Creating a Mixed Reality Common Operating Picture Across C2 Echelons for Human-Autonomy Teams

Christopher Reardon[1], Mark Dennison[1], Jason Gregory[1], Theron Trout[2], and John G. Rogers[1]

[1]U.S. Army Research Laboratory, Adelphi, MD, USA
[2]Stormfish Scientific Corporation, Silver Spring, MD, USA

ABSTRACT

One of the most significant challenges for the emerging operational environment addressed by Multi-Domain Operations (MDO) is the exchange of information between personnel in operating environments. Making information available for leveraging at the appropriate echelon is essential for convergence, a key tenet of MDO. Emergent cross-reality (XR) technologies are poised to have a significant impact on the convergence of the information environment. These powerful technologies present an opportunity to not only enhance the situational awareness of individuals at the "local" tactical edge and the decision-maker at the "global" mission command (C2), but to intensely and intricately bridge the information exchanged across all echelons. Complimentarily, the increasing use of autonomy in MDO, from autonomous robotic agents in the field to decision-making assistance for C2 operations, also holds great promise for human-autonomy teaming to improve performance at all echelon levels. Traditional research examines, at most, a small subset of these problems. Here, we envision a system that sees human-robot teams operating at the local edge communicating with human-autonomy teams at the global operations level. Both teams use a mixed reality (MR) system for visualization and interaction with a common operating picture (COP) to enhance situational awareness, sensing, and communication – but with highly different purposes and considerations. By creating a system that bridges across echelons, we are able to examine these considerations to determine their impact on information shared bi-directionally, between the global (C2) and local (tactical) levels, in order to understand and improve autonomous agents teamed with humans at both levels. We present a prototype system that includes an autonomous robot operating with a human teammate sharing sensory data and action plans with, and receiving commands and intelligence information from, a tactical operations team commanding from a remote location. We examine the challenges and considerations in creating such a system, and present initial findings.

Keywords: human autonomy teaming, command and control, mixed reality

1. INTRODUCTION

The future battlefield will be complex and contested, with information from heterogeneous systems and sensors crossing multiple domains simultaneously. To achieve operational dominance, Warfighters at all echelons must have access to a Common Operating Picture (COP) that ensures the right information is available at the right time. More so, this information needs to exist in a format that is easy to understand so that critical decisions can be made quickly and with high confidence in the success of intended effects. Existing battlefield information systems used for command and control (C2) may not be adequate for such multi-domain operations (MDO). Many of these systems are purposed towards a particular domain and are stove-piped, making data sharing and interoperability difficult or impossible. As data is exchanged across domains and systems, analysts will likely become overwhelmed with information, making it difficult for commanders to execute courses of action down to units at the tactical edge. The increasing use of autonomous systems to augment squads also requires new ways of managing information and orders. To create shared understanding and maximize situational awareness between humans and autonomous agents at the same echelon as well as up the chain of command, fundamental research on information mediation and the development of new information interaction methods are necessary.

Further author information: (Send correspondence to Christopher Reardon)
Christopher Reardon: christopher.m.reardon3.civ@mail.mil
Mark Dennison: mark.s.dennison.civ@mail.mil

2. CROSS-REALITY IN MILITARY SCENARIOS

Cross-reality (XR) technologies offer numerous capabilities that may prove useful in solving the problems presented by MDO-C2. Broadly, XR refers to the range of systems that augment virtual content over reality (augmented reality) to systems that immerse users in a completely virtual environment (virtual reality). However, XR may also refer to the fact that many of these immersive systems are "crossing" the gap between virtual and physical, using mapping technology like LiDAR to enable fluid interaction between digital content and the real world. In terms of C2, such technologies afford ways of presenting and interacting with battlefield information not possible in currently fielded systems. Moreover, robotic agents provide novel sources of information to the MDO picture that may necessitate the need for a COP that crosses physical and digital realities. Figure 1 shows the relationship between changes along the XR continuum and echelon.

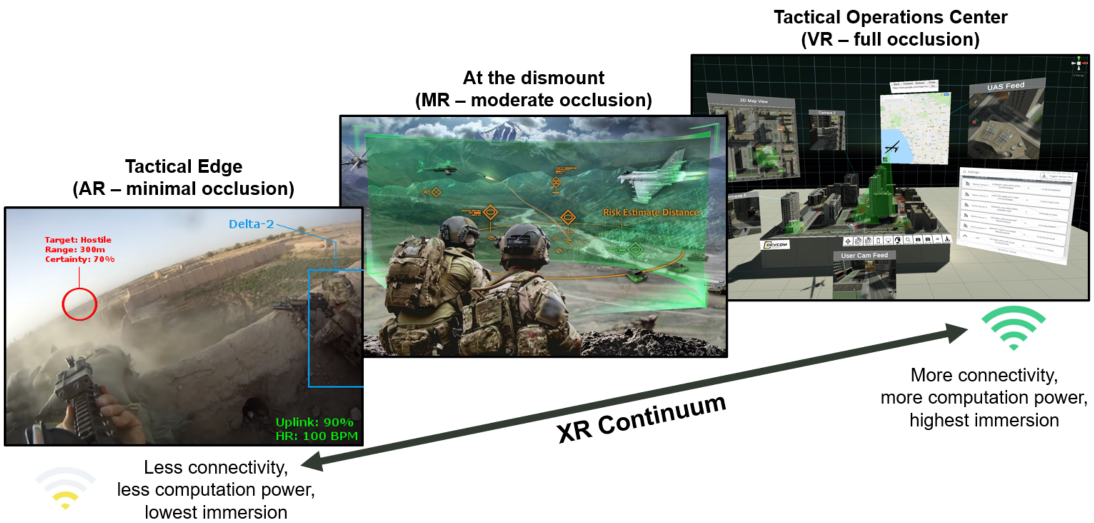

Figure 1. The cross-reality (XR) continuum depicted across military echelons from the tactical edge to the tactical operations center.

However, there is currently limited knowledge on best practices for using XR technology to augment C2 capabilities. This is due, in part, to the fact that in general best practices for using immersive technology in operational settings are largely unstudied. Recent work[1,2] surveyed research in these efforts and named this emerging field Immersive Analytics. They defined a research agenda which includes the following elements: Combining Human and Computer Intelligence, Evaluating the Utility of Immersion, Designing Immersive Analytics Systems, Facilitating Collaboration Through Immersion, and Changing the Analytical Process with Immersion. While each of these topics is critically important, for the purposes of this paper, we will focus on issues related to the design of immersive systems and how they may facilitate collaboration and shared understanding between humans and autonomous systems at the tactical edge and at the command center. Some open questions in this space are: How to create shared understanding of information between users using different display media (e.g. VR, AR, and tablet)?, What is the best way to represent 3D information on a 2D and 3D display simultaneously?, What methods of interaction enable effective manipulation of data at the tactical versus operational level?

The viability of information changes dramatically as one navigates requirements from the tactical edge to the tactical operations center (TOC). We argue that the following factors should be considered during bi-directional command and control in cross-reality: situation context, information content, timing of updates, and modality of display and interaction.

2.1 Situation Context

Situation context refers to the operational environment and the information recipient's role in that environment. Warfighters at the tactical edge are in a very different situation than commanders and analysts that are some distance from hostile ground forces. For example, consider a scenario where a platoon commander is directing squads to patrol particular routes within an urban environment. Those squads need information relevant to this context, which might include navigation way-points and operation phase lines. If a particular squad were to come into contact with the enemy, the context enters a combat situation and the relevance of particular information changes dramatically. Now, the situation dictates a shift in priority to information about the position of enemy forces, line of sight, etc. Additionally, within both of these scenarios, each Warfighter and autonomous system holds a unique role in executing the mission. Thus information that may be optimal for the objectives of a squad leader may not be optimal for each member of their team.

2.2 Information Content

Information content refers to the properties of the data for a particular piece of information. For example, a request for the location of a robot agent operating alongside a squad may consist of its grid location, its GPS coordinates, or something more colloquial like "behind the green building to our west". Each of these responses can also range in granularity or precision. For example, an autonomous system tracking squad health could query a Soldier's gear for an average heart rate report or it could ask for detailed raw data on blood oxygen, heart rate, and cognitive function. Thus, information content is intimately tied to the intended use of the data, the complexity needed or available for computation, and ease in understanding.

2.3 Timing of Updates

Timing of updates refers to the temporal frequency that changes to information occur at. In some cases, a user might prefer to be given as close to real-time updates in the state of some sensor, such as a target detection algorithm that is tracking the position of an enemy asset. However, in other cases, a commander in a TOC may only want to be alerted about a change in information when a particular event of interest has occurred. Systems and sensors might only report activities in this manner to conserve bandwidth or to prevent saturating a network. Timing may also consider the rate at which information is gathered, processed, or published by a sensor and can act as a limiting factor for operational requirements. Critically, when a human operator is involved in interacting with information, they have limited cognitive resources and are constantly at risk of becoming overloaded by balancing current and new information from the battlefield.[3]

2.4 Modality of Display and Interaction

Modality of display and interaction refers to the data visualization paradigm and associated methods for manipulating the data presented in that paradigm. The "display" must consider the realm of available human sensory systems beyond just the visual system. This means interfaces could convey information through sounds, speech, haptics and even direct biofeedback. Forces at the tactical edge need clear and concise information display to allow vigilance for threats around them. Thus, augmented reality displays may be more optimal than virtual reality as the complete visual occlusion in most VR head-mounted displays would degrade situational awareness. However, commanders and analysts who are not in direct combat can utilize virtual or mixed reality to view and interact with data in a completely immersive C2 system. Interactions with information systems are also multi-modal, and can include modalities such as key presses, gesture controls, gaze, voice commands, writing, etc.

Operators will require a new visual language that conveys information about the world around them through cross-reality displays. For example, is existing military symbology easily translated from 2D to 3D representations, and are these effective for tactical overlays in augmented reality? As cross-reality technology advances, there will be an increasing need for an intuitive and unambiguous way of ascertaining which elements in the user's view are interactable information overlays, and which are static, real-world elements.[4] However, the creation of intuitive and easy to use immersive interfaces is an open research challenge and has often been found to be scenario specific.[5–8]

3. XR-COP PROTOTYPE SYSTEM AND SCENARIO

We propose the design for a prototype system that will enable researchers to study the effects of manipulating the previously described factors of information context, content, timing, and modality on distributed-C2 of a human and autonomous agent team. The proposed scenario involves a human operator and robot team receiving orders and real-time intelligence from a remotely located human commander. The human operator will be wearing an AR system, such as the Microsoft HoloLens or IVAS*, and the human commander will be wearing a fully-occlusive head-mounted device. The robot is a small, man-portable ground robot capable of moving at human speeds and carrying a sensor and computation payload to support autonomous navigation and visual perception, such as an Endeavor Robotics PackBot. The tactical team and commander are connected over a secure but bandwidth limited and unreliable network. Figure 2 describes a high-level overview of the system and scenario.

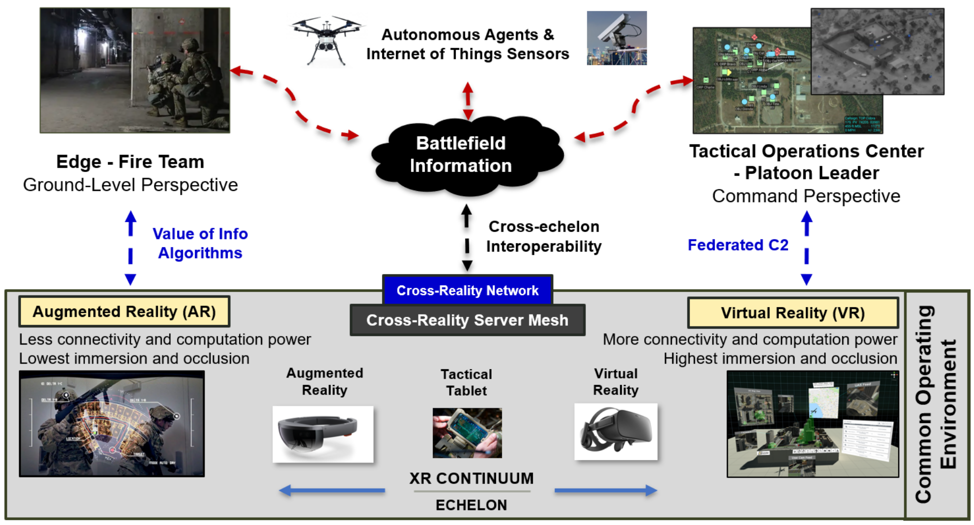

Figure 2. Information dynamics and mediation for multi-echelon communication between cross-reality systems, robot agents, and distributed command.

The mission objective is for the tactical team to navigate an urban environment to locate an asset of interest while avoiding enemy detection at all costs. The urban environment contains both neutral civilian sensors and sensors that are owned by the opposing force. These sensors may include things like cameras, motion detectors, or offensive devices such as jammers. The commander has access to a 3D model of the terrain that the tactical team will be operating in, but the information may not be current and completely accurate.

In the described scenario, we examine the factors of situation context, information content, timing of updates, and modality of display and interaction, and their potential impact on enabling effective C2 and autonomy at local and global echelons. We identify the following research questions and associated methods for exploring answers to them with respect to this scenario.

3.1 Research Questions

1. *What are the intrinsic and extrinsic differences in context between local and global?*

   First, there may be extrinsic contextual factors in the scenario that are shared across levels. At a high level for example, we would hypothesize that the current phase of MDO has a shared context across all performers at all echelons. E.g., when entering the disintegrate phase, the context of activity for all or most individuals will focus on rapidly degrading an enemy's capabilities.[9] On the other hand, there may be factors intrinsically present for an individual given his or her function. A simple example could be infantry being the natural target of IEDs. To create AI capable of interpreting context appropriately for

*The U.S. Army's Integrated Visual Augmentation System (IVAS) creating and sharing a COP across echelons, research should be performed to discover and recognize what fundamentally drives differences in context from intrinsic and extrinsic perspectives.

2. *How should content be aggregated at the squad level, at the TOC level, and across these echelons?*

   In this scenario, data will be generated from four sources: the human operator at the edge, the robot agent co-located with this human, the commander remotely located at the TOC, and the physical world where the experiment is taking place. The information generated from these data sources in its entirety is likely more than what is necessary or comprehensible by individuals at each echelon. The requirements for mission execution at each echelon must be considered to understand how information should be parsed, filtered, and aggregated accordingly. For example, when considering the robot agent, the tactical team may only need to know its current location and a semantic summary of its status, such as "All good." However, operators at the TOC may want access to more detailed sensor information, such as the entire history of the robot's location and detailed, real-time feeds from each of its sensor components. User studies must be conducted to determine information requirements and subsequent information understanding of echelon specific and cross-echelon content.

3. *How does information timing affect decision-making at the tactical and TOC levels?*

   Data from sensors on the battlefield will be moved between information systems at different rates. The speed of these updates, or how often operators will have access to updated information, will have an effect on situational awareness and decision-making, and will be limited by cognitive load. In a laboratory setting, we can precisely control these factors to understand: 1) the effects of network bandwidth limitations and latency on requests for information, 2) delays in timely information due to computational requirements (e.g., an algorithm providing a high confidence estimate slowly, or a lower confidence estimate quickly) , 3) how timing affects tactical decision making from the local perspective and C2 from the TOC perspective.

   Further, we believe there will be a strong relationship between context and timing, particularly for 3). As context changes, the frequency of information exchange may also change correspondingly. For example, the best rate of updates on an adversary's position may be directly related to one's distance from the adversary and the range of the adversary's offensive capabilities; the closer you are to their effective range, the more frequently you might want updates on their position. This is particularly the case when balanced against other burdens placed on cognitive load.

4. *How should modalities (presentation and interaction) be selected, and how should novel modalities or combinations thereof be capitalized upon?*

   Decision-making at the tactical edge and in the TOC serve different functions, and thus will likely require different modalities to both present and interact with information. Given a vast array of possible modalities for presenting and interacting with information, factors such as content, context, user preference, and cognitive load, combined modalities, should influence the selection of modalities to achieve optimal outcomes.

   In this scenario, we can compare visual, sonic, and haptic methods for conveying the position and status of friendly robotic assets and hostile assets to the tactical operator and at the TOC. We can also attempt to identify a combination of multiple modalities that can increase op-tempo and mitigate cognitive load. Conversely, we can seek to identify modalities or combinations that might create over-stimulation and increase cognitive load and confusion. It is also critical to measure the degree of shared understanding of synchronized battlefield information between the tactical team in AR and to the TOC in VR. It is important to consider that prior work has found that user preference does not always influence performance.[10,11]

   Finally, we note that as new modalities emerge, how they can be used innovatively, particularly in light of emergent MDO scenarios, to achieve e.g., more rapid convergence across domains or improved mission performance at the tactical level, should be investigated. For example, in our scenario we could compare the use of 2D vs 3D map visualizations populated by autonomous teammates.[12]

3.2 Technical Requirements

Performing this research necessitates solving a number of technical challenges. The first challenge requires creating interoperability between the information system being used at the tactical edge, the system being used at the TOC, and the autonomous robotic agent. Communication must then be established across these systems and different methods of sending and receiving data will have different capabilities in terms of bandwidth, latency, range, power consumption, etc. In some contexts, assured positioning, navigation, and timing (PNT) will not be available, so methods of synchronizing and aligning systems must be implemented.[13] Algorithms for combining disparate data frames into a common frame must also be developed, such that the information coming from the robot agent can be aligned to representations in the COP. It is important to note that these challenges have potential to impact the described factors and research questions, so trade-offs should be identified and considered. In our scenario, we would particularly identify issues of assured PNT and the need to combine data from different sources into a COP, for which our system has particular strengths.[14]

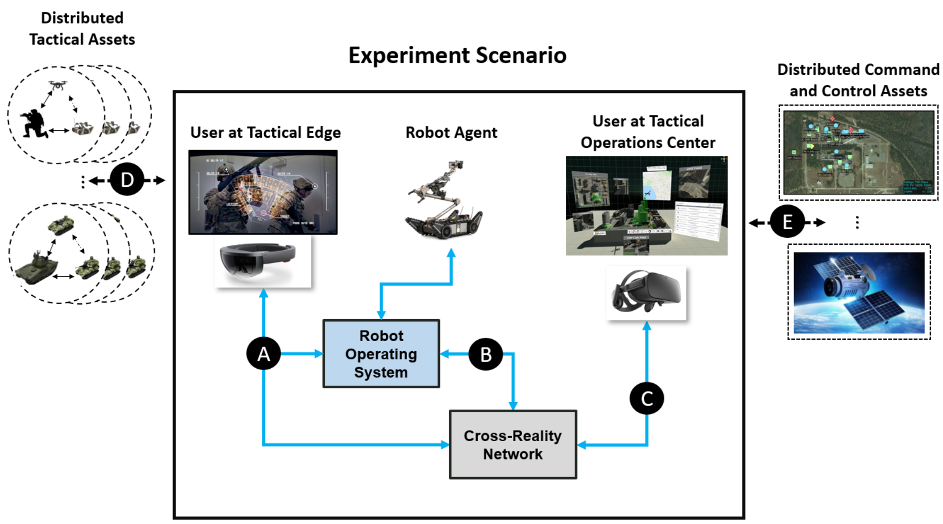

Figure 3. Diagram showing connectivity between systems in the experimental scenario and to other distributed systems in a larger battlefield context.

The experiment scenario shown in Figure 3 depicts a Tactical User wearing an AR system that is transporting information using the Robot Operating System (ROS)[†] middleware and over the XR Network (A), a Robot Agent sends information via ROS, and from ROS through the XR Network (B), and a remotely located User at the TOC wearing a VR system (C). Additionally, information is sent out to distributed tactical assets (D) and to distributed command and control assets (E). Each letter-element represents a significant technical challenge point that must be overcome to pursue the previously described research questions.

At the tactical level, the AR user requires timely information flow between the Robot Agent via ROS[15] (A) and from the TOC via the XR Network (C). Here, interoperability between systems must be solved through a common data framework or a data bridge which can covert messages not only between formats but between coordinate systems, e.g., when tactical teams use a locally-derived PNT solution, such as a 3D map generated by a Robot Agent. We envision this common data framework as a ROS-XR Network bridge (C) which will convert messages between the ROS format and a format such as Cursor on Target (CoT) which will be used in the XR-COP at the TOC. The XR-COP will send messages to the XR-Network in a proprietary format or directly in CoT format (D) which will be converted at the bridge and disseminated to the Robotic Agent via (C) or to the Tactical User via (A).

Communication links within and between echelons (A-D) and between distributed tactical (E) and C2 (F) assets will be subject to challenges such as bandwidth, power, range, and adversarial action, as noted above.

[†]https://ros.org

The XR-Network could utilize a variety of communication frameworks, such as wireless networks, radios, etc. to connect each of these systems. It could also utilize technology such as EMANE[‡] to represent different communication systems with different levels of bandwidth or packet loss.

While this is not an exhaustive or detailed list of all possible technical requirements, we believe this is the minimum set of technical issues that must be addressed to begin enabling an across-echelon XR-based COP system for human-autonomy teams.

4. CONCLUSIONS

In order to achieve the rapid and continuous integration of capabilities across domains and the information environment necessary for MDO convergence to occur, information must be optimally exchanged, presented, and acted upon across command-and-control echelons. We have presented a vision for research into enabling this critical capability using emergent cross-reality technologies and human-autonomy teaming. We have observed that the needs for information exchange, presentation, and interaction will be different as we move from the global C2 level to the tactical edge. We have identified key factors that will shape the requirements to fulfil these needs: situation context, information content, timing of updates, and modality of display and interaction. Four fundamental research questions that we believe require essential investigation are outlined, and the core set of technical challenges that must be addressed in the context of our prototype system have been enumerated. We believe that this publication lays the conceptual groundwork for using XR and AI to significantly advance MDO convergence and deliver potentially disruptive novel capabilities.

ACKNOWLEDGMENTS

The views and opinions expressed in this article are those of the authors and do not necessarily reflect the official policy or position of the United States Army or Department of Defense. The authors would like to thank Steve Russell and Steve LaRocca for their insightful discussions in the preparation of this document.

REFERENCES

[1] Marriott, K., Schreiber, F., Dwyer, T., Klein, K., Riche, N. H., Itoh, T., Stuerzlinger, W., and Thomas, B. H., [*Immersive Analytics*], vol. 11190, Springer (2018).

[2] Skarbez, R., Polys, N. F., Ogle, J. T., North, C., and Bowman, D. A., "Immersive analytics: Theory and research agenda," *Frontiers in Robotics and AI* **6**, 82 (2019).

[3] Tran, N., Mizuno, K., Grant, T., Phung, T., Hirshfield, L., and Williams, T., "Exploring mixed reality robot communication under different types of mental workload," in [*roceedings of the 3rdInternational Workshop on Virtual, Augmented, and Mixed Reality for HumanRobot Interaction (VAM-HRI'20)*], (2020).

[4] Batch, A., Cunningham, A., Cordeil, M., Elmqvist, N., Dwyer, T., Thomas, B. H., and Marriott, K., "There is no spoon: Evaluating performance, space use, and presence with expert domain users in immersive analytics," *IEEE transactions on visualization and computer graphics* **26**(1), 536–546 (2019).

[5] Raybourn, E. M., Stubblefield, W. A., Trumbo, M., Jones, A., Whetzel, J., and Fabian, N., "Information design for xr immersive environments: Challenges and opportunities," in [*International Conference on Human-Computer Interaction*], 153–164, Springer (2019).

[6] Winkler, P., Stiens, P., Rauh, N., Franke, T., and Krems, J., "How latency, action modality and display modality influence the sense of agency: a virtual reality study," *Virtual Reality* , 1–12 (2019).

[7] Santos-Torres, A., Zarraonandia, T., Díaz, P., Onorati, T., and Aedo, I., "An empirical comparison of interaction styles for map interfaces in immersive virtual environments," *Multimedia Tools and Applications* , 1–22 (2020).

[8] Thiry, K. E., Wolloko, A., Kingsley, C., Flowers, A., Bird, L., and Jenkins, M. P., "Designing federated architectures for multimodal interface design and human computer interaction in virtual environments," in [*International Conference on Human Systems Engineering and Design: Future Trends and Applications*], 404–410, Springer (2019).

[‡]EMANE is the Extendable Mobile Ad-hoc Network Emulator developed by the U.S. Naval Research Laboratory.

[9] US Army Training and Doctrine Command, "Tradoc pamphlet 525-3-1 "The US Army in Multi-Domain Operations 2028,"," *Training and Doctrine Command, Ft. Eustis, VA,(6 December 2018), viii–x* (2018).

[10] Wilkening, J., "User preferences for map-based decision making under time pressure," in [*COSIT 2009 Doctoral Colloquium Proceedings, Aber Wrac'h, France*], 91–98 (2009).

[11] Quispel, A. and Maes, A., "Would you prefer pie or cupcakes? preferences for data visualization designs of professionals and laypeople in graphic design," *Journal of Visual Languages & Computing* **25**(2), 107–116 (2014).

[12] Gregory, J., Baran, D., and III, A. W. E., "Evaluating the presentation and usability of 2D and 3D maps generated by unmanned ground vehicles," in [*Unmanned Systems Technology XV*], Karlsen, R. E., Gage, D. W., Shoemaker, C. M., and Gerhart, G. R., eds., **8741**, 135 – 144, International Society for Optics and Photonics, SPIE (2013).

[13] Reardon, C., Lee, K., Rogers, J. G., and Fink, J., "Augmented reality for human-robot teaming in field environments," in [*HCII Virtual, Augmented, and Mixed Reality (VAMR)*], 79–92, Springer (2019).

[14] Trout, T. T., Russell, S., Harrison, A., Dennison Jr, M., Spicer, R., Rosenberg, E. S., and Thomas, J., "Collaborative mixed reality (mxr) and networked decision making," in [*Next-Generation Analyst VI*], **10653**, 106530N, International Society for Optics and Photonics (2018).

[15] Gregory, J. M., Reardon, C., Lee, K., White, G., Ng, K., and Sims, C., "Enabling intuitive human-robot teaming using augmented reality and gesture control," in [*AAAI Fall Symposium on Artificial Intelligence for Human-Robot Interaction*], (2019).