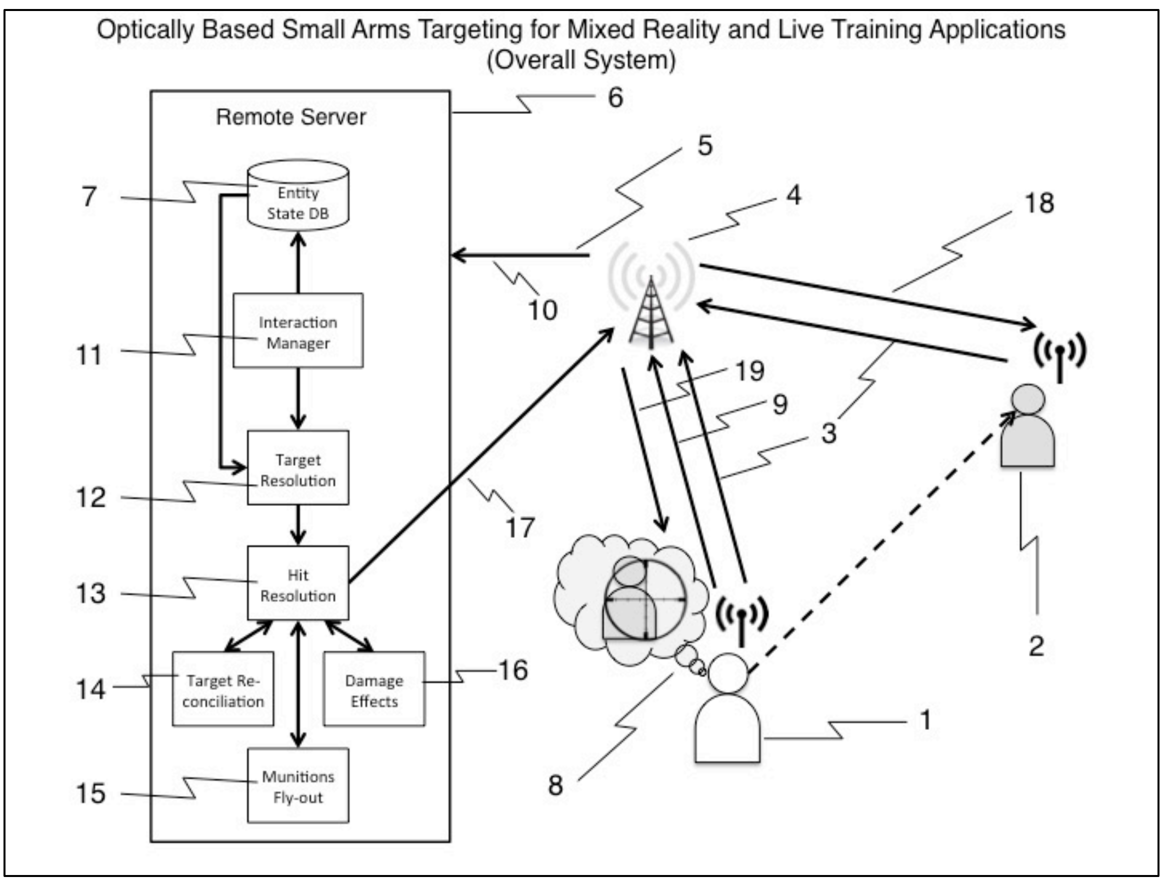

Figure 1: Overall System Architecture

stored in the Entity State Database (7) for later use. The shooter (1) aims his training rifle at his target (2) and pulls the trigger. The shooter's location, weapon orientation, and sight image (8) are transmitted (9) to the wireless relay. The sight image is a digital representation of the shooter's view through his weapon's sight when he pulls the trigger. The location and orientation of the shooter (1) and his sight image is transmitted (10) to the Remote Server (6) and the Interaction Manager (11). The Interaction Manager queries the Target Resolution module (12), which produce a list of possible targets from the Entity State Database based on the weapon location, orientation, and known orientation error. This list of possible targets is provided to the Hit Resolution module (13), which performs the following actions:

- It runs the computer vision algorithm to find targets in the sight image. This step includes processing the sight image to locate targets and determining the relationship between the aim point and the targets based on the sight image. For instance, did the shooter aim high, low, left, or right of center of mass of the target?
- It calls the Target Reconciliation module (14), which reconciles results from the computer vision computation with information from the Entity State Database. This step identifies which targets from the Target Resolution module (14) correspond to targets identified by the computer vision algorithm.
- It queries the Munitions Fly-out module for the flight time of the projectile and adjustments to the trajectory of the round. These adjustments can be based on range (e.g., drop of the round over distance), atmospheric effects, weather, wind, interactions with the terrain, and other factors as required to accurately predict the trajectory of the round.
- It computes whether the trajectory of the round intersects the target determined by the Target Reconciliation module based on the adjusted trajectory, time of flight, and velocity of the target.
- If the round struck the target, the Hit Resolution module calls the Damage Effects module (16). This module computes the damage to the target based on the weapons characteristics, such as whether the target was killed, sustained a minor wound or major wound, the location of the wound, etc.

A near miss is reported (17) through the wireless relay (4) and retransmitted (18 and 19) to the target (2) and the shooter (1), respectively, who are informed of the near-miss result. A hit result is reported (17) through the wireless relay (4) and retransmitted (18 and 19) to the target (2) and the shooter (1), respectively. The shooter is notified of

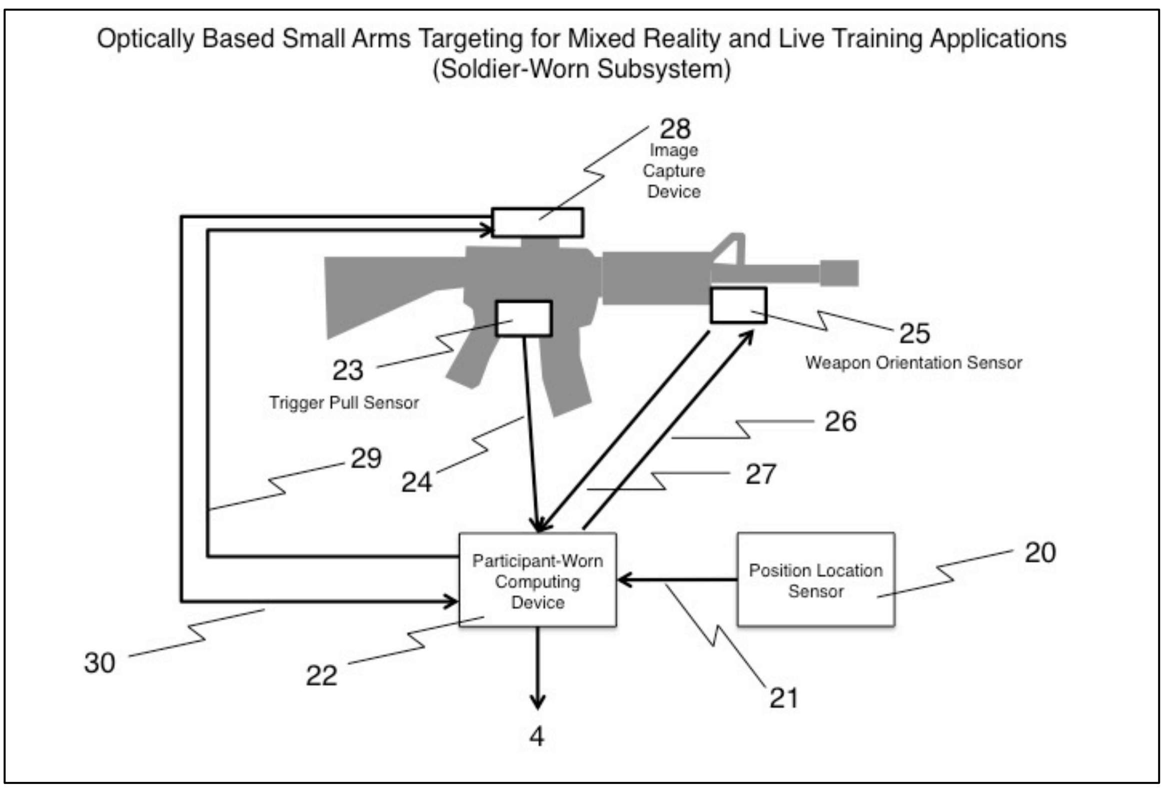

Figure 2: Soldier-Worn Subsystem

a hit, and the target is notified of that he was hit, with what weapon or round he was hit, and the severity of the damage.

OBSAT uses a central server to perform hit adjudication and maintain global simulation state. A central server allows the system to use more complex algorithms than are feasible with the client software on a mobile device with limited processing and battery resources. A central server also requires only O(n) network links to provide communication between connected systems. This client-server approach has a number of benefits, including easier software maintenance and improvement of “fair fight” issues. In addition, it is efficient with regard to bandwidth. For example, a peer-to-peer type architecture, in which the processing occurs on soldier-worn systems which must communicate with each other, requires $O(n^2)$. In a bandwidth- and latency-constrained environment, minimizing the number of active connections is a beneficial trade-off.

Communication between soldier-worn subsystem and the central server is conducted over a 4G LTE cellular connection. This channel was selected due to convenience and durability. The off-the-shelf mobile hardware employed by OBSAT supports 4G LTE as a baseline capability. The communication hub – a cellular tower – scales reliably with geographic distance. Alternative communication channels such as 802.11 Wi-Fi would require additional hardware and suffer from significant performance degradation at long ranges.

The 4G LTE technology suffers from several drawbacks, however. Performance degrades as the cellular towers become congested. This limitation is mitigated by the Army’s move towards dedicated cellular infrastructure at many training sites. LTE communications also involve higher latency than Wi-Fi technologies. This latency is only observed when data transmission begins, and can be mitigated by keeping the mobile devices’ cellular antennas operational. Such mitigation increases power consumption for the mobile devices, but an available external battery pack, such as that being developed by PEO Soldier (PEO Soldier, 2012, and McCaney, 2014) provides sufficient power for an extended training scenario.

OBSAT uses an off the shelf exercise control (EXCON) tool developed by Human Systems Integration (HSI) that was modified to support the OBSAT communications protocol and record events. This capability proved invaluable during development and testing; the per-shot AAR capability is very visual and helped developers debug OBSAT. It also enabled developers to demonstrate why shots were properly hits or misses during demonstration events.

Soldier-Worn Subsystem

Figure 2 displays the configuration of the soldier-worn subsystem. The sub-system operates as described below. The Position Location Sensor (20) provides periodic updates of the soldier’s location, orientation, and speed to the Soldier-Worn Computing Device (22). The Soldier-Worn Computing Device transmits these updates to the wireless relay (4) as described previously.

When the soldier pulls the trigger on his training rifle, the Trigger Pull Sensor (23) sends a message (24) to the Soldier-Worn Computing Device. The Soldier-Worn Computing Device sends trigger-pull events (27) and (29) to the Weapon Orientation Sensor (25) and the Image Capture Device (28). The Weapon Orientation sensor returns the weapon orientation (27) to the Soldier-Worn Computing Device. Similarly, the Image Capture Device (28) provides the sight image as seen by the shooter (30) to the Soldier-Worn Computing Device. The image capture device may provide:

- A mix of visible spectrum, non-visible spectrum, and multispectral images.
- A video image or a series of still images.
- Images from a single viewpoint or multiple viewpoints.

The Soldier-Worn Computing Device (22) sends the location and orientation of the weapon as well as the sight images to the Wireless Relay (4) as described previously.

The OBSAT proof-of-principle assumed a single image capture device (i.e., scope) mounted Picatinny rail, which has been zeroed to the weapon. This research employed both a daylight and infrared scope. The soldier-worn computing device for this proof of principle was a Samsung Galaxy Note One smart phone. This device was chosen, because it is probably going to be part of the soldier-worn ensemble for go-to-war equipment being developed by PEO Soldier (McCaney, 2014). For this proof-of-principle, an Airsoft M-4 surrogate was used. The primary image capture device is the ATN X-Sight HD Scope for visual spectrum. A thermal imaging clip-on was added in front of the visual spectrum scope to enable imaging in the IR spectrum. Currently the soldier-worn subsystem operates in either the visual spectrum or infrared spectrum. Eventually we expect to leverage image data from both spectrums to provide a more complete picture of the soldier's environment. This will allow the strengths of image processing in one spectrum to shore up the weaknesses of image processing in the other spectrum. The final integrated list of target features would be more complete (fewer false negatives), better culled (fewer false positives), and offers the best opportunity to define an accurate target silhouette.

The solder-worn computing device contains a GPS device that provides position updates at a rate of 1 per second (1 Hertz, or 1 Hz). For our stationary demonstration this was sufficient. At a 1 Hz update rate there was too much error in the calculated velocity for the system to resolve hits properly when the soldier is moving evasively. An external GPS sensor was found that connects to the soldier-worn computing device over a Bluetooth link and provides a 10 Hz update rate to the native Android location services. The device used was a Garmin GLO Portable Bluetooth GPS and GLONASS Receiver.

Orientation Sensor

The system uses an Inertial Labs OS3D sensor to provide the orientation of the weapon in three dimensions. The OS3D device connects to the soldier-worn computing device using a serial over USB protocol. The OS3D Service retrieves the data frames from the orientation sensor at a rate of 10 Hz, decodes them, calculates the weapon orientation, and broadcasts the data to the system.

Computation of Hit and Miss

Target Resolution

Target resolution is the first step in the hit detection pipeline. The target resolution algorithm uses shooter position, target positions, and scope field of view to determine which targets, if any, may be present in the sight picture. The determination is based on whether the target is alive or dead and whether the target's reported position lies within a sensor error cone. In the OBSAT approach location and orientation accuracy estimates only needs to be good enough to select the right target from the cloud-based database. OBSAT "flies out" an electronic bullet to determine whether the target was struck. The ballistic flight path can be as detailed and computationally expensive as required for specific applications. The flight path of the bullet is not straight, but elliptical. This means that even if the sight picture is on the target, the round might miss. Also, if the target is moving, and the shooter did not properly lead the target, he would not get credit for a hit. In this method, the orientation of the shooter's head is irrelevant; only the orientation of the weapon matters. If no living target candidates are within the error cone, the system records the shot as a miss due to the lack of targets and no further processing is done. This "early out" reduces both processing cost and latency in such cases.

Image Processing

The image processing module is responsible for identifying human targets within the sight picture. This is the most complex step in the hit detection pipeline, and multiple algorithms were evaluated for precision and recall. The image processor implementations leverage the OpenCV (Itseez.com, 2016) and AForge libraries (AForge.net, 2016), as well as the Emgu CLR (Emgu.com, 2016) wrapper for OpenCV. Image processors fall into two broad categories: non-temporal and temporal. Non-temporal processors operate on a single candidate image. Temporal processors operate on a candidate image, but are provided with a reference frame captured at a previous time to exploit temporal changes in position The OBSAT team evaluated several different algorithms in multiple spectrums.

The **Sobel** filter is a basic edge detection algorithm that uses discreet differentiation to compute an approximate gradient magnitude (Anonymous Sobel, 2016). The Sobel filter image processor performed well at detecting human silhouette edges, but produced poor results in areas with leaves or similar high-noise edges. Based on these results, the OBSAT team ruled out the use of Sobel filtering for target detection.

The **Canny** edge detection algorithm is a more complex detector than Sobel (Anonymous Canny, 2016). It also uses discreet differentiation, but includes steps to suppress non-maximum values and an iterative hysteresis function that includes borderline pixels connected to a valid edge. Using the Canny algorithm is challenging due to the difficulty in selecting the appropriate threshold and hysteresis values. The algorithm was of limited use for evaluating visible or IR images, but was a useful tool for silhouette matching against visible spectrum images. Its use in that algorithm is described below.

OBSAT evaluated the **OpenCV cascade classifier**, which uses Haar-like features for face detection (OpenCV, 2016). The classifiers are trained for faces that are 24x24 pixels or larger. The sight pictures are of insufficient resolution to meet this requirement at many valid target ranges. As a result, face detection was excluded from further algorithmic development. Should a higher-resolution sight picture become available, this classifier should be reevaluated.

Histogram of Oriented Gradient (HOG) **Pedestrian Detection** is one of the most successful pedestrian detection algorithms (Tsai, 2010). The algorithm computes the discrete differential gradients at each point in the image, quantizes the gradients into a histogram, and uses the histogram bin sizes across the scan window as inputs to a support vector machine(SVM). HOG showed very good results in the visible spectrum for non-occluded and horizontally occluded targets. Performance for vertically occluded and IR targets was suboptimal. Stochastic optimization via Monte Carlo simulation was applied to find the best parametric values for the SVM threshold and grouping size. The OBSAT team also used training images extracted from the Caltech pedestrian dataset to train new HOG classifiers for small targets and occluded targets. The trained classifiers showed decreased performance compared to the default due to the mismatch between the training data (pedestrians in an urban setting) and test conditions (pedestrians in a wooded environment). Accuracy could be improved with a larger number of domain-appropriate training images.

The **background subtraction** algorithm uses the Speeded Up Robust Features (SURF) feature detection algorithm to correlate feature points between the candidate image and a reference frame (Anonymous SURF, 2016). Reference frame pixels are transformed into candidate image space and then subtracted. This produces hotspots in the candidate image where an object is moving. Performance of background subtraction varied depending on the quality of the candidate and reference images. When both were of high quality and closely correlated, the background subtraction algorithm was effective at locating moving targets. Often, however, the combination of motion blur and image noise would leave apparent motion artifacts throughout the subtracted image.

Gaussian-smoothed high-pass filtering (GSHPF) is a custom algorithm developed by the OBSAT team to process IR images. This algorithm is effective at finding human targets in a scene and generating an approximate silhouette for the target. The first step in the algorithm isolates hot “islands” within the image and generates features from them. The second step then uses Gaussian smoothing to link together nearby islands. Different Gaussian kernel sizes were tested, with a 5x3 kernel yielding the best results. The surviving hot “islands” are culled through convolution with a silhouette kernel to detect regions that have the characteristics of targets.

Overall performance of GSHPF was very good; although, the IR sensor itself become unreliable at higher ambient temperatures. When the background temperature approached human body temperature, which occurs during the summer in Florida by 10:00 AM, the IR sensor no longer captures any contrast between human targets and non-human background objects. This causes significant degradation of the GSHPF algorithm. In fact, it became difficult for the human user to see targets through the sight in order to fire on them. The most accurate results were obtained using GSHPF for the IR spectrum and the custom visible spectrum processing for visible spectrum processing. These algorithms were used for the final demonstration.

The team built a custom algorithm that combines the GSHPF with background subtraction to improve IR spectrum performance against moving targets. Rapidly moving targets appear very dim in IR sight pictures due to the longer response time of the commercial-grade IR sensor. Background subtraction allows the algorithm to find these dim moving targets while GSHPF effectively locates stationary and occluded targets. The success of this algorithm was closely related to the success of the background subtraction algorithm. When background subtraction is able to correctly correlate the candidate and reference image, the aggregate algorithm performs better than GSHPF alone. The success rate of background subtraction was below 60%, however, causing the aggregate algorithm to perform worse than pure GSHPF on average.

The team developed a custom aggregate visible spectrum algorithm that combines the HOG pedestrian detector with a modified version of Canny edge detection for silhouette fitting. This algorithm produced the best overall hit detection results for visible spectrum images.

Target Reconciliation

The target reconciliation interface is responsible for mapping image processing candidate regions with target resolution candidate targets. Target reconciliation also performs silhouette fitting against mapped candidates when possible to do so. Several variant target reconciliation algorithms were tested. The final algorithms differ slightly for IR and visible spectrum images. Due to limitations in weapon orientation sensor precision, the minimum distance needed to reliably reconcile two targets is approximately one degree of arc. This translates to 5 meters of separation when targets are 150 meters distant and 12 meters at 375 meters, as shown in Table 9. Future work will use the sight picture to assist in target reconciliation, greatly reducing these distances. The IR and visible spectrum target reconciliation algorithms differ in how the silhouette fitting step is performed due to differences in the candidate region silhouette mask data available from the corresponding image processing algorithms.

Hit Resolution

Hit resolution is the final step in hit detection. It takes the reconciled targets and determines if a fired munition would actually hit any of the selected candidates. This includes accounting for munition rise or drop at the target's range, correcting for the target's velocity, and determining the effect of a hit. The hit resolution module uses a munitions fly-out interface to query the time of flight and distance above or below weapon sights for the fired munition at a specified time. At present, only 5.56mm rounds are supported. The data for the munition fly-out is taken from existing, authoritative sources. The 5.56mm round flight time data was taken from OneSAF, while rise and drop data were collected from military documentation. The hit resolution code calculates the target position when the bullet reaches that distance using the flight time retrieved from the munition fly-out interface. The position is then adjusted to account for rise or fall against the weapon sights. This data is then used to calculate the image-space location to test for a hit or a miss. If a silhouette mask is present, the value of the mask is compared to upper and lower threshold values. If above the upper threshold, the shot is resolved as a hit. If between upper and lower thresholds, the shot is a near miss. If below the lower threshold, the shot is a miss. For candidate targets without a silhouette, the position is tested against the bounding box. If the point is inside the box, the round hits; if outside, it misses. Due to time constraints, OBSAT uses a very basic damage effects model. Any hit is considered lethal.

EXPERIMENTAL RESULTS

The team provided three demonstration events for the Government customer during development of the proof of principle system. The purpose of the first demonstration was to show that the overall system architecture was feasible. Two targets stood downrange from the shooter. The shooter fired multiple shots at the targets, attempting alternately to hit and miss each target. This process was repeated with targets at 150 meters, 250 meters, and 360 meters from the shooter. System accuracy at all ranges was good, with accuracy at 150 and 250 meters being comparable. Accuracy at 360 meters was slightly lower. The most common failure mode observed was sensor inaccuracy due to changes in the electromagnetic environment. The most effective algorithm for stationary targets depends on ambient temperature. When ambient temperature is different from human body temperature, the IR algorithms performed better than those for the visible spectrum. When ambient temperature is very close to human body temperature, the visible spectrum algorithms perform better.

The second demonstration event focused on improving issues found during the first event and also expanding the demonstration to include moving targets. The target sprinted back and forth across a narrow range in front of the shooter. The shooter fired the weapon multiple times, attempting to both hit and miss the target. The shooter led the target appropriately to compensate for fly-out time. This process was repeated with the target at 150 meters, 250 meters, and 360 meters from the shooter. System accuracy at all ranges was comparable to that with stationary targets. Similar issues with sensor accuracy were observed. The sensor margin or error also increased when the weapon was in motion, leading to higher aggregate error. The visible spectrum algorithm was the most effective for moving targets. Slow IR sensor response with the commercial, non-military-grade IR sensor caused rapidly moving targets to appear very dim, making it difficult for the computer vision algorithms to locate the target in the sight picture. A more responsive, military-grade, IR sensor may mitigate this limitation.

The third demonstration event validated previous findings and expanded the system to partially occluded targets behind foliage. The target moved to positions that occluded him horizontally (aligned with the transverse anatomical plane) and vertically (aligned with the mid-sagittal anatomical plane). The shooter fired the weapon multiple times, attempting to both hit and miss the target. This process was repeated with the target at 150 meters, 250 meters, and 360 meters from the shooter. The target remained stationary for this demonstration. System accuracy at all ranges was lower than that of non-occluded targets. The OBSAT proof-of-principle system was able to detect partially occluded targets at all ranges. This indicated that with additional maturation, the technology would be able to fire through foliage that would block a laser, but not a bullet. Improved sensors, larger training sets, and improved computer vision will address the shortcomings of the system against partially-occluded targets. As with stationary targets, the IR spectrum algorithm performed best for occluded targets. It was more accurate at identifying the target and provided better silhouette fitting for located targets. Visible spectrum performance was adequate, but below the accuracy of the IR algorithm. In particular, while in the visible spectrum OBSAT performed well against horizontally occluded targets, in the IR spectrum OBSAT was able to detect both horizontally and vertically occluded targets. Occluded performance can be improved further through additional work on the computer vision portion of the system. Developing an occluded target data set will allow the training of custom classifiers to handle this case. The introduction of deep neural network (DNN) classifiers may further improve accuracy for occluded cases.

An experiment was performed in which sixty shots were taken against a mix of stationary, moving, and occluded targets. Eleven samples were discarded either due to their use for system calibration or due to configuration errors. The forty-nine remaining data points were analyzed (n = 49). The aggregate results are reported in Table 1: Analysis Results. The test results show a positive predictive value (PPV) approaching 100% and a negative predictive value (NPV) of 55%. The system displayed an overall accuracy of 75% during the experiment, with all failures resulting from Type II errors. The causes of observed Type II errors are silhouette misfit (n = 5), sensor drift (n = 4), and failure of the computer vision algorithm (n = 3). The high incidence of silhouette misfits is partially a result of budget constraints preventing the full tuning of the algorithm.

Table 1: Analysis Results

	Ground Truth Hit	Ground Truth Miss
OBSAT Reported Hit	22	0
OBSAT Reported Miss	12	15

Bandwidth Estimates

During development, bandwidth was questioned several times as a limiting factor of this approach. Latency and bandwidth are, of course, related. Given the size of the packets and the bandwidth of a standard LTE network, the team determined the number of soldiers who could train with OBSAT simultaneously. This back-of-the-envelope computation assumes that some percentage of the users would fire their weapon within the exact same second and that the rest merely would be sending 10 Hz position updates. Given those assumptions, if 5% of the soldiers were firing their weapons in the same second, that OBSAT could support more than a battalion of soldiers at the same time, as shown in Table 2.

Table 2: Required Bandwidth for OBSAT 1

	%	Platoon	Company	Battalion
Total number of soldiers		45	140	600
% Firing within the same second	5%	3	7	30
% Sending 10 Hz position updates	95%	42	133	579
Total bandwidth required		3,711,000	9,271,500	39,735,000

If one assumes that 20% of the soldiers would fire their weapons in the same second, which is unreasonably high, fewer than two companies can participate in the same exercise, as shown in Table 3, unless the network is upgraded to LTE Advanced. No attempt was made to optimize OBSAT bandwidth requirements, but optimizations, such as sending fewer than five captured images per trigger pull, compressing messages, and other techniques will improve scalability.

Table 3: Required Bandwidth for OBSAT 2

	%	Platoon	Company	Battalion
Total number of soldiers		45	140	600
% Firing within the same second	20%	9	28	120
% Sending 10 Hz position updates	80%	36	112	480
Total bandwidth required		9,558,000	29,736,000	127,440,000

FUTURE WORK

Throughout development, the team identified a number of research efforts that would improve the performance of OBSAT. The areas that will improve performance most dramatically occur in areas related to computer vision. This is an active field of research in the community, and the OBSAT team is prepared to leverage advances as they are known. While OBSAT displays performance consistent with the state of the art in computer vision, there are opportunities to further improve the accuracy of the system.

Enhance Image Capture

One option is to enhance the image capture process. OBSAT currently operates on 848x480 images captured from an H.264 stream and converted to PNG. The IR capture includes an additional conversion from a smaller 640x480 IR sensor to an intermediate 800x600 display resolution in front of the visible scope. Improving the resolution of the captured weapon sight pictures would improve overall system performance and enable OBSAT to apply additional algorithms such as cascading Harr-like feature classifiers for face detection. This requires the identification and integration of higher resolution visible and IR sensors into the OBSAT system. Optimally, image resolution should provide approximately 80 pixels across the shoulders of the target at 375 meters without the use of digital zoom.

The temporal image processing algorithms would be further improved by fitting the system with a sensor that detects an aim posture and begins capturing reference frames prior to the trigger being pulled. When the weapon is raised to a firing position and is held steady (while the soldier is aiming) the system can capture some number of reference frames. This would greatly improve the reliability of currently borderline algorithms such as background subtraction.

Image capture could be further enhanced by supporting multi-spectral imaging. Currently, the system can capture IR or visible spectrum images, but not both simultaneously. Enabling concurrent capture of multiple spectra would enable fusion algorithms that provide the benefits of each individual spectrum while mitigating the limitations of each technique. While this will result in increased bandwidth requirements, which themselves can be mitigated with other techniques, applying both the daylight and IR algorithms to the sight picture and combining the results should result in marked improvements in accuracy.

Improve Computer Vision Accuracy

The most significant limit on computer vision performance is the lack of training data. With appropriate training data, new computer vision classifiers could be trained using the existing algorithms. This would improve system performance against targets in a variety of poses. Classifiers could also be developed to detect weapons, enhancing accuracy in exercises where camouflage might otherwise hinder computer vision performance. Training data can be captured using live subjects, but technological advances now make it possible to generate realistic synthetic training data using DirectX 12-class rendering hardware. This would facilitate the development of training data across a wide variety of lighting conditions, posts, background environments, and degrees of occlusion.

Bandwidth and Latency Optimizations
While the client-server approach has a number of advantages, scalability must be considered. The team has identified a number of techniques that could be employed to improve the bandwidth requirements of OBSAT. These techniques and others should be explored to improve the scalability of the approach.

Refine Spatial Precision
Any improvements in position or orientation accuracy improve the overall OBSAT performance. The introduction of mobile ad hoc network (MANET) localization techniques may improve location accuracy. In general these techniques rely on time of arrival or direction of arrival techniques (Wang and Zekavat, 2006) as well as a fixed number of anchor points to accurately locate the various moving entities within the network. Additional research is needed on combining these techniques to compensate for each technique's weaknesses.

Damage Effects
OBSAT currently uses the simplest possible damage model – any hit is automatically a kill. This can be enhanced to provide more realistic damage effects based on hit location and munitions type. Further enhancements could be achieved through the integration of an open source physiology engine such as BioGears (ARA, 2016), allowing for a higher fidelity damage model and integrative battlefield triage and treatment training.

Additional Weapons Effects
An advantage of the OBSAT approach is that it has the potential to simulate not just rifle fire but additional weapons, such as machine guns, shotguns, and grenade launchers. It will likely also work for vehicle weapons. The team has begun to explore methods for representing burst and fully automatic fire, non-line of sight weapons, and other effects.

REFERENCES

AForge.net (2016). *AForge.net Framework*. Retrieved from http://www.aforgenet.com.

Anonymous. (2016). *AN/PAS-13*. Retrieved from https://en.wikipedia.org/wiki/AN/PAS-13.

Anonymous Canny. (2016). *Canny Edge Detector*. Retrieved from https://en.wikipedia.org/wiki/Canny_edge_detector.

Anonymous Sobel. (2016). *Sobel Operator*. Retrieved from https://en.wikipedia.org/wiki/Sobel_operator.

Anonymous SURF. (2016). *Speeded up Robust Features*. Retrieved from https://en.wikipedia.org/wiki/Speeded_up_robust_features.

Applied Research Associates, Inc. (ARA) (2016). BioGears. Retrieved from https://biogearsengine.com.

Emgu.com (2016). *Emgu CV Main Page*. Retrieved from http://www.emgu.com/wiki/index.php/Main_Page.

Itseez.com (2016). *Open Source Computer Vision*. Retrieved from http://opencv.org.

McCaney, K. (2014). Army's Move to Samsung Reflects a Flexible Mobile Strategy. Retrieved from https://defensesystems.com/articles/2014/02/24/army-nett-warrior-samsung-galacy-note-ii.aspx.

OpenCV. (2014). *Open CV API Reference*. Retrieved from http://docs.opencv.org/2.4/modules/objdetect/doc/cascade_classification.html.

Program Manager Training Devices (PM TRADE) (2012). *Army-Tactical Engagement Simulation (A-TESS) Industry Day Briefing*. Retrieved from https://www.lt2portal.mil/FG/FileGatekeeper.aspx?file=LT2_L0/Briefings/b5fde86a-52df-456e-89b5-ad8c63674e7b/ATESS_Industry_Day_Brief.pdf.

Program Executive Officer Soldier (PEO Soldier) (2012). *Nett Warrior*. Retrieved from http://www.peosoldier.army.mil/docs/pmswar/Nett-Warrior-Poster-061512.pdf.

Tsai, Grace. (2010). *Histogram of Oriented Gradients*. Retrieved from http://web.eecs.umich.edu/~silvio/teaching/EECS598_2010/slides/09_28_Grace.pdf

Wang, Z., & Zekavat, S.A. (2006). Manet Localization via Multi-Node TOA-DOA Optimal Fusion. *Proceedings of the 2006 IEEE Military Communications (MILCOM) Conference*, 1-7.