ARL-TR-10022 ● Nov 2024

Human–Machine Integration (HMI) Approach for Small Unit Single Target Tracking in Urban Environment

by Andrew Tweedell, Paul Sabbagh, Sangjin Han, Joseph Conroy, Jinho Kim, Osben Toulson, Richard Diego, William Maslin, and Jonathan Touryan

NOTICES

Disclaimers

ARL-TR-10022 ● Nov 2024

Human–Machine Integration (HMI) Approach for Small Unit Single Target Tracking in Urban Environment

Andrew Tweedell, Paul Sabbagh, Joseph Conroy, and Jonathan Touryan
DEVCOM Army Research Laboratory

Sangjin Han
Booz Allen Hamilton

Jinho Kim
Iron-Triangle Defense

Osben Toulson, Richard Diego, and William Maslin
DCS Corporation

REPORT DOCUMENTATION PAGE

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED START DATE	3. DATES COVERED END DATE
November 2024	Technical Report	1 October 2023	10 September 2024

4. TITLE AND SUBTITLE
Human–Machine Integration (HMI) Approach for Small Unit Single Target Tracking in Urban Environment

5a. CONTRACT NUMBER	5b. GRANT NUMBER	5c. PROGRAM ELEMENT NUMBER

5d. PROJECT NUMBER	5e. TASK NUMBER	5f. WORK UNIT NUMBER
0602184A/CN2		

6. AUTHOR(S)
Andrew Tweedell, Paul Sabbagh, Sangjin Han, Joseph Conroy, Jinho Kim, Osben Toulson, Richard Diego, William Maslin, and Jonathan Touryan

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)	8. PERFORMING ORGANIZATION REPORT NUMBER
DEVCOM Army Research Laboratory ATTN: FCDD-RLA-FC Aberdeen Proving Ground, MD 21005	ARL-TR-10022

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
DISTRIBUTION STATEMENT A. Approved for public release: distribution unlimited.

13. SUPPLEMENTARY NOTES
ORCIDs: Joseph Conroy, 0000-0002-1236-6267; Jonathan Touryan, 0000-0001-7343-7869

14. ABSTRACT
Near future US Army Infantry squads will soon deploy with increasingly intelligent technologies including small arms, unmanned vehicles (UAVs and UGVs), integrated augmented-reality (AR) heads-up displays (HUDs), and other tactical sensors to provide small-unit lethality overmatch on the ground. Currently, these system or platform concepts are not well integrated with each other or with the squad. This effort seeks to establish technologies and methodologies for investigating how these systems should interact with each other and with Soldiers in operational settings. The purpose of this report is to share results demonstrating real-time, multisystem (HUD, weapon, UAS), cross-platform integrated data capturing capabilities during military operations on urbanized terrain, featuring a single target acquisition and handoff exercise between squad members. Two sensor-ed rifle systems, two HoloLens 2 (HL2) AR wearable glasses, and one BlueUAS Teal UAS were simultaneously used by a small team (two riflemen, one UAS operator, and one squad leader [SL]) to identify and track a single moving human target as the target navigated a complex environment. The UAS and rifle systems successfully identified and tracked the target while providing GPS coordinates across the data network. This localization data were successfully ingested by the HL2 systems and used to 1) display horizontal and aerial video to the SL to determine the target threat level, and 2) display target location and threat level to the riflemen relative to their spatial location. The culminating proof-of-concept demonstration successfully demonstrated the need for a Soldier-centric integration plan for deploying small-unit lethality technologies and provides a technological pathway for future human–machine integration research for dismounted operations.

15. SUBJECT TERMS
Humans in Complex Systems; Photonics, Electronics, and Quantum Sciences; human–machine interaction; HMI; wearables; unmanned aerial system; UAS; small arms; artificial intelligence; AI; target tracking

16. SECURITY CLASSIFICATION OF: a. REPORT	b. ABSTRACT	C. THIS PAGE	17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UU	30

19a. NAME OF RESPONSIBLE PERSON	19b. PHONE NUMBER *(Include area code)*
Andrew Tweedell	(817) 914-5168

STANDARD FORM 298 (REV. 5/2020)
Prescribed by ANSI Std. Z39.18

Contents

List of Figures

1. Introduction

The US Army is modernizing ground combat forces to include a myriad of miniature and tactical sensors and robotics on the battlefield for Multi-Domain Operations (MDO; DOA 2021). Infantry squads will soon deploy with increasingly intelligent small arms, unmanned aerial and ground vehicles (UAVs and UGVs), integrated augmented-reality (AR) heads-up displays (HUDs), and other tactical sensors to provide small-unit lethality overmatch on the ground. The motivation behind modernizing ground combat forces with this vast array of intelligent systems surrounding the Soldiers "with the goal of always having robots, not Soldiers, make first contact with the enemy" is a sentiment shared throughout the US Army Futures Command (AFC; Heckman et al. 2023).

Currently, these system or platform concepts are not well integrated with each other or with the squad. While considerable work is being done in the areas of platform development and wireless communications to share data between different platforms, less work is being done to optimize these platforms together for the squad's lethality and survivability capabilities. For example, a small UAS (e.g., Program Manager [PM] Soldier Lethality's Soldier Borne System), can provide real-time aerial video to the squad; however, it still requires a Soldier using a fly-by-wire interface to pilot, which may reduce situational awareness. If future HUDs, such as PM Soldier Lethality's Integrated Visual Augmentation System (IVAS), do not selectively visualize information to the operators, the platform risks distracting or cognitively overloading the operators. Future small arms will have camera-based fire control technologies with Aided Target Recognition (ATR) algorithms to highlight threats and automatically calculate ballistic solutions. Small-arms platforms like these are currently only deployed as an isolated system and only work on targets they see. In a federated fire control context where an algorithm is trying to allocate weapons to targets, the optimal weapon–threat pairing may not be within the field of view. While simplistic, the examples given above highlight how siloed development of these platforms may result in reduced effectiveness when they are all deployed to a squad in real-world operational environments. This is evident in small-unit lethality situations where resources are limited, time pressure is high, and multisystem coordination is critical. Thus, the human–machine integration (HMI) challenge remains in scaling-up information density, transforming information intelligently, and speeding up decision-making.

The Humans in Complex Systems Competency within the US Army Combat Capabilities Development Command (DEVCOM) Army Research Laboratory (ARL) proposes developing novel HMI approaches for integrating intelligent dismounted platforms into the squad for enhanced lethality and survivability. Real-time Soldier–system information collected from distributed sensors could provide individual, squad, and environmental data *in the context of the current fight* as well as provide that data to intelligent systems. The broader goal of this effort is to establish technologies and methodologies for investigating how these systems should interact with each other and with Soldiers in operational settings. The purpose of this current report, however, is to share experimental results demonstrating real-time, multisystem (HUD, weapon, UAS), cross-platform integrated data capturing capabilities during military operations on urbanized terrain (MOUT) features a single-target acquisition, and handoff exercise between squad members. Two custom instrumented rifle platforms, two HoloLens 2 (HL2) headsets, and one BlueUAS Teal Golden Eagle UAS were deployed for the scenario. This report contains the concept of operations and technical details surrounding the multisystem exchange of relevant information during a dynamic scenario to enhance target acquisition and tracking. This effort was performed as part of ARL's Intelligent Weapons and Concepts Technology project, which conducts applied HMI research for dismounted Soldier technologies. The rest of this report is divided into the following sections: 1) background on component technology, 2) methods for system configuration, 3) qualitative results and lessons learned, and 4) conclusion and discussion on future research directions.

2. Background

The primary roles of the various component technologies in this demonstration either include geolocation of the target of interest or intuitive visualization of target information to the operators in the scenario. All hardware used for this proof of concept is commercial off-the-shelf. Instrumented rifles and aerial robotics provide the means to localize targets from different vantage points on the battlefield, while AR HUDs allow the operator to visualize in a common spatial reference frame and interact with information collected from the other systems. The unified objective of the systems should be to enable operators in the field to access critical information outside their capabilities and make decisions about threats faster.

2.1 Rifles

The purpose of the instrumented rifles is to provide ground-level environmental data (GPS, electro-optical [EO], and ATR) as well as HMI data. The HMI data includes measures of Soldier behavior with the weapon. For example, small sensor modules such as inertial measurement units (IMUs), have been used to characterize weapon motion during shooting events (Antipov et al. 2022) and has previously been used by our group to distinguish single shooter engagement types, e.g., static aiming versus smooth tracking versus quick transitions (Lee et al. 2022). Collecting this data from multiple rifle systems across a squad may allow us to qualify emergent behaviors at higher orders (echelons); however, this requires further research. Furthermore, our group has implemented intelligent fusing of IMU motion data and ATR tracking to fill gaps in tracking caused by partial target occlusion.

Viewing the instrumented rifles as UAS-like platforms and including sensor suites such as GPS, magnetometers, and flight control can provide unique capabilities (Han and Sabbagh 2022). For this effort, GPS, ATR, and the magnetometers were leveraged to geolocalize targets via triangulation from two rifle systems. Thus, the custom instrumented rifles are an integrated sensor fusion system with ATR that includes the following capabilities:

- Collecting sensory data from sensor devices (i.e., IMU, GPS, magnetometer, imagery, and a flight controller), that are attached to a rigid body (e.g., the rifle body)
- Adhering to the framework where operator behavior is labeled by classification algorithms using shooting behavior (Han and Sabbagh 2022; Lee et al. 2022)
- Running ATR algorithms to detect an object and fuse IMU data to fill the gap between target detections that are caused by occlusions
- Aggregating data and displaying information to monitor the instrumented rifle information

2.2 Unmanned Aerial System

As small UAS proliferate in increasingly autonomous roles for a variety of applications, one particular re-occurring desire is for such systems to automatically detect, classify, and localize mission-relevant objects within an area of interest. Appropriate theoretical approaches geared toward monocular vision-based localization have been developed (Barber et al. 2006; Redding et al. 2006; Barber 2007). While a detailed derivation of vision-based localization is outside the purview of this report, a summary schematic of the coordinate system translation for vision-based localization utilized by this effort is shown in Fig. 1.

The purpose of the UAS is to provide aerial environmental data (EO and ATR) as well as geolocalization data of the target during movement across the network. Humans are an ideal target class—for both their relevance for many applications and their ubiquity within the academic research world. A popular classifier model, the YOLOv5 detection model, comes pretrained on the Microsoft Common Objects in Context (COCO) dataset (Lin et al. 2014) wherein "person" is the first class of object. Here, we enhance the detection range by leveraging a model that is fine-tuned using additional aerial viewpoints (Maxey et al. 2023).

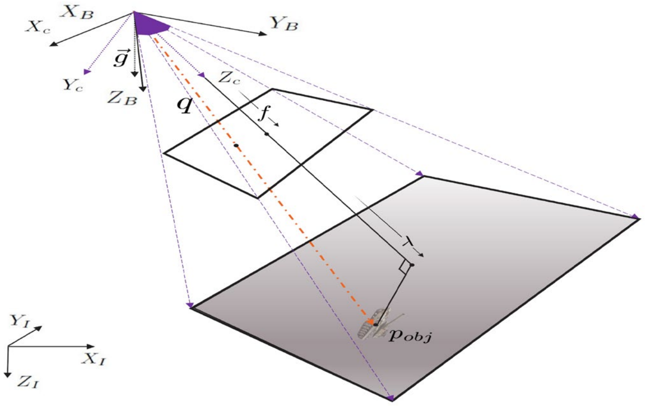

Fig. 1 Visualization of coordinate system for vision-based localization from UAS. *q* is camera orientation quaternions aligned with the projection of the point p_{obj} of the object onto the image plane. *f* is camera focal length; *g* is the gravity vector; λ is a nondimensional depth scale factor; *X, Y, Z* represent the coordinate axes for camera (*C*), UAS body (*B*), and inertial systems (*I*).

2.3 AR HUD

On a battlefield where the relevance and timeliness of information is crucial for success, AR devices will inevitably be used to provide information to the Soldier and gather information from the Soldier's immediate surroundings (Gans et al. 2015; see also, the US Army's IVAS program). The purpose of the AR HUD is to provide visualization of target information to operators, as well as to capture additional ground-level environmental (EO) and HMI (eye-gaze) data. The HL2 is a head-worn mixed-reality device developed by Microsoft that allows operators to see holograms integrated into their environment and play 3D spatial and audio data. However, a prevalent concern with deploying AR HUDs is the potential decrease in situational awareness when presenting visual information in real time. As noted by Geuss et al. (2020), ". . . the effectiveness of providing information through AR displays to improve situation awareness and decision-making is limited by an operator's ability to translate data displayed for them into actionable information." While the effect of visual stimuli on situational awareness is not a new research area, developing technologies such as wearable AR platforms allows researchers and human factors engineers to further investigate the "what" and "when" of information presentation on the battlefield. Current state-of-the-art HUD devices can record and stream a forward-facing point-of-view camera, binocular eye tracking, a map of the device's surroundings, and audio data from low-level sensors. Additionally, external data can be ingested by these devices to supplement visualization, such as information from the weapon or wearable sensors. Thus, a key deliverable from this effort is to develop an initial technical framework for examining the relationship between battlefield data, team situational awareness, and small-unit tactical performance.

3. Methods

3.1 Vignette

The demonstration described in this report was developed and implemented at ARL's Robotics Research Collaboration Campus (R2C2) located in Graces Quarters, Maryland. The operational vignette used for this proof-of-concept demonstration was for a small team (i.e., four individuals) to track a single human target walking through a MOUT environment, to decide the threat level of the target, and propagate that decision throughout the team. Two riflemen (RFL) were positioned on the perimeter of the MOUT facility while the squad leader (SL) was positioned in a building nearby, as depicted in Fig. 2. The purpose of sectioning the SL away from the rifle team was to prevent line of sight to the target such that they must rely on the information propagated through the HL2. The two RFL were each provided identical instrumented rifle platforms on tripods. One RFL wore a calibrated HL2 headset, while the other HL2 headset was given to the SL. The UAS was launched and piloted to maintain an aerial view of the target from within the airspace in accordance with the R2C2 Standard Operating Procedure. System-specific instrumentation and calibration details are provided later in the report.

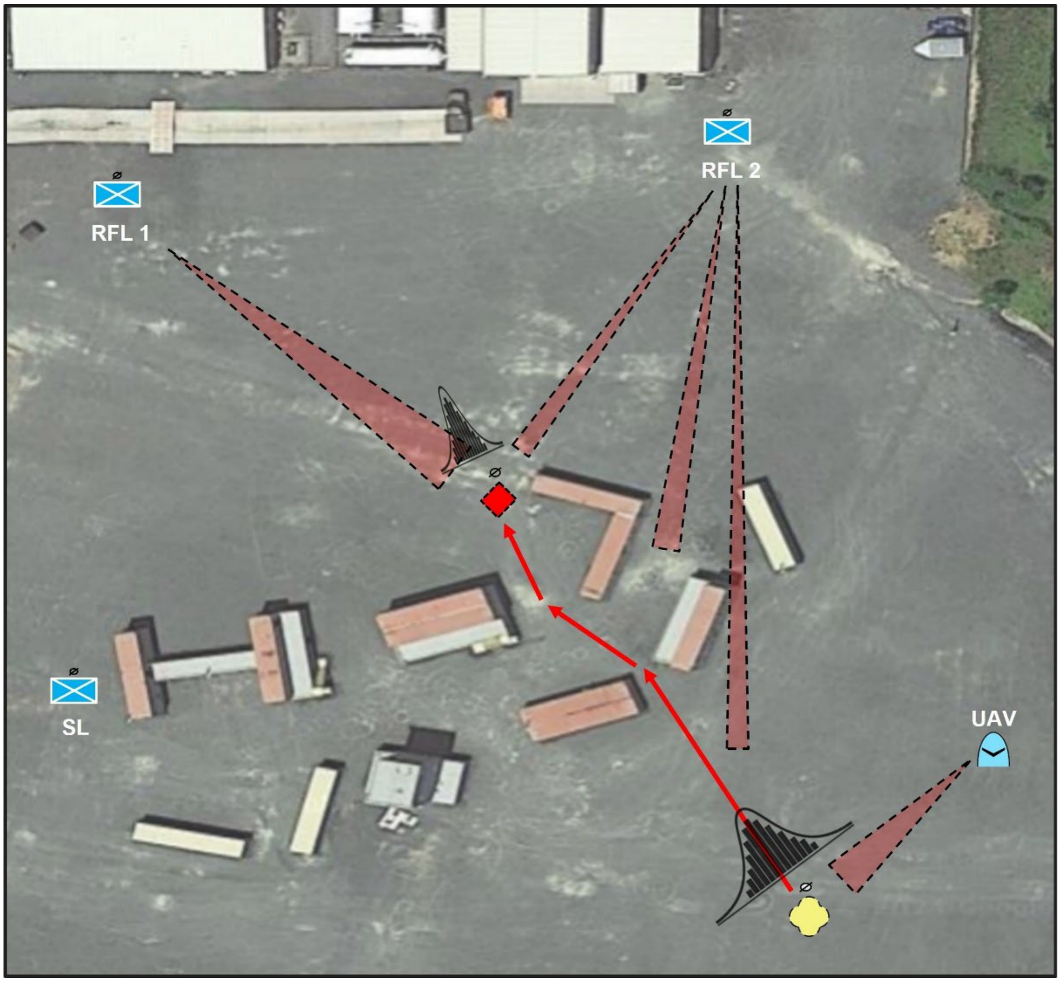

Fig. 2 Aerial view of MOUT environment and operational vignette for single-target tracking by a mixed-technology small team. MIL-STD-2525 (DOD 1994) icons are used to designate the location of the unknown entity spotted by security UAV. Distributions show representative uncertainty in location and state of target as it progresses through the MOUT environment. Red line indicates path of entity through MOUT buildings. Icon changes to hostile by end as SL makes decision regarding threat characterization.

The start of the scenario sequence consisted of the UAS initially finding and visualizing the target entering the area of operation. The UAS performs ATR to classify the target as human and geolocalization to provide GPS coordinates of the target to the network. The GPS coordinate was ingested by the HL2 and transformed to the relative coordinate frame of the operator such that an icon was projected (in AR device) in the direction of the target point of view (see Fig. 10), allowing the SL to track the target's position relative to the RFL. A MIL-STD-2525 (DOD 1994) icon (Fig. 9) for an unknown individual was displayed to the first RFL. The instructions were to track the target through the environment with the instrumented rifles. The instrumented rifles provided barrel-aligned EO imaging as well as GPS and weapon orientation. Similar to the UAS, the rifles perform both

ATR to classify targets as humans and use vector convergence for estimating geolocalization of the target. Additionally, the EO feeds from both rifles were transmitted to the SL HL2 and could be displayed for the operator in AR. The MIL-STD-2525 icon was also displayed to the SL in the direction of the target relative to their perspective. With the RFL and target locations visualized in their field of view and the live video feeds from the rifles, the SL was prompted to decide whether the target was hostile or friendly. Once a decision was received by the HL2 interface via holographic button, the decision was propagated to the RFL HL2 and the MIL-STD-2525 icon would change according to the SL's decision. The RFL would then continue tracking the target until the end of the sequence.

3.2 Network

The network architecture for data communications can be divided by system components into three auxiliary systems: the rifle, UAS, and HL2 networks (Fig. 3). This subsection describes the strategy for real-time data transmission to enable effective information exchange between component technologies and humans. The central hub of the network through which the three auxiliary network branches communicated was the Orbi router mesh network (NETGEAR Orbi 960 Series Quad-band Wi-Fi 6E Mesh). The base Orbi bridged the UAS network and the rifle network together via Ethernet connections to a switch and a laptop, respectively. It also broadcast a local Transmission Control Protocol (TCP)/IP Wi-Fi network to which the RFL HL2 connected. Additionally, the base formed a Wi-Fi wireless mesh with a remote Orbi access point that serviced the other HL2 (worn by the SL). Network endpoints were assigned static IP addresses to allow sensor nodes to be set up and identified consistently.

The UAS and rifle networks used the Robotic Operating System (ROS) messaging framework to communicate data to the central hub. The two instrumented rifles were connected via 100-m Cat5 Ethernet cables to a switch. Each rifle afforded a certain vantage point of the target in the MOUT site and published ROS 1.0 topics from its onboard sensors on the network. A laptop node, also hardwired to the switch, subscribed to the rifle topics and served as an intermediary, republishing the rifle sensor data again under ROS 1.0 onto the downstream Orbi network. For simplicity, instead of running an ROS node on the HL2's limited ARM processor, a laptop was used as a bridge to receive ROS messages and then port them to the HL2. The laptop also runs an application that can preview ROS messages as they are being forwarded to the HL2 device. We refer to this application as the "Server App." The Server App listens to ROS messages published from the two rifles and UAS. These messages comprise GPS location data from both weapons, the UAS, and their calculated target locations. There are also aggregated video feeds from the rifles and a video feed from the UAS. Furthermore, the Server App provides a preview of the incoming rifle video feed as well as a means for controlling the state of any detected target (as described in Section 3.1).

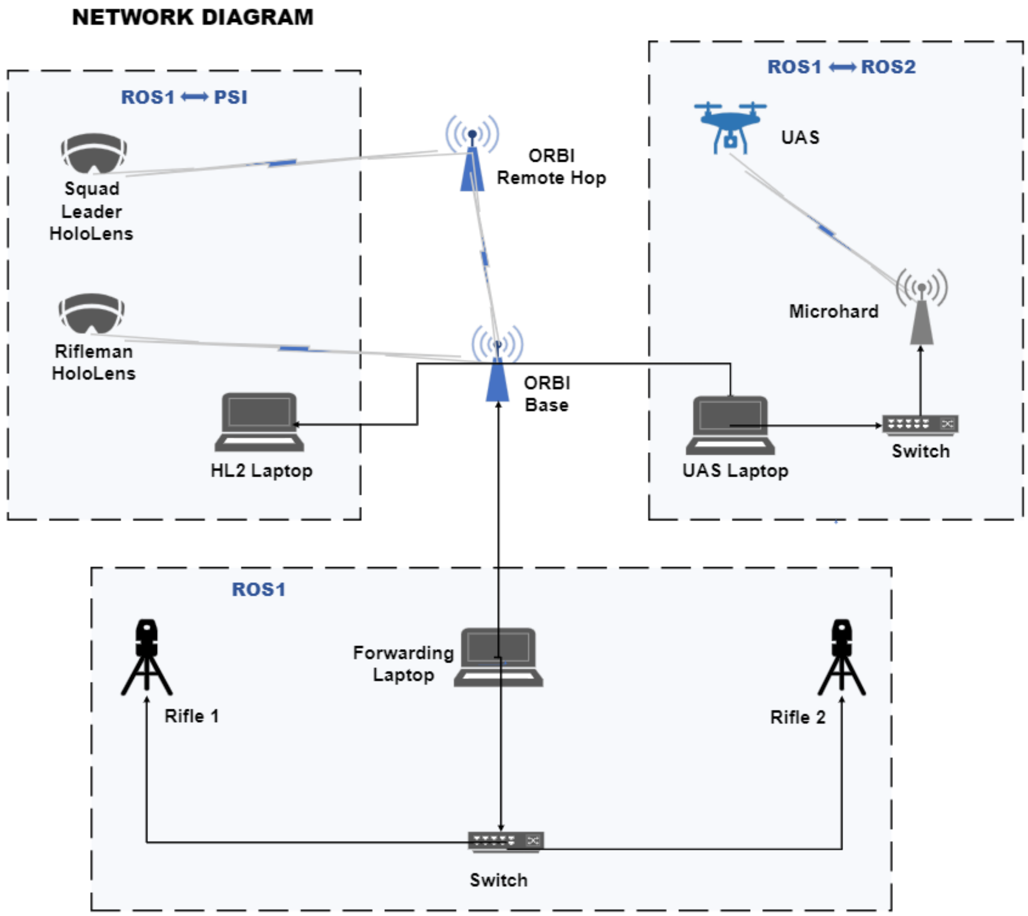

Fig. 3 Network system architecture for perception and tracking pipeline

3.3 Custom Instrumented Rifles

The instrumented rifle has three essential hardware components: a bore-aligned EO camera, a rigidly attached sensor suite, and a scope attachment (Fig. 4). The custom software components include a modified IMU device driver, camera device driver, and communications software. The camera is a film-through-scope camera with resolutions of 1080p when live streaming over a USB connection. The scope attachment allows the film-through-scope camera to be housed next to the scope. The scope attachment allows a rifle operator access to the scope while simultaneously allowing the camera to capture the same imagery seen by the rifle operator. The simultaneous and identical capture of the imagery allows the framework to process imagery in real time. Custom 3D-printed housings with white acrylonitrile butadiene styrene plastic allowed for easy mounting of the sensor suite to the Picatinny rails of the rifle. The sensor suite is composed of an IMU, GPS, and magnetometer, packaged together as an autopilot in a small form factor.

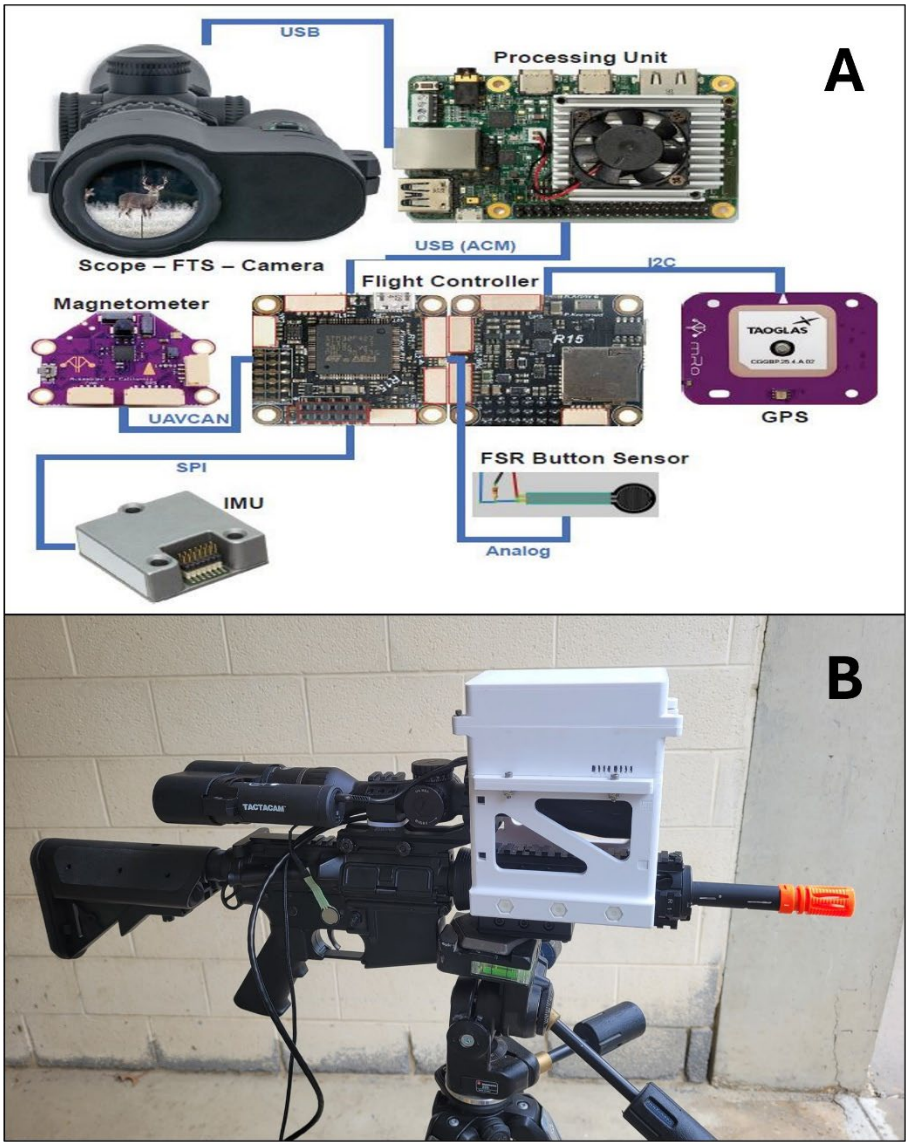

Fig. 4 (A) Instrumented rifle hardware component diagram and (B) rifle platform on tripod. Force sensitive resister was not used in this demonstration.

For software, a device driver for a PX4 Autopilot is added to use an external IMU (i.e., ADIS16465 IMU). The IMU features a tactical-grade performance with Angular Random Walk <0.2°/√h and Gyroscope Noise ≤2°/h. The details of the device driver are described by Han and Sabbagh (2022); the driver can be updated with the latest firmware of the PX4 Autopilot with minimal effort. An open-sourced Linux device driver for a webcam is used to capture the imagery from the external camera. With the vendor, product, and index ID, the device driver connects to the external camera and provides the imagery as a topic through ROS. The external camera, mounted on the rifle, produces an HD 720p resolution imagery at 30 fps in a motion JPEG (MJPEG) compressed format over a wired USB-C connection, consuming 85 MB/s of network bandwidth. Finally, a Node.js application was written to configure a graphical user interface (GUI) to comprehensively monitor all data streams from both rifles.

Using the "roslibjs" library,[*] WebSockets were made with a rosbridge server (rosbridge_server package) and provided publishing and subscribing, among other essential ROS functionality (Toris 2022). Chart.js[†] supplied the charts in the interface (Chart.js 2023). Additional charts, such as bar graphs, can be implemented to visualize more data streams. The ROS package web_video_server opens a local port and waits for any incoming HTTP requests to be made. When a video stream of a ROS image topic is requested with an HTTP request (e.g., http:localhost:8080/stream_viewer?topic=/camera/image_raw), the web_video_server subscribes to the requested topic and provides a video player playing the image stream. Video configurations can be specified by appending parameters to the query string.

The GUI (Fig. 5) displays live streams of the scope imagery from both rifles simultaneously, overlapped by bounding boxes from the object detection algorithm running in real time. A 2D map-tile view is placed in the center of the GUI. The map incorporates sensor measurements from the rifles and displays them to produce a geolocalization estimate of a target. The locations of the arrow icons represent the locations of the rifles updated by the GPS sensors; the headings of the icons represent the headings of the rifles updated by the magnetometer sensors. Each icon has a vector originating from the rifles' latitude and longitude in the heading.

[*] roslibjs is a JavaScript library for interacting with ROS from the browser.
[†] Chart.js is an open-source JavaScript application development charting library.

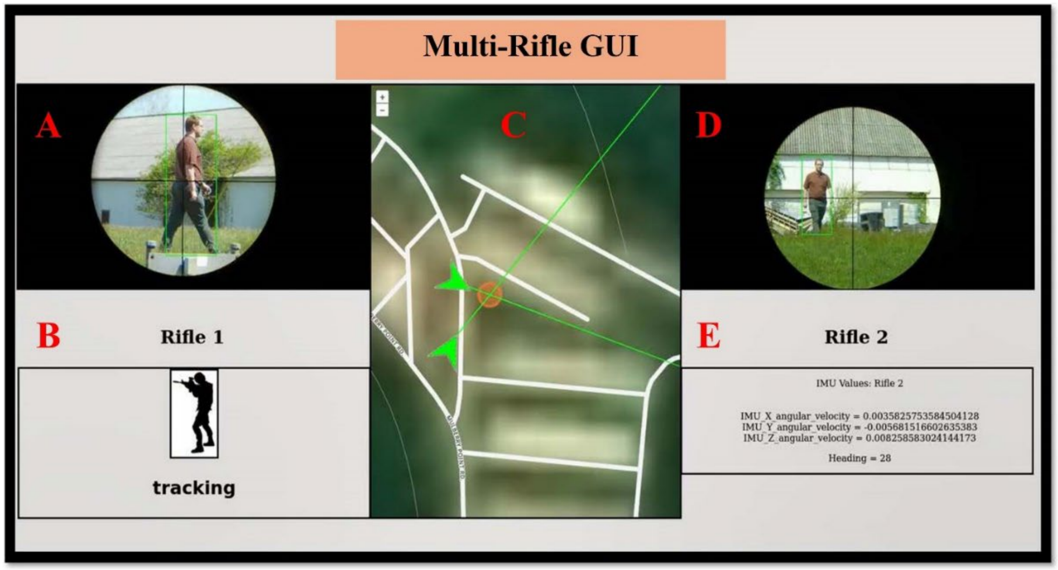

Fig. 5 Example image of multi-rifle GUI for visualization of target geolocalization. From left to right: A) Scope view of rifle 1 with ATR target highlighting. B) Behavior classification algorithm results. C) Map tile displaying rifle orientation vectors and target location estimation. D) Scope view of rifle 2 with ATR target highlighting. E) Raw IMU data from rifle 2 display.

3.4 UAS

UAS implementation used a Teal Golden Eagle V2 vehicle (Fig. 6). The Teal Golden Eagle V2 is representative of the current commercial state of the art, but requires all processing to occur on the ground control station computer. Alternatively, the ModalAI VOXL2 vehicle requires more effort in terms of custom setup, but all processing can occur onboard the vehicle, which has led to a great increase in overall frame rate for image processing. Thus, both vehicles were tested (nonsimultaneously) during network integration to ensure adequate information bandwidth for delivery and processing by human operators.

Fig. 6 BlueUAS Teal Golden Eagle V2

When using the Teal Golden Eagle V2, the onboard radios were modified from point-to-point mode to "mesh" architecture to permit inclusion of a third radio to downlink the imagery while in flight. The Teal Golden Eagle vehicles are Robotic and Autonomous System-Air Interoperability Profile–compliant and hence can interface using MAVLink (Linne 2022)—i.e., the QGroundControl station software. For target detection, the highest-quality image is desired; however, when flying drones, typically only a compressed video stream is available, which is prone to artifacts. For this application, we assume we can endure a reduction in frame rate to achieve higher image quality. Rather than use the compressed video stream, we set the vehicle "Camera Trigger" parameter to "time-based; always on" and set the interval to 1000 ms. This sets the drone to take imagery at a regular interval without an explicit trigger each time. A Python-based ROS1 node was used to subscribe to the mavlink-based "CAMERA IMAGE CAPTURED" message and download automatically collected JPG imagery to the ground control station and then subsequently publish as a ROS1 image message. The ROS1 image message is consumed at the ground control station by a ROS1-based YOLOv5 script that provides the object detection. While this method results in a detection frame rate of about 1 Hz, the advantage is that the systems are commercially available and require very little modification to integrate the above detection pipeline.

The VOXL2-based setup (Fig. 7) represents a more tightly integrated setup, resulting in a much higher frame rate as the object detection is performed on the vehicle "at the edge." A custom detection classifier is modified to operate with the hardware acceleration available on the VOXL2 Adreno GPU by formatting the model to the ".tflite" standard and dropping in place of the existing YOLOv5 classifier. Since there are bandwidth limitations between the UAS and the ground station with ROS2, a developed ROS2 package (TCP bridge) is used to receive image streaming from the ground station with ROS2. The TCP bridge is based on TCP, and therefore, guarantees more-reliable data transfers. The image streaming using TCP bridge is limited to 4–5 Hz as the data are sent as uncompressed image topics; however, this is sufficient to enable situational awareness on the ground.

Fig. 7 ModalAI VOXL2-based UAS

With either system, the localization algorithm provided the inertially referenced position in local inertial coordinates. The "/fmu/out/vehicle local position" topic provided the required information to translate back to latitude and longitude from local coordinates (in meters). The following formulas were used:

$$latitude = \frac{Y_m}{R_E} \times \frac{180^\circ}{\pi}$$

$$longitude = \frac{X_m}{R_E \cos(lat^\circ)} \times \frac{180^\circ}{\pi}$$

As stated above, the YOLOv5 detection model, comes pretrained on the Microsoft COCO dataset (Lin et al. 2015), wherein "person" is the first class of object. We enhanced detection range by leveraging a model that is fine-tuned using additional aerial viewpoints (Maxey et al. 2023). Furthermore, for the custom VOXL2 vehicle, we used a relatively uncommon lens style with an extremely long focal length, which theoretically enables enhanced localization accuracy at longer distances when using a nonvariable focal length lens. A typical localization operation is presented in Fig. 8, wherein a target is placed on a local inertial map. Some tracking and bookkeeping operation will be required to ensure multiple targets in the field of view can be independently tracked; however, this capability is not currently in the pipeline. It is also yet to be determined how to best apply the tracker within the local inertial space.

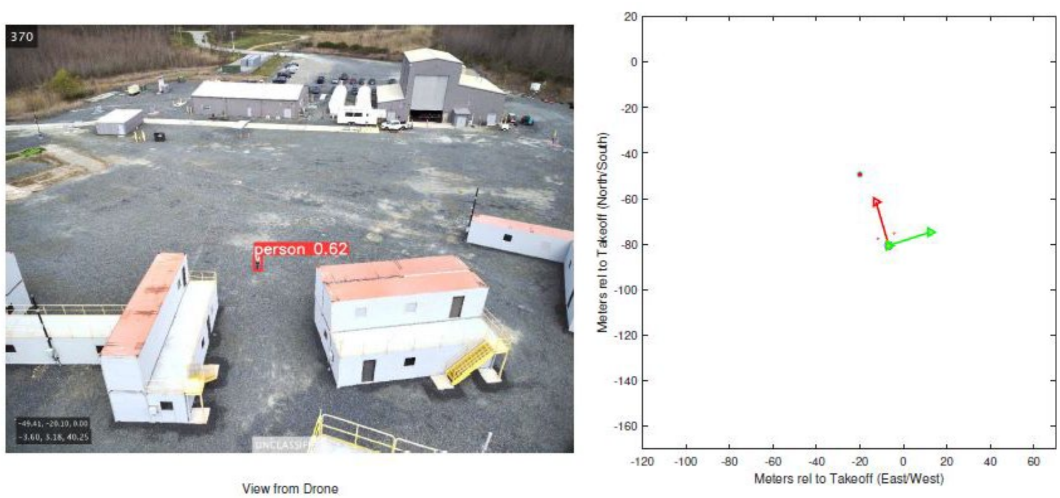

Fig. 8 Exemplar frame of imagery with corresponding localization estimate from BlueUAS Teal Golden Eagle

3.5 AR HUD

The HL2 was initially designed as an indoor device and does not have GPS sensors. To translate GPS coordinates into a local coordinate space for accurate holograms displays, a quick-response (QR) code was used to set a known GPS coordinate point. From this point, the Haversine function (see formula below to calculate distance between points A and B) was used to calculate distances between the operator and the target in both north–south and east–west directions.

$$a = sin^2\left(\frac{lat(B) - lat(A)}{2}\right) + cos\big(lat(A)\big) * \cos\big(lat(B)\big)$$
$$* \, sin^2\left(\frac{long(B) - long(A)}{2}\right)$$
$$c = 2 * atan2(\sqrt{a}, \sqrt{1-a})$$
$$Geospatial\ distance\ between\ points\ A\ and\ B = 6371000 * c$$

During calibration, a menu with slider bars was displayed to the operator to provide a way to rotate the in-device coordinate system around the origin point. This allowed the operator to align the in-device coordinate system to the north-facing GPS coordinate frame. The icons displayed in the device are 3D renderings of the MIL-STD-2525 (DOD 1994) icons for unknown, hostile, and friendly (Fig. 9). The icons were created using Blender 4.0. By using a standard size, the icons give a perception of depth to the operator as the icon is displayed at varying distances (Fig. 10).

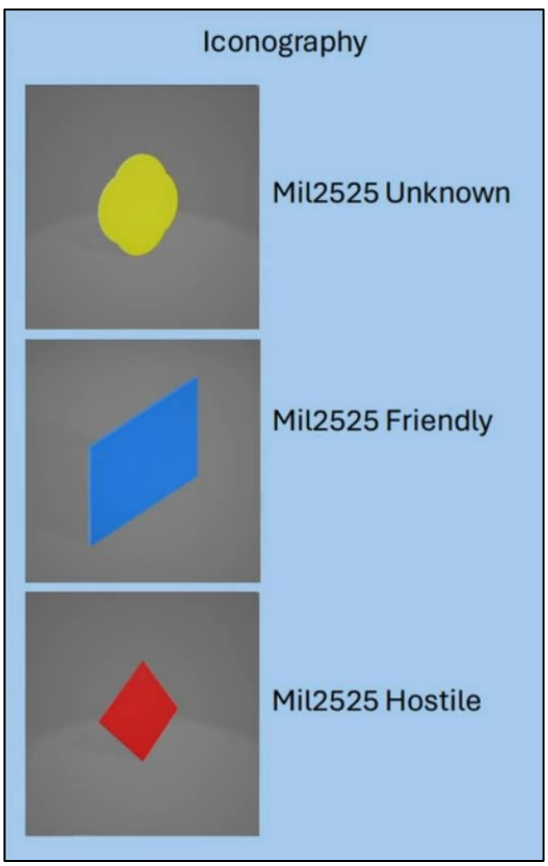

Fig. 9 AR projections of MIL-STD-2525 (DOD 1994) iconography

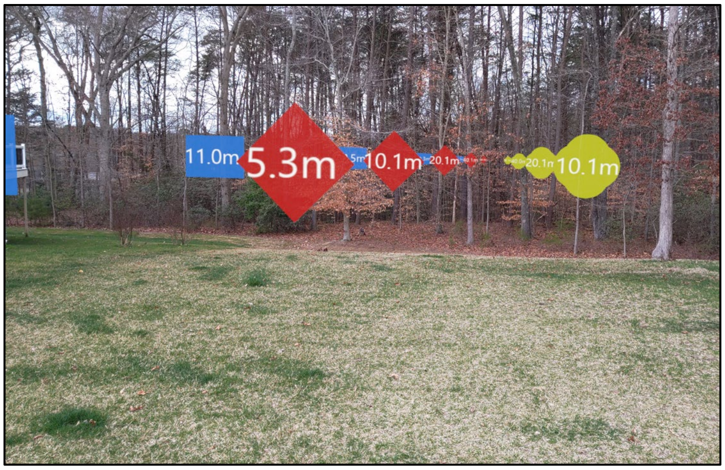

Fig. 10 MIL-STD-2525 (DOD 1994) iconography visualized at various distances using HL2. Distance from observer is noted in each icon.

A state machine controls whether the icon displayed is unknown, hostile, or friendly. In all states, coordinates marked as known friendlies (e.g., RFL positions) are displayed with the friendly icon. In state one, the "unknown" state, unmarked coordinates display the unknown icon. When a signal from the rifles or drone indicate that a target has been detected, the state machine transitions to state two. At this point, the SL receives information from the rifles as video feeds of the target (described in Section 3.1). Upon the designation as hostile, the state machine transitions to state three. In state three, the target's coordinate point is displayed as the hostile icon. From here, the state can be set back to state one or state two from the Server App.

4. Proof-of-Concept Demonstration

The multisystem proof-of-concept demonstration was successfully executed in April 2024 at R2C2 in Graces Quarters, Maryland. A small team of operators with instrumented rifles, HUDs, and a UAS positively tracked a human walking through a MOUT environment with several buildings disrupting the line of sight (Fig. 11). Information was shared between systems (e.g., UAS and HL2 or RFL and HL2) that allowed an SL to survey the target and adjust the threat level (Fig. 12). Full video of the vignette is publicly available upon request to primary author.

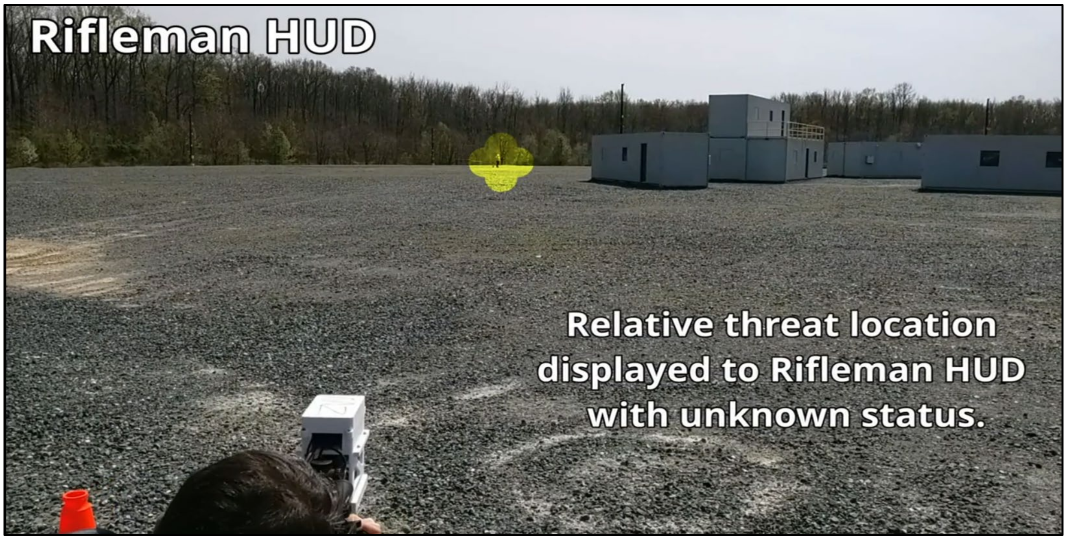

Fig. 11 Image from RFL 1 HL2 with MIL-STD-2525 (DOD 1994) unknown icon aligned with target location

Fig. 12 Image from RFL 1 HL2 with squad leader HL2 view overlaid in picture-in-picture

Analysis is still ongoing to quantify the error between geolocalization of the target via the multi-rifle system, the Teal Golden Eagle UAS, and the ground truth global navigation satellite system modules, which may require additional test runs to establish an error distribution. However, during development and testing of this vignette, the multi-rifle geolocalization maintained fairly good accuracy (as determined by AR overlay) and did not cause any issues visualizing the icons in the HUD.

Despite the success of this proof-of-concept demonstration, there were several issues. Primarily, hologram projection contrast in sunlight is still an ongoing issue despite the consistent development in the AR HUD display capabilities. The difficulty that sun luminance levels cause for operators to perceive overlaid information in the AR display has been noted since 1999 (Azuma 1999) and continues to be explored (Erickson et al. 2020). Secondly, as with any heterogenous ecosystem approach, network and data schema differences preclude easy integration. This is especially true with research-grade systems such as the ModalAI UAS and instrumented rifles. The current setup configuration required multiple data bridges, including ROS1 to ROS2 (Teal Golden Eagle UAS) and ROS1 to PSI (HL2). Additionally, the Teal Golden Eagle V2 vehicle—while easy to deploy, reliable in flight, and suitable for a wide range of weather conditions—does not support onboard tight time-stamping of imagery and metadata, which greatly aids the localization estimation. Furthermore, the design precludes any AI/ML-based onboard processing of the visual data. The result of this is that processing must occur at the ground station, which greatly increases the network bandwidth and decreases the image processing rate. Follow-on efforts shall use a platform, such as the QRB5165-based VOXL2 or similar, which supports onboard processing for enhanced time-stamping and update rate. Onboard processes can also significantly decrease the network bandwidth by only transmitting frames with a positive detection. These could be critical failure points in a deployed setting; however, they were necessary for the proof-of-concept demonstration. Future iterations of testing should include more cohesive network infrastructure (i.e., ARL's AURORA; Dennison and Trout 2020) for seamless integration of external systems.

5. Conclusion

In order to modernize the Army's ground forces by 2040, a more holistic, integrated approach is necessary during the early phases of research and development. This holistic approach includes the recent push for better HMI into the Infantry formation (HMI-F). This proof-of-concept demonstration showed that an integrated approach to dismounted Soldier–AI technologies, centered around the Soldiers' innate needs and capabilities, is not only feasible but necessary. The simple vignette shown in this effort was for single target detection, tracking, and threat evaluation; however, the utility of this integrated approach will likely increase as the number of targets and the complexity of the environment increases. Thus, a key outcome for this effort was to establish a means to research HMI in more realistic environments.

5.1 Near-Term Operational Capabilities

A networked system of weapons, HUDs, UASs, and other AI tools may allow for some near-term small unit operational capabilities. The utility of visualizing information to an operator in the field via HUD operator is obvious and has been fielded in limited capacities (Shear 2023). However, data coming in from other sources in the field, such as target location via UAS or other rifle systems, can help alleviate the need for "manual" communication between leaders and the squads. Manual communication refers to text, radio, smoke, or other means that require the operators to explicitly communicate the information and understand its meaning, content, and location, in the specific context of the fight.

We can better utilize the technology in a squad fire control setting by using the HUDs to visualize the following: 1) target handoffs via ATR, 2) target reference points based on spatial anchoring to objects in the environment, 3) shift fires indicators based on visualization of known assault elements and estimated enemy locations, and 4) initiating changes in rate of fire based on operator–weapon behavior and status. UAS elements that are persistently aware of the relative location of the Blue Force elements in relation to itself and enemy location, based on ATR may have defensive and offensive maneuvering capabilities to best support the operators on the ground by providing 360° security and coverage of blind spots. These UAS capabilities are greatly enhanced by tracking and understanding the operators' behaviors in the field, which can be done via intelligent weapons and HUD technologies.

5.2 Future Research Directions

To push the bounds of the operational capabilities mentioned above, research is still required in the HMI-F domain. This includes expanding to higher-level concepts, such as command and control at higher echelons. The data provided by this proof-of-concept system can be further integrated in future cross-reality common operating pictures (Dennison and Trout 2020), which give commanders wider situational awareness by providing a distributed, virtualized environment to control assets in and have those commands propagated down to the squad. While outside the scope of the current project, research is still needed to understand how humans (commanders) ingest 3D field information in virtualized settings.

As alluded to in Section 5.1, robotic capabilities will be leveraged extensively in near-term operations. A key area of research relevant to small-unit performance will be the bidirectional adaptation of humans and autonomous systems in shared, complex environments. Research questions such as, "How do humans adapt their roles during operational tasks in the presence of autonomous agents?" and "What learning strategies must autonomous agents implement to maintain or increase performance of their human counterparts?" will be explored. These questions require a mix of novel human–AI experimental methodologies and an increase in behavioral inference capabilities from humans.

6. References

Antipov VM, Badarin AA, Grubov VV, Kazantsev VB, Hramov AE. Portable tracker for neurophysiological research of sport shooting. Proc. SPIE 12194, Computational Biophysics and Nanobiophotonics; cApril 2022; 121940M. https://doi.org/10.1117/12.2626382

Azuma RT. The challenge of making augmented reality work outdoors. In: Ohta Y, Tamura H, editors. Mixed reality: merging real and virtual worlds. Springer-Verlag; c1999. p. 379–390.

Barber DB. Accurate target geolocation and vision-based landing with application to search and engage missions for miniature air vehicles [master's thesis]. Brigham Young University (US); 2007 Feb. https://scholarsarchive.byu.edu/etd/1313

Barber DB, Redding JD, McLain TW, Beard RW, Taylor CN. Vision-based target geo-location using a fixed wing miniature air vehicle. Journal of Intelligent and Robotic Systems. 2006 Nov;47(4):361–382. http://dx.doi.org/10.1007/s10846-006-9088-7

Chart.js. [accessed 2023 Feb]. https://www.chartjs.org/docs/master/

Dennison M, Trout T. The Accelerated User Reasoning for Operations, Research, and Analysis (AURORA) cross-reality common operating picture. DEVCOM Army Research Laboratory (US); 2020 Apr. Report No: ARL-TR-8940. https://apps.dtic.mil/sti/pdfs/AD1096359

[DOA] Headquarters, Department of the Army. US Army modernization strategy: investing in the future. Headquarters, Department of the Army (US); 2021.

Erickson A, Kim K, Bruder G, Welch GF. Exploring the limitations of environment lighting on optical see-through head-mounted displays. Proceedings of the 2020 ACM Symposium on Spatial User Interaction; 2020. Association for Computing Machinery, New York; c2020. Article 9, p. 1–8. https://doi.org/10.1145/3385959.3418445

Gans E, Roberts D, Bennett M, Towles H, Menozzi A, Cook J, Sherrill T. Augmented reality technology for day/night situational awareness for the dismounted Soldier. Proc. SPIE 9470, Display Technologies and Applications for Defense, Security, and Avionics IX. Head- and Helmet-Mounted Displays; c2015 May. p. 947004. https://doi.org/10.1117/12.2177086

Geuss M, Cooper L, Bakdash J, Moore S, Holder E. Visualizing dynamic and uncertain battlefield information: lessons from cognitive science. Proc. SPIE 11426, Virtual, Augmented, and Mixed Reality (XR) Technology for Multi-Domain Operations; c2020 Apr. https://doi.org/10.1117/12.2558509.

Han S, Sabbagh P. Integration of the ADIS16465 tactical-grade inertial sensor with the open-source Pixhawk autopilot and PX4 flight stack. DEVCOM Army Research Laboratory (US); 2022 May. Report No.: ARL-TR-9463. https://apps.dtic.mil/sti/citations/AD1169845

Heckman L, Magnuson S, Park A. Army lays out plan for robotic combat vehicles. National Defense Magazine; 2023 Dec.

Lee M, Tweedell A, Dennison M, Sabbagh P, Conroy J, Trout T, Freeman J, Lance B. Modeling and analysis of motion data from dynamic Soldier state estimation to enable situational understanding. Proc. SPIE 12125, Virtual, Augmented, and Mixed Reality (XR) Technology for Multi-Domain Operations III; c 2022 May. p. 121250E. https://doi.org/10.1117/12.2618634

Lin TY, Maire M, Belongie S, Bourdev LD, Girshick RB, Hays J, Perona P, Ramanan D, Dollar P, Zitnick CL. Microsoft {COCO:} common objects in context. Computer Vision and Pattern Recognition; c2014 Feb; p. 1405.0312. http://arxiv.org/abs/1405.0312

Linne B. Streamlined development pipeline for MAVericks, ARL's unmanned autonomous vehicle (UAV) software. DEVCOM Army Research Laboratory (US); 2022 Sep. Report No: ARL-TR-1134. https://apps.dtic.mil/sti/pdfs/AD1181447

Maxey C, Choi J, Lee H, Manocha D, Kwon H. UAV-sim: NERF based synthetic data generation for UAV-based perception. Computer Vision and Pattern Recognition; c2023 Oct. p. 2310.16255. https://arxiv.org/abs/2310.16255.

MIL-STD-2525. Common warfighting symbology (version 1). Department of Defense (US); 1994 Sep 30.

Redding J, McLain T, Beard R, Taylor C. Vision-based target localization from a fixed-wing miniature air vehicle. American Control Conference; c2006.

Shear F. Army accepts prototypes of the most advanced version of IVAS. Department of Defense (US); 2023 Aug. https://www.army.mil/article/268702

Toris R. roslibjs - the standard ROS JavaScript library. n.d. [accessed 2022 May]. http://wiki.ros.org/roslibjs

List of Symbols, Abbreviations, and Acronyms

2/3D	two-/three-dimensional
AFC	Army Futures Command
AI	artificial intelligence
AR	augmented-reality
ARL	Army Research Laboratory
ATR	Aided Target Recognition
COCO	Common Objects in Context
DEVCOM	US Army Combat Capabilities Development Command
EO	electro-optical
fps	frames per second
GPS	global positioning system
GUI	graphical user interface
HL2	HoloLens 2
HMI	human–machine integration
HMI-F	human–machine integrated Infantry formation
HTTP	Hypertext Transfer Protocol
HUD	heads-up display
ID	identification
IMU	inertial measurement unit
IP	Internet Protocol
IVAS	Integrated Visual Augmentation System
MDO	Multi-Domain Operations
MJPEG	motion JPEG
ML	machine learning
MOUT	military operations on urbanized terrain
PM	Program Manager

QR	quick-response (code)
R2C2	Robotics Research Collaboration Campus
RFL	riflemen
ROS	Robotic Operating System
SL	squad leader
TCP	Transmission Control Protocol
UAS	unmanned aerial system
UAV	unmanned aerial vehicle
UGV	unmanned ground vehicle
USB	Universal Serial Bus

1 (PDF) DEFENSE TECHNICAL
INFORMATION CTR
DTIC OCA

1 (PDF) DEVCOM ARL
FCDD RLB CI
TECH LIB