Special Issue: Dynamic Large-Scale Swarm Systems in Urban Environments: Results from the DARPA OFFSET Program

Field Report

Command and Control of a Large Scale Swarm Using Natural Human Interfaces

Brian M. Williamson[1], Eugene M. Taranta II[2], Yasmine M. Moolenaar[1] and Joseph J. LaViola Jr[1]

[1]Department of Computer Science, University of Central Florida, Orlando, FL 32816
[2]Northrop Grumman Corporation, Mission Systems, Orlando, FL 32817

Abstract: While swarm robotics contain varying levels of autonomy, it is still important for human operators to command and control the swarm before and during mission operations. As such, human-in-the-loop interfaces must be developed that are intuitive to the operator and able to relay accurate commands to flight and ground robots in a large-scale swarm network. Furthermore, the tremendous amount of information gathered by the swarm must be sorted and presented back to the user in a way that reduces cognitive load so that they may determine the relevancy of the information and its effect on the mission. In this paper, we present sketch-based and augmented reality interfaces that both allow for command and control (C2) of unmanned ground and flight vehicles and for processing of data that the swarm has gathered in the field. These interfaces were tested in two field experiments with multiple aerial and ground robots deployed on a mission. Finally, we discuss the results of the field experiments, lessons learned, and areas for future work.

Keywords: human robot interaction, robot teaming, control

1. Introduction

With the advancement of unmanned ground and flight vehicles, a reality where human-robot teams can be deployed into the field becomes a possibility (Hoffman and Breazeal, 2004). In the situation where a swarm of autonomous robots act based on the commands from a human operator, it becomes imperative to develop intuitive interfaces between the two systems. In this realm of research, many interfaces have been attempted (Sarkar et al., 2016; Kolling et al., 2012; Nourmohammadi et al., 2018). The human operator's interface must include both the means to quickly send commands to the swarm and to process the immense amount of information returned from the robots. In terms of military applications this means a limited number of operators commanding up to a 100 robots simultaneously in a nonlinear dispersed operation (Edwards, 2005). While interfaces have been developed for human and swarm interaction (Naghsh et al., 2008; Bashyal and Venayagamoorthy, 2008; Podevijn et al., 2012), rarely have they been tested at such a large scale.

Received: 31 March 2022; revised: 14 August 2022; accepted: 6 January 2023; published: 21 February 2023.

Correspondence: Brian M. Williamson, Department of Computer Science, University of Central Florida, Orlando, FL 32816, Email: Brian.M.Williamson@knights.ucf.edu

DOI: https://doi.org/10.55417/fr.2023009

1.1. Research Question

We ask whether advanced user interfaces can effectively be used in a large-scale deployment of robot systems and whether human-swarm interactions that were previously developed for a small number of robots can scale to the deployment of a hundred active robots within a swarm? Furthermore, can human-swarm interactions effectively task a large swarm in a simulated mission scenario?

1.2. The Developed User Interfaces

While the keyboard and mouse is the standard interface, there is still the question of how these hardware components can best be utilized by an operator. Furthermore, with advancements in augmented reality headsets, there is now the question of how this technology can be incorporated to provide enhanced command and control (Moon et al., 2014). In this paper, we go over the design of such interfaces, their deployment in field experiments, and discuss how their operations work in the field. Our interfaces drew from existing research on human and robot interactions (Sakamoto et al., 2009; Cacace et al., 2016; LaViola, 2015; Bashyal and Venayagamoorthy, 2008), iterating upon them to meet the large scale needs of the field experiment.

Displaying a map of the environment that a user sketches onto has been shown as an effective means to display information to a user and reduce cognitive load (Larochelle et al., 2011). This leads to our sketch-based interface, which can utilize either a stylus or a mouse. This interface works to display information gathered by the robots onto a map which the user can easily navigate. For commands, they can quickly draw gestures and fill in any additional parameters needed either with further gestures or via keyboard input. Gestures were designed to be as intuitive as possible given the command they were correlated to.

The concept of displaying virtual assets in a mixed reality application to users has been seen as an effective means to convey information (Holzbach, 2008). As such, we created an augmented reality interface app deployed to the Hololens 2. This interface utilizes hand tracking technology to both recreate features available in the 2D sketch interface, and provide a 3D sketch interface to the user. It displays map information in multiple forms to the user: as a top-down view, as a scaled 3D map, and from the grounds-eye view. Furthermore, the augmented reality interface allowed for voice commands, creating a multimodal interface for robot tasking LaViola Jr et al. (2014).

Field experiments were conducted with both interfaces. The sketch interface was used as the primary command and control system for the swarm while the augmented reality interface was used for field operations given its enhanced situational awareness. We demonstrate the capability of the interfaces to work together, with the sketch interface supplying route information to a field operator and the augmented reality interface marking off no-go zones for rear command. Both interfaces show themselves to be intuitive and easy to use, capable of handling multiple robots simultaneously and presenting the information back to users without distracting them from physical observations. This demonstrates how new technology may be utilized to create novel interfaces with field robotics.

2. Related Work

There has long been the idea that even autonomous robots require some level of supervision. In Sheridan (1992), research went into the requirements for supervisory control of robots in a human-robot team. It calls for the need for automation combined with a high level language and situational awareness for the operator. This has led to the concept of humans being able to control robots remotely, as seen in Luo and Chen (2000). This is further expanded in Kulakov et al. (2016) where long range operations are considered when supervising robots. In Koh et al. (2018) full body and hand gestures are analyzed as potential teleoperation controls of robots to complete key tasks.

In the realm of aerial robotics, there has been research into the ideal way to control unmanned aerial vehicles, particularly with 3D gestures and gaze control. Pfeil et al. (2013) looked at the possibility of using 3D gestures to control a robot and examined the efficacy of this interface over the normal game controller. In Cauchard et al. (2015) this research was expanded by adding in more user defined gestures for robot controls. In Hansen et al. (2014) gaze controls were used to control the flight path of a robot. Peshkova et al. (2017) considered a combination of input modalities such as 3D gestures, voice control and gaze control in an attempt to determine the ideal natural interface for a supervisory human and a singular robot. In Nagi et al. (2014) multiple aerial drones are controlled via methods to select one for direct control and then switch to others as needed. In the IMPACT project (Behymer et al., 2017) we see the desire to utilize intelligent agents as a portion of cooperative control algorithms (CCA) when planning multiple unmanned vehicles and the use of a map interface for mission planning and execution, similar to our own sketch interface. Furthermore the COUNTER project demonstrated the use of micro aerial robotics for surveillance purposes that are combined into a single visual system (Vigilant Spirit Control Station) built around an intuitive user interface (Feitshans et al., 2008).

Regarding sketch interfaces and robotic controls, there is a growing body of research to consider. In Skubic et al. (2007) sketching for robot controls was utilized, with the user drawing routes and other commands for a team of robots. While a map interface was not used, it did employ color coded information for the operator as only range detection feedback was available. In Sakamoto et al. (2009), sketching was used to control a home vacuum robot, which included the concept of the operator drawing out a region signaling to the robot as the "vacuum" zone. This is expanded in Liu et al. (2011) where tasks over time could be layered for a team of robots to execute, such as vacuuming before mopping. These concepts are similar to our zone drawing capability in the sketch interface. Shah et al. (2010) developed a fully probabilistic commanding interface that utilized Hidden Markov Models to recognize which gesture an operator was making and provided them the opportunity to confirm the command or decline and try the gesture again.

Multimodal interfaces have also proven effective with robots, such as in Correa et al. (2010) where a combination of sketching and voice commands were used to supervise an autonomous forklift. This concept is expanded in Taylor et al. (2012) where sketching and a dialogue between the robot and user are utilized to develop context toward the command. An intuitive supervisory control system can be utilized for an unmanned system. Most relevant to our task is Cacace et al. (2016) which developed a multimodal interface to control unmanned ground and flight vehicles for search and rescue operations in the Alpine mountains. This research abstracted direct control over the robots into more of generalized tasks that robots would then carry out. In general for aviation, augmented reality has shown promise such as in the SESAR-RETINA project (European Commission and Directorate-General for Research and Innovation, 2018) which utilized mixed reality for the purpose of improved air traffic control via enhanced situational awareness.

The field of swarm robotics has a rich history beginning with the concept of swarm intelligence (Bonabeau et al., 1999) in nature which focused on the idea of division of labor among creatures in the animal kingdom. From there an evolution took place of how these concepts can be applied into robotics (Beni, 2004). While there are many methods to accomplishing a robot swarm, several shown in Brambilla et al. (2013), there exist core principles in how they will operate. Primarily, the robots may operate without a centralized control or some external infrastructure, instead relying upon the principles of swarm intelligence (Dorigo et al., 2014). In recent swarm robotics research it is noted that all robots must be working together cooperatively on the same network to be considered a swarm network and that supervisory human operators may be introduced (Arnold et al., 2019).

While the ideal is total autonomy, there is still a need for certain operations which require a small team of human supervisors. As such, a field of research has grown in how human beings work specifically with the swarm robots. In McLurkin et al. (2006) the challenges of communicating to swarm robotics is outlined. It is a combination of understanding the information the robots present back to the operator and translating human commands in a way the swarm as a whole can understand. Other challenges involve human perception of swarm data are shown in Brown et al. (2016) where they focus on the span of the drone network and the duration of the interactions. Two particular challenges, selection and beacon control, are covered in Kolling et al. (2012) and how they can allow for a human operator to control a swarm. Naghsh et al. (2008) looked at the concept of human and robot swarm teams working together in a fire emergency, where humans are able to take control of individual robots as needed. In Bashyal and Venayagamoorthy (2008) an approach was taken where human operators were identified by the swarm as a fellow robot that can communicate with them, similar to the concept of human agents employed within our work. In Vasile et al. (2011) swarm interactions and their challenges are classified based on robot-environment interactions, robot-robot interactions and robot-human interactions. Podevijn et al. (2012) worked toward self-organized feedback in a swarm system, utilizing natural user interfaces such as 3D gestures via a Kinect to select and organize a series of ground robots. In both Giusti et al. (2012) and Labazanova et al. (2018) special glove devices were used to control a swarm, while high fidelity tracking devices can be replaced with hand tracking capabilities, as seen in Kim et al. (2020) and in the Hololens 2 tracking system that we utilized. Furthermore, field experiments have been conducted before to test human swarm interactions with real life scenarios, as seen in the MAGIC 2010 competition (Olson et al., 2012; Butzke et al., 2012). This experiment brought together robotics experts to build human-robot teams that would operate in previously unknown urban environments to fulfill mission goals.

For our system we utilized the combined research of 3D gestures and sketching to assist with swarm control. Regarding our sketch classifier, the work done in Taranta et al. (2017) was utilized as it is shown as adaptable to many forms of signal input and relies on little training data, ideal for our combination of 2D and 3D gesture data. Where hand tracking was available, such as in the augmented reality interface, we utilized principles outlined in LaViola (2015) as the ability to navigate and control virtual assets in VR could be translated well to the swarm network.

3. System Design

Two interfaces were designed to evaluate the use of advanced user interfaces in a command and control application of a large-scale flight and ground swarm. These interfaces were not seen as mutually exclusive as they could represent different operators working in different contexts on the same mission. The first interface is a sketch-based system upon which 2D data could be gathered from the user and translated into commands for the swarm network. The second interface utilizes augmented reality to recreate a 2D tablet on the user's hand and allows for 3D sketching data to be performed. While the interfaces were built from lessons learned in previous research (LaViola, 2015; Sakamoto et al., 2009), they were combined in a unique multimodal approach for command and control of a large-scale swarm network.

A centralized server application was built in PYTHON (PYC2) which handles the distribution of data between the interfaces and the swarm network. This system also keeps track of robot health and general status so that it can make decisions on which robots are capable of performing tasks together. This interface existed as a conduit into the swarm's mesh network capable of issuing tasks to the swarm which individual robots would bid upon to complete.

We go through both interface's capability to relay information back to the operator for situational awareness and command the robot swarm. We also identify key features within each interface useful for navigation and mission operations.

3.1. Sketch Interface

The sketch interface utilizes a 3D map designed in Unity 3D that reflects the mission environment. With this map, information can be relayed from the swarm and back to the user in a manner that reduces cognitive load (Larochelle et al., 2011). This information can be placed directly onto a location of the map relevant to what the robot discovers and utilizes common military iconography and color coding systems that the operator would be familiar with. This interface was designed to run on either a tablet PC with a stylus or a normal laptop with a mouse or touchpad. Figure 1 shows an example of the map that a user would first see, which included preloaded boundary lines and virtual buildings of the field experiment.

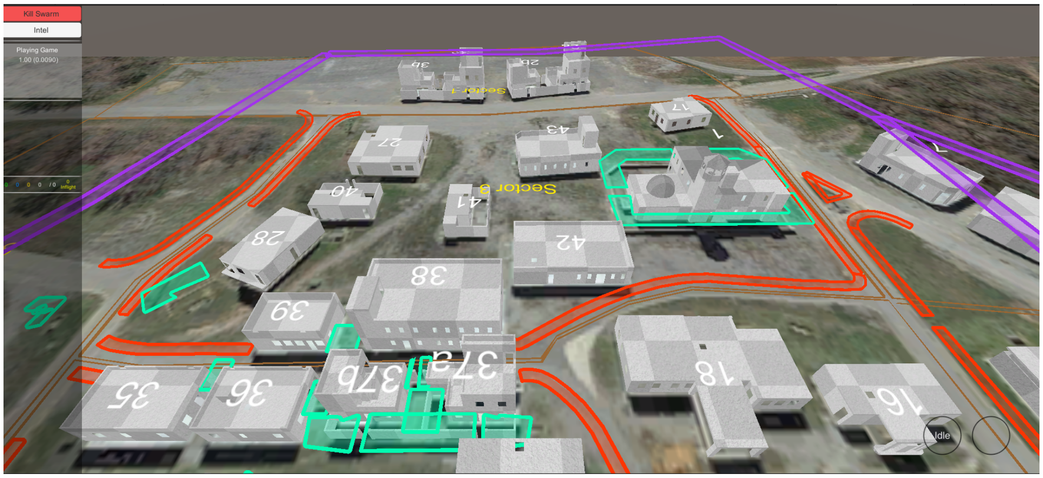


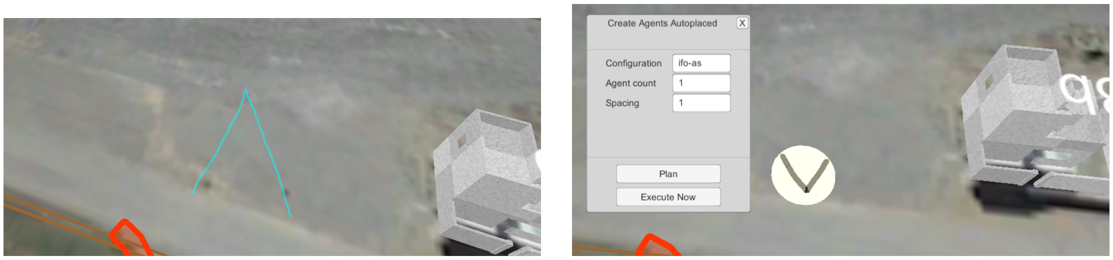
Figure 1. Tactical display an operator would work within for the field experiment.

Figure 2. Temporal coverage map with darker areas having not been visited. Over time the colors will shift showing areas that were visited, but should be updated.

3.1.1. Tablet Situational Awareness

As the mission is conducted an operator can view blue and red force positions, terrain layout and mission planning all with spatial context of the mission area around them. Furthermore, they can monitor a single robot within the swarm by selecting it then performing gestures to gather more information as needed. For example, by swiping right they can translate their camera to directly over the robot. Further gestures allow operators to view robot health in a popup window, display logs from the robot, or enable a view of the robot's camera and depth video feeds. All artifacts discovered by the swarm network are also displayed upon the map at the location the artifact was discovered and with appropriate military iconography.

A grid layout system was developed utilizing voxel overlays of the map. This creates color coded information that the operator can quickly switch to for enhanced situational awareness of the mission area. A threat probability overlay is available which presents high threat areas reported back by the swarm network in red. A temporal coverage overlay (Figure 2) is also available which shows areas



(a) Drawing of a gesture onto the map.

(b) Entering further parameters by clicking on the gesture of the map.

Figure 3. Issuing of a tactic by drawing directly onto the map. Location information is derived from the center location of the gesture.

of the map that have not been visited yet or were last visited some time ago. As time advances the colors shift so that new coverage tasks can be commanded.

3.1.2. Robot Command and Control via Tactics

Robots can be commanded via issuing high level tactics to the swarm network. These tactics are designed to be intuitive to human operators and representative of common mission tasks. The central server application, PYC2, decomposes tactics down into robot primitives which are then sent to the swarm network. The swarm then bids on which individual robots within the network are best suited to carry out the tactic and begin execution.

The issuance of tactics is done by the operator sketching a gesture directly onto the tactical display as shown in Figure 3a. The sketch must correspond to a command, a subset of which are shown in Table 1. The 2D data gathered from the mouse/stylus are run through the Jackknife classifier (Taranta et al., 2017) and matched to a sketch based on prerecorded training data. The centroid of the sketch data is used to locate the finalized tactic into real world GPS coordinates. Configuration parameters, such as how the recognized sketch relates to a tactic command, are loaded from the centralized server software (PYC2) to the sketch application. This allows for sketches to take on different meanings given different operators or mission parameters.

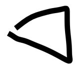
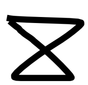
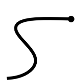
Once a sketch is recognized the gesture data is erased and replaced by a tactic icon. The user can then click on the icon to open up a parameter window. If additional parameters are required by the tactic, the user can enter them as shown in Figure 3b. If a tactic ends in the wrong location the user can simply click and drag it to a new position or erase the tactic completely by scribbling over it.

Tactics can be chained together by drawing links between the tactics on the ground which creates a temporal chain on which order the tactics will be executed (Figure 4a). Furthermore the user can select a subset of robots within the swarm to carry out a tactic by drawing a lasso over them, as shown in Figure 4b. The tactic would then be bid to those selected robots rather than the entire swarm network.

Once the user is ready they can select "Execute Now" and the tactic will be broken down into a series of robot primitive commands, such as "move-to location," to be bid on by the swarm network. As it is executed the status of the tactic is displayed in color coded information with blue meaning it is in progress, green that it was successful and red that the tactic failed. If needed the operator can click on a tactic to pull up relevant logs if they need more information on its status.

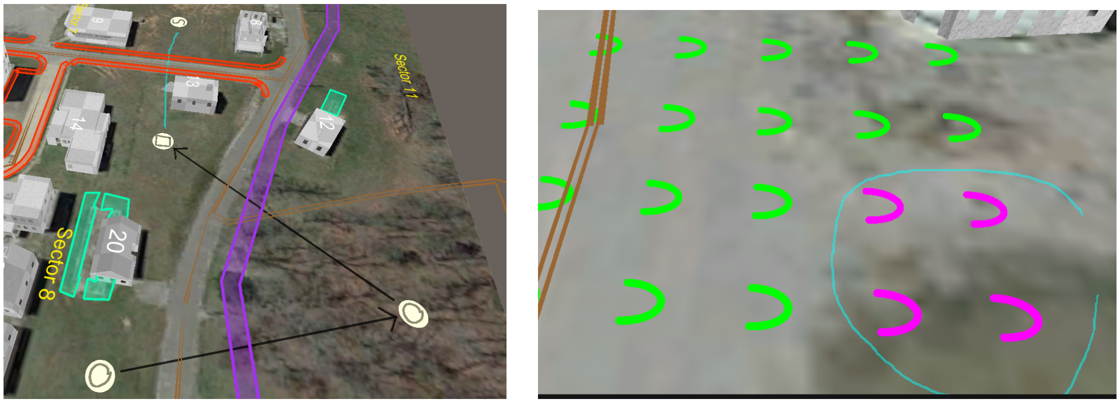
The user is also able to gesture map features to be distributed to the swarm network. These include singular points of interest that can be used in specific tactics such as "Examine Object" and polyline commands which were used for route drawing and boundary zones. Boundaries can include exploration zones, deployment zones, and no-go zones all of which create a closed loop area. If needed any polyline may be edited by selecting it and going into an editing mode where its boundaries can be modified.

Table 1. A subset of intuitive sketches that an operator could perform along with corresponding tactics to distribute to the swarm network.

Tactic Name	Gesture	Parameter	Type	Description
Acoustic Spoof		Context	None	Agent must move to this point and then make firefight sounds
Deploy		Context	DeploymentZone	Deploy autonomous agents for field use.
		UGV count	int	Number of ground agents to deploy.
		UAV count	int	Number of aerial agents to deploy.
Examine Object		Context	POI	Use UAV to scan an object of interest.
		Radius	float	Radius of sphere around object.
Follow Route		Context	Path	Request an agent to traverse the nearest path.
		Altitude	float	Height in meters that UAV will assume.
		Distance	float	Distance in meters between points. Zero to force simplification.
		Use Chaining	bool	Issue one point at a time, for testing only.
Hold Position		Context	Path	Move a set of agents to points along the perimeter and hold.
		Altitude	float	Height in meters that UAV will assume.
		Duration	float	How long to hold.
		Agent Count	int	Number of agents to place along perimeter.
Overhead Scan		Context	Explore	Fly UAVs over area to find artifacts.
		Altitude	float	Height in meters that UAV will assume.
		Cell Size	float	Minimum linear distance between waypoints in meters.
		Agent Count	int	Number of agents used to scan area.
Safe Land		Context	None	For a given air vehicle, find nearby safe location to land.
Secure Artifact		Context	None	Send ground agent to secure artifact.
		cf	int	Number of coverfire agents to send to the artifact.
Timer		Context	None	Wait a specified amount of time before returning tactic status.
		Seconds	float	Time in seconds to deplay completion response.
		Success	bool	Should tactic report success (True) or failure (False)
Scan Building		Context	Building	Scan outside of building using available scene geometry to guide path.
		Standoff	float	Distance from building walls when scanning in meters.
		Between	float	Distance between points in meters.
		Radar Scan	bool	Scan building for dismounted solders.

3.1.3. Single Point of Contact Camera Controls

Iterating from the Unicam approach (Zeleznik and Forsberg, 1999), users can quickly control the camera by interacting with one of the two circles stationed in the bottom right corner of the screen. If the user clicks or places, their stylus within the left circle, their next movement (either vertical or horizontal) would change the context to either fly or look. If they click into the right circle, their movement method would be to pan the camera as shown in Figure 5. In all of these modes, the



(a) By drawing a line between tactics users can chain them temporally.

(b) Users can select a subset of the robots in the swarm to carry out a task by lassoing them with their drawing.

Figure 4. Users have the capability to enhance tactic execution by chaining tactics together in sequences and selecting subsets of the robot swarm.

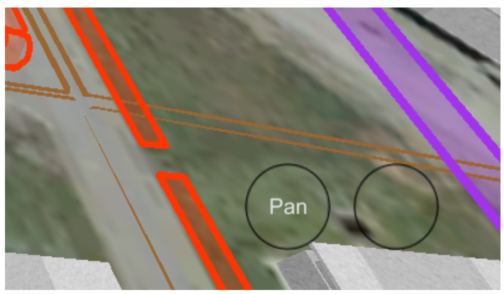


Figure 5. Camera control display as the user pans the camera with their mouse.

distance from the user's mouse/stylus to the center of the circle indicates the velocity with a further distance translating to a higher speed.

While in fly mode, if the user begins by moving the pointer upwards they would be set to fly forward and downwards would set them to fly backwards. Forward and backwards are relative to the current rotation of the camera. Subsequent movements of the pointer in the vertical direction translate to adjustments in the pitch rotation of the camera. Movements on the horizontal axis become adjustments to the yaw rotation of the camera. For the look mode, the camera remains at the same position and all movements of the cursor translate into changes of the pitch/yaw orientation with the same axis from the fly mode mapped to the same axis of rotation. Finally, the pan mode keeps the orientation fixed and moves the camera either vertically or horizontally based on the same axis of movement as the pointer. This panning motion is relative to the current orientation.

3.2. Augmented Reality Interface

For the augmented reality interface we utilize a Hololens 2 (HL2) as our hardware component. This provides built in tracking for the user, mapping their real world movements in the virtual environment. It also allows for displaying of virtual assets, and hand/joint tracking. The interface was further enhanced by Microsoft's speech recognizer set to a dictionary of commands.

The software is based on the same architecture used for the sketch interface, built in Unity 3D and adapted for a Universal Windows Platform (UWP) build so that it could be deployed to the Hololens 2 hardware. As such, the Hololens 2 was set to show up as a human agent on the network, be in communication with the PYC2 server, and receive all network traffic from the swarm and other human agents.

3.2.1. Augmented Reality Situational Awareness

To correctly utilize augmented reality situational awareness we first had to ensure proper localization of the user in the virtual world. An Android app was developed to read the GPS coordinates from a phone and transmit them to the AR application via a TCP connection. This is normally used only as a starting point as Hololens 2's moment to moment tracking is generally more accurate than commercial GPS. A speech command was added for the user to synchronize to the latest GPS coordinates when they first start the application or if they feel tracking in the virtual environment has drifted.

A set of hand gestures were added to allow a user fine-point calibration of orientation and position to reality. For orientation this is done by holding up two closed fists then performing a full body rotation. Buildings and other virtual landmarks are temporarily shown to the user so they can align them with real-world counterparts. For position calibration, if the user holds a closed left fist and an open right palm, the virtual scene will slowly move forward relative to their current orientation. As such, it is advised to calibrate orientation first before any position adjustments.

Once position and orientation are aligned, the Hololens 2's built-in tracking system can then monitor user movements and orientation changes and map them in the virtual environment. This allows virtual assets, such as buildings, vehicle markers, and drawings to maintain a position that corresponds closely with the real-world assets. There is a margin of error present from GPS error both in the initial positioning and in GPS error from devices themselves. Despite these errors operators can still utilize virtual markers to know the general location of a physical asset. Figure 6 illustrates virtual markers being near physical robots and how visibility can be maintained even through the roof and other visual obstructions.

With virtual assets overlaying real-world assets, enhanced situational awareness is now possible. Users can track small robots at a distance via their much larger virtual markers. Furthermore, they can be tracked through real-world buildings as the virtual buildings were hidden from the user's view. Grid overlays were not implemented for this interface as the data were seen as less useful from a ground view.

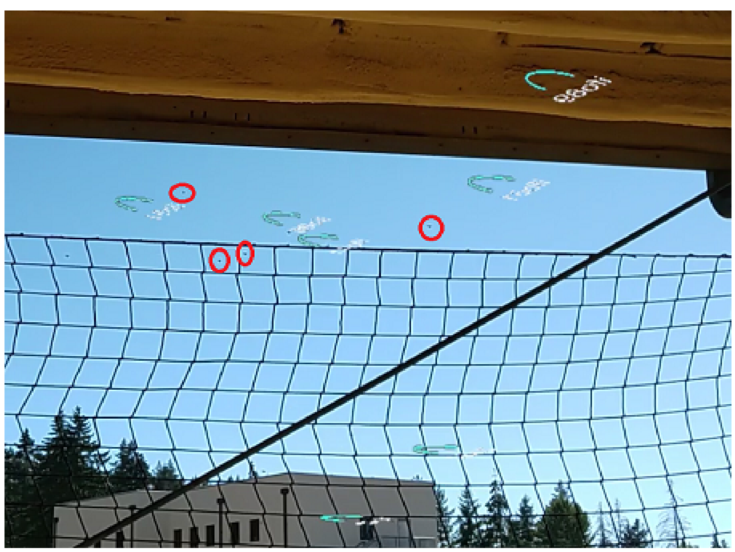


Figure 6. An example of enhanced situational awareness via augmented reality. Red circles were added in to illustrate the physical location of unmanned aerial vehicles that were reporting their GPS coordinates to the network. Photo was taken at FX4 event in Washington as live footage from FX6 event in Fort Campbell, Kentucky was not available.

(a) Sketching a tactic command directly onto a building.

(b) Pinching a building to select it. A green glow was used to confirm selection.

Figure 7. Interface for issuing tactics into virtual space. For testing purposes, buildings were shown to the user in these examples.

3.2.2. Robot Command and Control

Two techniques were incorporated to allow the issuance of tactics to robots. The set of tactics to be issued were identical to the ones provided in the sketch interface as the configuration of which gesture corresponds to which tactic command was supplied by the same centralized server.

The first technique involves sketching 3D gestures into the air. The user can hold their right index finger into the air and wait for a chime signaling the start of gesturing. The movements of their index finger are then tracked to create a series of 3D points, as shown in Figure 7a, that can be fed to the gesture classifier (Taranta et al., 2017). Gestures are ended by making a flat palm. A dehooking algorithm (LaViola, 2005) was implemented to clean up and improve recognition of the sketch data.

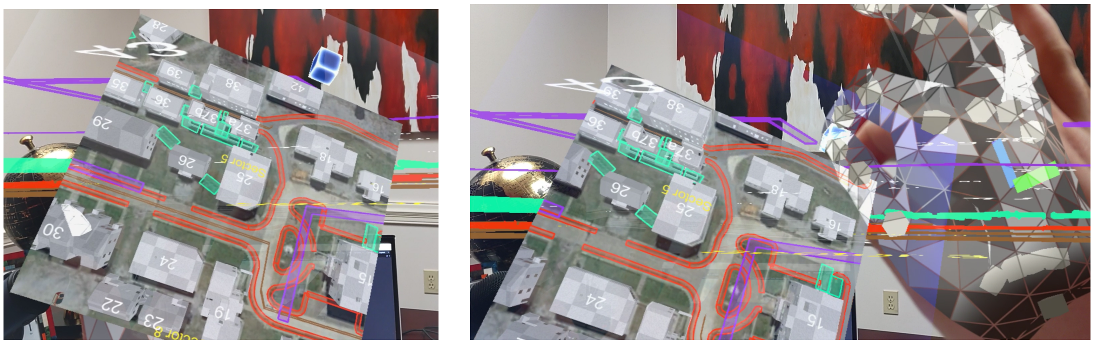
The center of the gesture is projected out into the virtual world to see if it collides with any virtual assets, such as buildings. Following the position and orientation alignment, virtual buildings will be placed near their real-world counterparts, so drawing toward a physical building will result in a collision. If the center of the gesture collides with a virtual asset its central point is then used as the location of the tactic to be executed. If no virtual assets are within the center of the gesture then the user's current position is used as the location of the tactic. Once the location is determined the tactic is executed on the network and distributed to the swarm. Unlike the sketch interface, there is no opportunity for further parameters with a keyboard. As such, default parameters are always used.

The second technique incorporates hand tracking with voice commands. Shown in Figure 7b, users can make a pinch gesture at a building to select it and a ray is cast from the pinch position toward any virtual assets. Once selected, the virtual copy of the building will be shown to the user in a green hue so they can confirm they have selected the correct building. They can then name a tactic via voice command, and the system will execute the tactic at the building's location. To deselect a building, the user only has to make a pinch gesture toward empty space where there are no virtual assets.

3.2.3. Virtual Tablet

Key capabilities of the tablet in the sketch interface were recreated in a virtual object in the Hololens 2 application. This is done by attaching a virtual plane displaying a second camera to the user's left hand as shown in Figure 8a. Not every feature of the physical tablet was available in the virtual tablet given its limitations in tactile feedback and field of view. However, key features were implemented that were seen as vital for command and control of the swarm network.

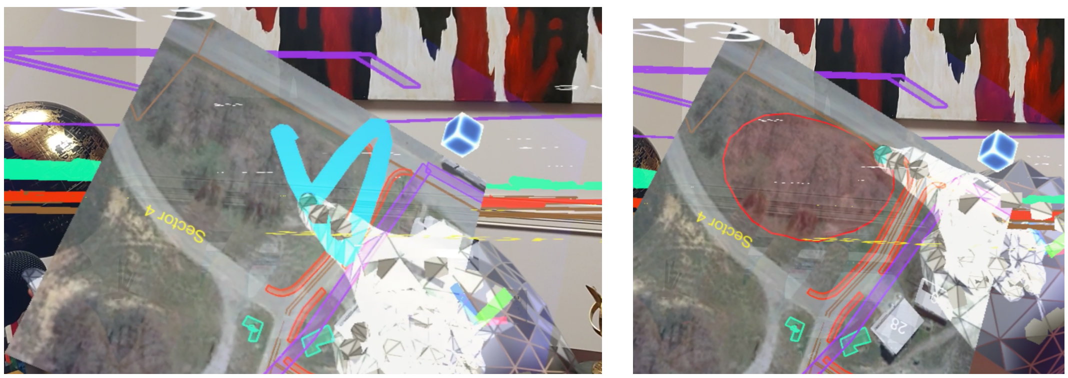
The virtual tablet's camera can be controlled via panning motion only. Full 6DOF movement was not included as the top-down view was seen as sufficiently informative and pan-only movements were less likely to lead to the camera losing view of key information in the scene. This panning motion is controlled by pinching a small cube that is floating slightly above the virtual tablet.



(a) Virtual tablet showing pre-drawn map information.

(b) Panning of virtual tablet's camera by pinching the cube.

Figure 8. Example of virtual tablet and camera control system.



(a) Drawing the no-go zone gesture.

(b) Drawing the subsequent no-go zone directly onto the virtual tablet.

Figure 9. Drawing a gesture onto the virtual tablet then drawing the subsequent information onto the tablet that would then be broadcast over the network.

While pinching the cube, translation movements made by the hand are replicated in the camera as shown in Figure 8b.

Utilizing the Hololens 2's hand tracking, the virtual tablet can detect when the user is touching it with their right hand. If the user holds their hand with a single index finger pointing out they can begin drawing a gesture which can be ended by holding out a flat palm. Gesture data are gathered by tracking the user's index finger and projecting it into the camera showing the virtual tablet, as shown in Figure 9a. By mapping the data to the same plane as the virtual tablet, it can be transformed from 3D data to 2D data that can then be fed to the same gesture recognizer (Taranta et al., 2017) as the sketch interface. For gestures that require further drawing, the user can point with their index finger again and draw on the virtual tablet, shown in Figure 9b. By performing a pinching gesture on the virtual tablet the entire gesture will be treated as complete and sent to the network for distribution. In particular, the virtual tablet was able to place key map information necessary for the swarm network such as points of interest, zone boundaries, and route information.

The AR interface's virtual tablet lacks in tactile feedback, relying on audio cues to help users identify state changes. As such gesturing data are more prone to error than with the tablet interface, particularly at the start and end points of the gesture. This is due to the user having to await visual feedback along with audio feedback to determine that the sketch has begun. Similarly, when they need to end the gesture the user has to change their hand to signal the end which can also result in error data. To resolve this, a dehooking algorithm from LaViola (2005) was implemented.

(a) Example of the scaled map.

(b) Example of the user positioning themselves to better see the simulated robots.

Figure 10. Example of the scaled map system with three simulated robots providing live updates through the swarm network.

3.2.4. Virtual Scaled Map

A virtual scaled map was created that can be summoned in front of the user via a voice command. This map contains a copy of key virtual assets that are scaled by the same value including re-creations of marker drawings. This is a situational awareness tool that allows an operator to view the entire mission as it is playing out in real time. An example can be seen in Figure 10. Once summoned the virtual scaled map will remain in the same location. It can be dismissed and brought to a new location with further voice commands. While not a part of the interface used in the field experiment it would be possible for the location of this map to be sent over the network to other operators wearing their own Hololens 2, creating a shared experience.

4. Field Experiments

Two field experiments were run with both interfaces. The first was held at Joint Base Lewis-McChord (known as FX4) and the second was at Fort Campbell (known as FX6). The simulated mission in both experiments was to assume an opposition force has seized control of a town with several high value targets present. The simulation used small visual tags known as April Tags (Figure 11) that the robots would identify and establish communication with via Bluetooth. The visual tag would inform the robot whether it has found a civilian, a high value target, an enemy, or a defensive structure.

Figure 12 shows a series of robots set up at the Fort Campbell event, but not yet deployed. Initial data of the scene including safe landing zones, power lines and boundary limits were created for both interfaces. An example of this initial information can be seen in Figure 13.

The sketch interface was used for primary command and control of the robots and was maintained on a tower overlooking the mission area and representing rear command. From here tactics could be issued to the swarm network via sketch commands and data received by the robots would be made visible on their tactical display.

For FX4, the augmented reality interface was tethered via an Ethernet cable into the system network as wireless services could not be provided to the device without interfering with the field robots. This was corrected for FX6 where the augmented reality interface was deployed into the field. From their ground position, this operator could view the virtual drawings (polyline and robot assets) in the physical world or the overhead map on their left hand via the virtual tablet.

4.1. Field Experiment 4 at Joint Base Lewis-McChord

Field Experiment 4 was conducted by the Defense Advanced Research Projects Agency (DARPA) at the Joint Base Lewis-McChord and involved the robot swarm operating within a town setting. April

Figure 11. Example of April tags used in the field experiment.

Figure 12. Robots awaiting deployment into the mission area.

Figure 13. Map of scene as seen in the Sketch interface. Colored polylines were shown to users in the augmented reality interface for enhanced situational awareness.

tag visual fiducials were used by DARPA to indicate where high value targets, civilians, and hostile targets were located. The mission goal of the robot swarm and its interfaces was to eliminate hostile threats to the robots while locating and securing high value targets. Other performance indicators to monitor were how well the swarm could obtain coverage of the area, accurately recognize April tags, and relay that information back to operators. From an interface standpoint this coverage information and how it was utilized by operators were a good indication of how the interface worked.

The experiment ran in July and August of 2020 for 14 days with four days for setup. Teams would receive one to two runs in a day with downtime for improvements in-between. Runs were two and a half hours and typical mission activities began with searching out the region to obtain maximum coverage then responding to threats and high value targets as they were found. For normal runs one operator, a male researcher, used the sketch interface and one operator, also a male researcher, utilized the augmented reality interface both in rear command. Periodically during runs military personnel and other researchers of varying ages and backgrounds were given the chance to test the interfaces. Approximately 15 personnel were given this chance and their comments were noted by the primary operators. Reconnaissance performed by the swarm was relayed back on the network to both interfaces. For the sketch interface this appeared as symbols upon the map, for the augmented reality interface this appeared as virtual symbols overlaid in the real world.

This experiment utilized up to 80 ground and aerial robots. All were online and available for tasking though only small numbers would be tasked at a time. The aerial robots were effective at providing coverage, though entering buildings and neutralizing targets with the ground robots encountered several difficulties. Wireless radios were attached to the robots, but the initial design had communication issues between the robots and operators which caused further issues during the experiment and limited the number of robots that could be tasked at a time. Furthermore battery life proved to be an issue as it introduced a time limit that a given robot could remain active during a run. Artifact discovery rate was low, nearly 20 percent, though this was explained by DARPA in debrief as an expected percentage.

The sketch interface was used as the primary means of operation during this experiment. The operators noted it as being intuitive to use and easy to task robots, though they found the communication issues caused problems that the interface could not overcome. The main issue was that information relayed back from the robots was delayed and tasking of robots was difficult if the network suddenly took them offline.

The AR interface was tethered via an Ethernet cable into the network as a wireless interface could not be provided during the experiment. Furthermore, there were concerns that the second operator appearing on the network may strain an already burdened system so the AR interface was left in a passive mode. The interface received information from the network and displayed it to the user, but did not transmit anything. This still demonstrated the situational awareness capabilities of the system which was shown to several program members and DARPA officials. An example of this enhanced SA can be seen in Figure 14. Users commented how useful it was to see assets behind buildings and to see the real-world locations of the virtual objects. The sketch interface operators also found this useful as they could communicate with the AR operator to determine the accuracy of their own drawings. In one example, the sketch interface outlined the boundaries of the game area which included a ditch. The AR interface was then used to determine how accurate the drawing was to the actual ditch.

A repeated comment was a desire to add a “Zoom In” feature for the AR interface. That is, they wished to be able to zoom in on a virtual object that they were seeing in the distance so that it could be clearer to them. While this presents technical problems beyond the scope of our work it is worth noting that multiple users from different groups requested this feature during the experiment.

4.2. Field Experiment 6 at Fort Campbell

The sixth field experiment was conducted by DARPA at Fort Campbell and contained similar mission objectives to FX4. A simulated town was used containing hostile enemies, civilians and high

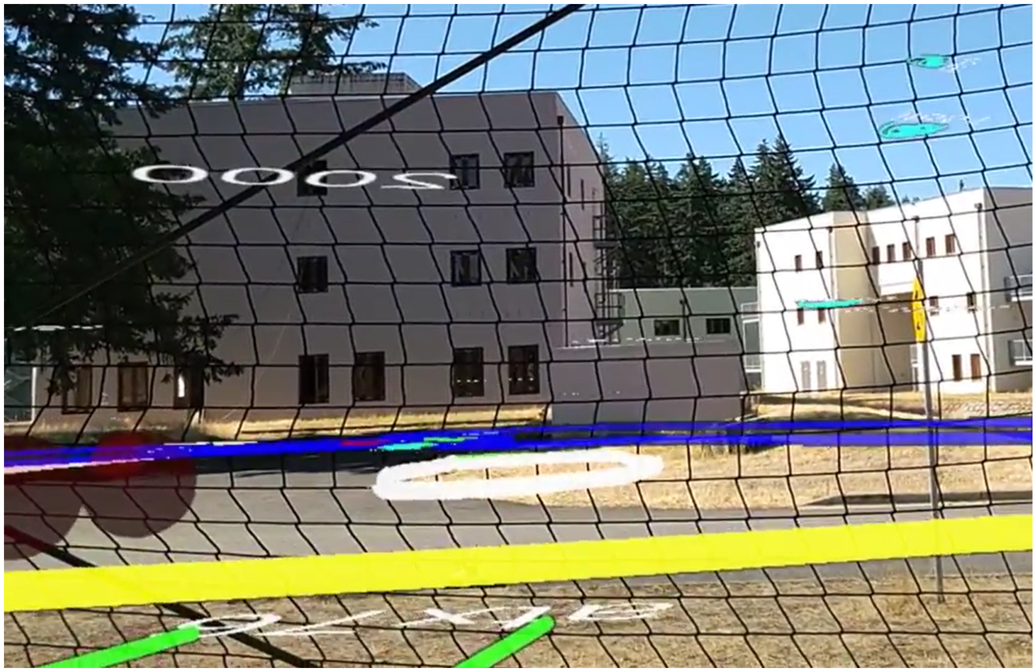
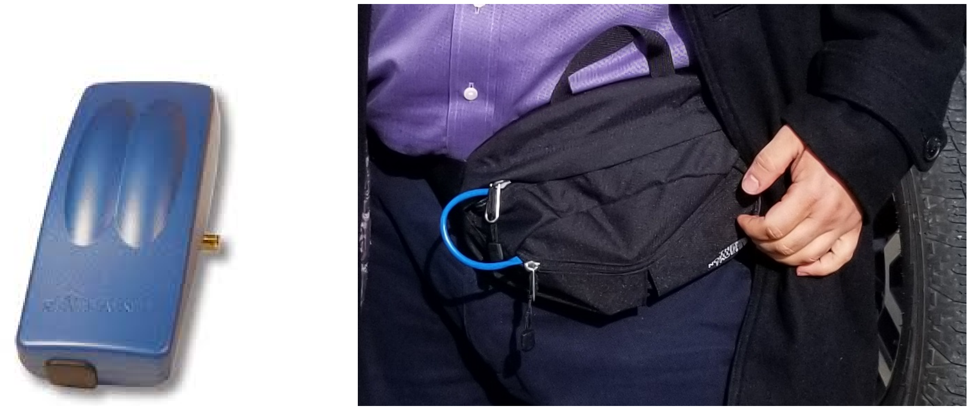


Figure 14. Situational awareness available during Field Experiment 4 in the augmented reality interface.

value targets. The robot swarm was to be used to provide situational awareness coverage, identify high value targets and neutralize them. The sketch interface was used as a part of rear command, who would task the swarm and the augmented reality interface was used for in-the-field operations. Previous wireless signal issues had been resolved allowing for more robots on the network and for the AR interface to act as an operator capable of tasking the swarm. A supply of charged batteries were also maintained to allow for longer mission runs.

The experiment ran in November 2021 for 18 days with five days for setup and included one run for each team per day. A run began at one and a half hours then quickly grew to three and a half hours over the course of the experiment. As in FX4 normal operations consisted of one operator, a male researcher, using the sketch interface as rear command and one field operator, also a male researcher, using the augmented reality interface. In this experiment the rear command was situated atop a tower with visibility over most of the simulated town while the field operator remained on the ground near the deployment zone of the robots with freedom to move through the town as needed. During demonstration periods, military personnel and other researchers were given the chance to test the interfaces. Approximately 30 personnel tested the interfaces and their comments were gathered for later presentation.

Over 200 total robots were used to maintain operations during the field experiment. In normal operations forty aerial robots and twenty ground robots would be enabled from a starting location, though during stress tests this number would approach one hundred total robots in the field. Once tactics were received by one of the interfaces to PYC2 it would begin bidding out tasks to the swarm network. As they scanned the mission area they could detect artifacts via April tags that were relayed back to rear command. This information was presented in the sketch interface as a symbol upon their map while the augmented reality interface could see the symbol in the real world and on their virtual tablet. With this reconnaissance, rear command would modify the mission plan as necessary and continue issuing tactics to the swarm to secure high value targets and neutralize or avoid hostile threats. Missions normally involved sending waves of robots out into the mission area with a large initial wave to gather as much information as possible and subsequent smaller waves to fill in gaps in information or to fulfill mission objectives. The field operator with the augmented



(a) The Rajant Dx2 Breadcrumb Radio. **(b)** Rajant radio worn by field operator during experiments.

Figure 15. Rajant wireless radios were used to maintain connectivity with the field operator in the augmented reality interface and the rest of the swarm network. This was done through a mobile ad-hoc network design.

reality interface could relay information they have gained back to rear command via handheld radio and add map information using their virtual tablet which would appear on the sketch interface's map. Artifact discovery rate and overall coverage was greatly improved with one run reporting over 600 artifacts uncovered within the first twenty minutes.

A Rajant radio network was used to maintain wireless communication between the swarm and rear command. For the augmented reality interface, a Rajant BreadCrumb DX2 was carried (Figure 15) by the field operator. The robots and human agents formed into a mobile ad-hoc network (MANET) so that communication could be maintained as long as the wireless radios were within close proximity of another wireless radio. As the swarm expanded out over the town the wireless communication mesh expanded with it, allowing more mobility for the field operator. This proved effective over the FX4 wireless system, capable of supporting a large number of robots, and only showing communication issues during stress tests.

The sketch interface performed well at relaying vital information from the mission back to rear command and accepting tactics so that the robots could be tasked to fulfill mission objectives. The capability to sketch tactics proved effective at quickly tasking robots to a specified location in the mission area. Furthermore, the operator was able to easily use the camera controls to navigate the virtual field to observe markers relayed back from the swarm network and to place tactics at exact locations. The operator was able to effectively utilize features such as lassoing robots within the swarm and chaining tactics together to carry out the tasks they had in mind and complete the mission. For example, in one case, a point of interest near the training facility needed to be investigated so nearby agents were lassoed to task them specifically. An example of chaining tactics was for setting up initial flight path for vertical take-off aerial units that were set to patrol overhead during the mission.

A test was run of launching up to a hundred unmanned vehicles simultaneously to determine the operator's capability of maintaining command and control. This both stress tested the network and the interface. The experiment was successful as the operator was able to process the information provided and effectively command the swarm to carry out required tasks.

Regarding tactics, the AR interface proved effective as operators could quickly gesture toward a building and have the robots carry out the tactic. Only tactics that would be applied to buildings were tested as other tactics involving positions would either require a virtual asset to be at that location already or for the user to walk there. Once a tactic was issued, the user was able to easily view it being carried out via monitoring the physical location and the virtual tablet.

For conveying information between the AR operator and the rear command, tests were conducted in which the AR operator would sketch no-go zones on their virtual tablet and confirm that rear command and the swarm were aware of their additions. In another test, rear command drew a route on the sketch interface, which appeared for the AR operator in the field as a virtual line drawn along the physical road. They were then told via radio that the route was added for them to walk which they did so successfully. In another example the field operator was able to see low coverage of a building via the virtual tablet and task nearby idle robots to scan it by drawing an S gesture over the building.

Small issues were observed when guest operators attempted to sketch gestures using the AR interface. The lack of tactile feedback and user's unique ways of drawing gestures proved to lower the probability of the recognizer correctly classifying the gesture. The issue was resolved when operators learned how to draw the gesture as the developers had when creating the training data, but this was not ideal for the interface's intuitiveness. The PYC2 configuration capabilities allow the system to be customized to the operator's preferred methods of drawing, but this was not utilized in the experiment.

4.3. Sketch Interface Performance

Over the two field experiments the sketch interface demonstrated its capability of easily relaying information from the swarm back to the operator at rear command and accepting tactics from them. While improvements were made between the two experiments and more customization options were available, operators commented how the sketch interface made "the gameplay feel the same." This was important as the interface was also designed to feel intuitive and easy to pick up which it maintained.

For the primary operator, drawing gestures was seen as easy as the system's gesture recognition algorithm had been trained to their form of drawing. For other user's the recognition did not work as well. As such an often heard comment was to use a button rather than a gesture. However, as one user noted "it looks easy when [the primary operator] does it" and they would like the functionality if it worked as well for them. This was ultimately the result of the nature of the demonstrations and field experiments. Temporary operators and guests were visiting for a short period of time and could not be sat down to train the algorithm. However, in normal operations this would not be an issue as time could be taken to produce training data for an operator that is loaded onto the system before the mission.

Other comments from operators during the field experiment were the following:

- "In general, the interface works great."
- "It can be hard to see the interface in the sun."
- "I struggled to remember all of the sketches. A help screen would be useful."
- "This could use some built-in checks. If I forget to draw a safe-landing zone, it should notify me of that before starting."
- "I like the interface when it works."

4.4. Augmented Reality Interface Performance

During the first field experiment the functionality of the augmented reality interface was less developed and could not be utilized as a field operator since it was tethered via an Ethernet cable. As such, operators found it to be vastly different between the two experiments with FX4 demonstrating the situational awareness and FX6 showing this functionality with the capability to interact back with the system.

During FX4 operators generally found the enhanced SA to be useful. They could monitor in real-time the status and location of drones that were not visible. For FX6, field operators had more capability and enjoyed the two map displays that were made available. They found this enhanced their situational awareness even further by being able to correlate real-world locations with the representations either in front of them on the scaled map or on their hand in the field map. In one part of the experiment, an operator who had no prior knowledge of the mission was able to quickly determine the location that rear command suspected of a high value target by monitoring drone movements in the scaled map.

Similar to the sketch interface, gestures drawn by the primary operator who the system was trained for were almost always recognized. For other operators, however, it could quickly lead to the wrong classification or no recognition at all. As such, a comment again was that people enjoyed seeing the primary operator work the gesture interface, but did not enjoy it as much themselves. Also, since the system always tracked the hands of the users, if operators were not careful with what they were doing they could trigger a command on accident. As such, operators were advised to keep their hands away from the front of their face if they were not trying to issue any command.

Other comments from operators during the field experiment were

- "This is a useful tool for situational awareness."
- "I like how the buildings are transparent. That I can see what is behind the building."
- "I can task with more context. This feels more tactical."
- "What I see applies immediately to the situation. This is better than looking at something else and trying to correlate it to reality."
- "Can I zoom in on that location?"
- "Can't see some of the icons in the sun. I have to block the visor with my hand to see clearly."
- "I drew at the wrong location. Now there's a tactic at my feet."
- "Have to be real careful when drawing."

5. Discussion and Future Work

The field experiments were seen as a successful run of the two interfaces, as both operators were able to fulfil their duties in observing information from the swarm network and relay information back to that network in the form of tactics. During the experiment, research and military personnel had the opportunity to use both interfaces and found the interfaces easy to use. Particularly they liked how easy it was to bring themselves up to speed on current mission status and monitor mission progress as commands were carried out by the swarm.

A common issue was noted of having users the system was not trained on drawing gestures. These operators found the system to be stressful and commonly asked for a button instead. As noted previously, this would not be an issue in real operations as every operator would have their training data loaded onto the system and recognition would greatly increase. Furthermore, it is easy to request a button, but the number of commands available would result in a distracting number of buttons or a complex menu system that the operators were not considering. Since users and visitors enjoyed seeing the primary operator work the interface that was trained for it, we stand by the assertion that gesturing is the correct way to intuitively contain several configurable commands.

There are still improvements to be made to the system. The field experiment showed architectural challenges to overcome, such as the bandwidth on the network and GPS issues on some of the robots. The experiment also showed improvements to be made to the sketch interface and augmented reality interface.

For the sketch interface there is future work in applying situational context to tactics that are drawn to enhance the experience. For example, when the operator wants to add a zone or boundary, they must be precise in their drawing. However, additional context analysis of the environment, the swarm network reporting, and additional situational information could be utilized to assist the operator in adding these boundaries. Furthermore, the capability of modifying boundaries or other polylines involved additional gestures to enter a modification mode which could be improved by analyzing context to determine if the user was attempting to change a boundary. Furthermore, users recommended a help menu to assist them in remembering all of the gestures available, perhaps one similar to those shown in Gesturebar (Bragdon et al., 2009). Also, while the sketch interface could be used with a stylus or a keyboard/mouse, users greatly preferred the stylus and felt the keyboard/mouse interface could be improved via multimodal input.

Several features were recommended for the augmented reality interface. First is the concept of a shared scaled map with annotation features. The idea being that any individual standing in an area, such as rear command, with a Hololens 2 device could see the same mission map in the same location. Also, that annotations could be made to this map which all people wearing the Hololens 2 would be able to see within their headsets. We also seek to expand the placement of tactics in the virtual scene so that a virtual building or other asset is not required. One concept is to draw the tactic then use hand gestures to move it forward and backward from the user. During the experiment, we noticed that the reliance on hand tracking could easily lead to users that tend to "speak with their hands" accidentally triggering various features of the system. Thresholds were adjusted, but there could be future work exploring the ideal ways to "engage" such a system and its hand tracking rather than monitoring hand movements at all times. Finally, some features of the real tablet were not replicated on the virtual tablet that could be brought in. For example, grid overlays were not as useful in the ground view, but could have been added to the virtual tablet for more situational awareness.

6. Conclusion

In this paper, we presented two viable interfaces for command and control over a large-scale swarm network in the field. One was a sketch interface which was seen as very intuitive and easy to use. It was also shown as a viable method to process the information presented back to the user with minimal cognitive load. The second interface was in augmented reality, which provided enhanced situational awareness and was shown as effective for field operations and providing information back to rear command. Both interfaces used gesturing techniques to easily send commands to the swarm and could observe as the unmanned vehicles carried out their tasks.

Two field experiments were conducted with both interfaces. During these experiments they demonstrated their effectiveness and was shown to a variety of users. Feedback was gathered from these experiments which showed their successes and their shortcomings that can be fixed in future iterations. While there were some criticisms, the overall impression was positive and showed promise in the use of these interfaces for human swarm interactions.

While there is work to be done in enhancing these interfaces, they show how it is possible for a small team to command a large-scale robot swarm to carry out mission objectives. As autonomous and semi-autonomous vehicles become more common for commercial and military use, such interfaces will be valuable in leveraging the intuition of human operators with the capabilities of swarm robotics.

Acknowledgments

This work was supported in part by DARPA under Contract N66001-17-C-4066.

ORCID

Brian M. Williamson https://orcid.org/0000-0002-4411-6633
Eugene M. Taranta II https://orcid.org/0000-0001-5615-2597
Joseph J. LaViola Jr https://orcid.org/0000-0003-1186-4130

References

Arnold, R., Carey, K., Abruzzo, B., and Korpela, C. (2019). What is a robot swarm: a definition for swarming robotics. In *2019 IEEE 10th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON)*, pages 0074–0081. IEEE.

Bashyal, S. and Venayagamoorthy, G. K. (2008). Human swarm interaction for radiation source search and localization. In *2008 IEEE Swarm Intelligence Symposium*, pages 1–8. IEEE.

Behymer, K., Rothwell, C., Ruff, H., Patzek, M., Calhoun, G., Draper, M., Douglass, S., Kingston, D., and Lange, D. (2017). Initial evaluation of the intelligent multi-uxv planner with adaptive collaborative/control technologies (impact). Technical report, Infoscitex Corp. Beavercreek.

Beni, G. (2004). From swarm intelligence to swarm robotics. In *International Workshop on Swarm Robotics*, pages 1–9. Springer.

Bonabeau, E., Theraulaz, G., and Dorigo, M. (1999). *Swarm intelligence*. Springer.

Bragdon, A., Zeleznik, R., Williamson, B., Miller, T., and LaViola Jr, J. J. (2009). Gesturebar: improving the approachability of gesture-based interfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 2269–2278.

Brambilla, M., Ferrante, E., Birattari, M., and Dorigo, M. (2013). Swarm robotics: a review from the swarm engineering perspective. *Swarm Intelligence*, 7(1):1–41.

Brown, D. S., Goodrich, M. A., Jung, S.-Y., and Kerman, S. (2016). Two invariants of human-swarm interaction. Technical report, AIR FORCE RESEARCH LAB ROME NY ROME United States.

Butzke, J., Daniilidis, K., Kushleyev, A., Lee, D. D., Likhachev, M., Phillips, C., and Phillips, M. (2012). The university of pennsylvania magic 2010 multi-robot unmanned vehicle system. *Journal of Field Robotics*, 29(5):745–761.

Cacace, J., Finzi, A., Lippiello, V., Furci, M., Mimmo, N., and Marconi, L. (2016). A control architecture for multiple drones operated via multimodal interaction in search & rescue mission. In *2016 IEEE International Symposium on Safety, Security, and Rescue Robotics (SSRR)*, pages 233–239. IEEE.

Cauchard, J. R., E, J. L., Zhai, K. Y., and Landay, J. A. (2015). Drone & me: an exploration into natural human-drone interaction. In *Proceedings of the 2015 ACM international joint conference on pervasive and ubiquitous computing*, pages 361–365.

Correa, A., Walter, M. R., Fletcher, L., Glass, J., Teller, S., and Davis, R. (2010). Multimodal interaction with an autonomous forklift. In *2010 5th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 243–250. IEEE.

Dorigo, M., Birattari, M., and Brambilla, M. (2014). Swarm robotics. *Scholarpedia*, 9(1):1463.

Edwards, S. J. A. (2005). *Swarming and the Future of Warfare*. PhD dissertation, Pardee RAND Graduate School, Santa Monica, CA.

European Commission and Directorate-General for Research and Innovation (2018). *SESAR-RETINA : delivering Aviation 4.0*. Publications Office.

Feitshans, G., Rowe, A., Davis, J., Holland, M., and Berger, L. (2008). Vigilant spirit control station (vscs): The face of counter. In *AIAA Guidance, Navigation and Control Conference and Exhibit*, page 6309.

Giusti, A., Nagi, J., Gambardella, L. M., and Di Caro, G. A. (2012). Distributed consensus for interaction between humans and mobile robot swarms. In *AAMAS*, pages 1503–1504. Citeseer.

Hansen, J. P., Alapetite, A., MacKenzie, I. S., and Møllenbach, E. (2014). The use of gaze to control drones. In *Proceedings of the symposium on eye tracking research and applications*, pages 27–34.

Hoffman, G. and Breazeal, C. (2004). Collaboration in human-robot teams. In *AIAA 1st Intelligent Systems Technical Conference*, page 6434.

Holzbach, M. (2008). Evaluation of holographic technology in tactical mission planning and execution. Technical report, ZEBRA IMAGING INC AUSTIN TX.

Kim, L. H., Drew, D. S., Domova, V., and Follmer, S. (2020). User-defined swarm robot control. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pages 1–13.

Koh, S. L., Pfeil, K., and LaViola, J. J. (2018). Exploring the potential of full body and hand gesture teleoperation of robots inside heterogeneous human-robot teams. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, volume 62, pages 2008–2012. SAGE Publications Sage CA: Los Angeles, CA.

Kolling, A., Nunnally, S., and Lewis, M. (2012). Towards human control of robot swarms. In *Proceedings of the seventh annual ACM/IEEE international conference on human-robot interaction*, pages 89–96.

Kulakov, F., Sokolov, B., Shalyto, A., and Alferov, G. (2016). Robot master slave and supervisory control with large time delays of control signals and feedback. *Applied Mathematical Sciences*, 10(36):1783–1796.

Labazanova, L., Tleugazy, A., Tsykunov, E., and Tsetserukou, D. (2018). Swarmglove: A wearable tactile device for navigation of swarm of drones in vr environment. In *International AsiaHaptics conference*, pages 304–309. Springer.

Larochelle, B., Kruijff, G.-J. M., Smets, N., Mioch, T., and Groenewegen, P. (2011). *Establishing human situation awareness using a multi-modal operator control unit in an urban search & rescue human-robot team*. IEEE.

LaViola, J. J. (2005). *Mathematical sketching: a new approach to creating and exploring dynamic illustrations*. Brown University.

LaViola, J. J. (2015). Context aware 3d gesture recognition for games and virtual reality. In *ACM SIGGRAPH 2015 Courses*, pages 1–61.

LaViola Jr, J. J., Buchanan, S., and Pittman, C. (2014). Multimodal input for perceptual user interfaces. *Interactive Displays: Natural Human-Interface Technologies*, pages 285–312.

Liu, K., Sakamoto, D., Inami, M., and Igarashi, T. (2011). Roboshop: multi-layered sketching interface for robot housework assignment and management. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 647–656.

Luo, R. C. and Chen, T. M. (2000). Development of a multi-behavior based mobile robot for remote supervisory control through the internet. *IEEE/ASME Transactions on mechatronics*, 5(4):376–385.

McLurkin, J., Smith, J., Frankel, J., Sotkowitz, D., Blau, D., and Schmidt, B. (2006). Speaking swarmish: Human-robot interface design for large swarms of autonomous mobile robots. In *AAAI spring symposium: to boldly go where no human-robot team has gone before*, pages 72–75. Palo Alto, CA.

Moon, S., Cho, D., Han, S., Rew, D., and Sim, E.-S. (2014). Development of multiple ar. drone control system for indoor aerial choreography. *Transactions of the Japan Society for Aeronautical and Space Sciences, Aerospace Technology Japan*, 12(APISAT-2013):a59–a67.

Naghsh, A. M., Gancet, J., Tanoto, A., and Roast, C. (2008). Analysis and design of human-robot swarm interaction in firefighting. In *RO-MAN 2008-The 17th IEEE International Symposium on Robot and Human Interactive Communication*, pages 255–260. IEEE.

Nagi, J., Giusti, A., Gambardella, L. M., and Di Caro, G. A. (2014). Human-swarm interaction using spatial gestures. In *2014 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 3834–3841. IEEE.

Nourmohammadi, A., Jafari, M., and Zander, T. O. (2018). A survey on unmanned aerial vehicle remote control using brain–computer interface. *IEEE Transactions on Human-Machine Systems*, 48(4):337–348.

Olson, E., Strom, J., Morton, R., Richardson, A., Ranganathan, P., Goeddel, R., Bulic, M., Crossman, J., and Marinier, B. (2012). Progress toward multi-robot reconnaissance and the magic 2010 competition. *Journal of Field Robotics*, 29(5):762–792.

Peshkova, E., Hitz, M., and Kaufmann, B. (2017). Natural interaction techniques for an unmanned aerial vehicle system. *IEEE Pervasive Computing*, 16(1):34–42.

Pfeil, K., Koh, S. L., and LaViola, J. (2013). Exploring 3d gesture metaphors for interaction with unmanned aerial vehicles. In *Proceedings of the 2013 international conference on Intelligent user interfaces*, pages 257–266.

Podevijn, G., O'Grady, R., and Dorigo, M. (2012). Self-organised feedback in human swarm interaction. In *Proceedings of the workshop on robot feedback in human-robot interaction: how to make a robot readable for a human interaction partner (Ro-Man 2012)*.

Sakamoto, D., Honda, K., Inami, M., and Igarashi, T. (2009). Sketch and run: a stroke-based interface for home robots. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 197–200.

Sarkar, A., Patel, K. A., Ram, R. G., and Capoor, G. K. (2016). Gesture control of drone using a motion controller. In *2016 International Conference on Industrial Informatics and Computer Systems (CIICS)*, pages 1–5. IEEE.

Shah, D., Schneider, J., and Campbell, M. (2010). A robust sketch interface for natural robot control. In *2010 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 4458–4463. IEEE.

Sheridan, T. B. (1992). *Telerobotics, automation, and human supervisory control*. MIT press.

Skubic, M., Anderson, D., Blisard, S., Perzanowski, D., and Schultz, A. (2007). Using a hand-drawn sketch to control a team of robots. *Autonomous Robots*, 22(4):399–410.

Taranta, E. M., Samiei, A., Maghoumi, M., Khaloo, P., Pittman, C. R., and LaViola, J. J. (2017). Jackknife: A reliable recognizer with few samples and many modalities. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pages 5850–5861.

Taylor, G., Frederiksen, R., Crossman, J., Quist, M., and Theisen, P. (2012). A multi-modal intelligent user interface for supervisory control of unmanned platforms. In *2012 International Conference on Collaboration Technologies and Systems (CTS)*, pages 117–124. IEEE.

Vasile, C., Pavel, A., and Buiu, C. (2011). Integrating human swarm interaction in a distributed robotic control system. In *2011 IEEE International Conference on Automation Science and Engineering*, pages 743–748. IEEE.

Zeleznik, R. and Forsberg, A. (1999). Unicam—2d gestural camera controls for 3d environments. In *Proceedings of the 1999 symposium on Interactive 3D graphics*, pages 169–173.

How to cite this article: Williamson, B. M., Taranta II, E. M., Moolenaar, Y. M., & LaViola Jr., J. J. (2023). Command and control of a large scale swarm using natural human interfaces. *Field Robotics, 3*, 301–322.