Human Factors in Robots, Drones and Unmanned Systems, Vol. 57, 2022, 138–145
https://doi.org/10.54941/ahfe1002321

Extreme Reality (EXR) Telemetry Interfaces

Yang Cai[1], Haocheng Zheng[1], Mel Siegel[1], and Lenny Weiss[2]

[1]Carnegie Mellon University, Pittsburgh, PA 15213, USA
[2]University of Pittsburgh Medical Center, 3600 Forbes Ave., Iroquois Bldg. Suite 400A, Pittsburgh PA 15213, USA

ABSTRACT

Extreme Reality (EXR) is an extended augmented reality that incorporates telemetry interfaces that can be operated in live or simulated extreme environments. In this study, we explore a few architectures of EXR, including real-time multimodal first-person view video streaming from the thermal, stereo, and panoramic sources, and incorporate live IoT (Internet of Things) data into extreme reality models via MQTT (Message Queue Telemetry Transport) on an AR headset. This technology can be applied to real-time operations and simulation training for incident command posts (ICP) and first responders with head-up display (HUD) helmets in extreme environments such as flood, fire, smoke, and shooting. It can also be applied to other applications such as the telemetry interface for assisting in training low-vision drivers.

Keywords: Telemetry, Biometrics, Thermal, Thermography, Breath rate, Breath pattern, Face tracking, Face detection, Face detection, Head tracking, Helmet, Augmented reality, XR, AR, Breath rate, Breath pattern, Thermal imaging, Helmet, HADR, First responder, Emergency medicine, Pattern recognition, Object tracking, Real-time, Vital signs, Emergency medicine

INTRODUCTION

Telemetry is a growing information technology that would have a great impact on our everyday lives, from online shopping, remote work, to telemedicine. It is an especially critical component in emergency response services to enable first responders to be aware of situations about the teammates, victims, targets, and the environment. With the latest increase in both the amount and types of Internet of Things (IoT) presented to public safety entities, such as location-based sensors and video streaming content, ergonomic and practical user interfaces are needed to enable responders to interact effectively without inducing significant cognitive overload.

Augmented Reality (AR) interfaces provide information tailored to the context and space in which a user works. The near-eye display (NED) overlays information on top of the real world, avoiding an “eyes-off-the-road” problem. For first responders, enhancing on-demand perception, information acquisition, or critical decision cues could save lives. However, AR is still an emerging technology in the professional worlds of firefighting, emergency medical services (EMS), and SWAT operations. Most AR devices are designed for home or office usage. They don’t have adequate sensors to expand situation-awareness, communication, or operations in extreme conditions, such as smoke, dark, wet, or poor connectivity climates. Current AR systems also lack realistic training content, connections to massive IoT data, and real-time feedback during a training or live operations. In addition, many AR systems are proprietary and lack interoperability and scalability with other emerging systems, such as unmanned robots, such as drones, cars, and submarines.

There is an emerging trend that AR technologies are moving toward more field operational applications with more sensors and enhancements for extreme conditions. For example, the hyper-reality helmet for first responders (Cai, 2020), head-mounted LiDAR with sensor fusion for special operations [Cai, 2020a; Cai and Arunachalam, 2021), action recognition for firefighters (Hackett and Cai, 2019), wearable sensors for localization and mapping (Cai, Alber, and Hackett, 2020b), biometric data sensing such as breathing pattern detection from helmet (Cai and Vatura, 2022), remote multibody temperature detection (Hackett, et al., 2022), the haptic interface on the helmet for firefighters to communicate and navigate through the smoke-filled dark building (Hackett et al., 2020), gait recognition from a drone [8], and biologically-inspired algorithms of Instinctive Computing (Cai, 2016) and so on. Some enterprises are geared toward that direction as well. For example, C-Thru head-up display with a holographic thermal image for firefighters (Fink, 2020), Microsoft's next-generation tactic head-up display with enhanced night vision (Capaccio, 2022), and Mira's helmet-mounted AR goggle (Mira, 2022). The NIST Public Safety Communication Research Division also organized grand challenges to catalyze the next-generation AR technologies for public safety applications, including the haptic interface challenge (NIST, 2019), CHARIoT AR interface, and IoT challenge (NIST, 2020), and the indoor localization challenge (NIST, 2022).

In this study, we extend AR technology further to telemetry, and to extend the reality with live data streaming, sensor fusion, and immersive display in extreme environments. We anticipate that this technology can be applied to real-time operations and simulation training. Figure 1 illustrates the concept of Extreme Reality in the spectrum of realities. Take a photo of the railroad for an example. AR is to overlay the virtual objects on the real-world image; VR is to create a virtual world without much communication to the real world; Hyper-Realty, on the other hand, is to fuse the real-time data and overlay on top of the real world image; Extreme Reality is to further telemetry the data to peers or command posts and superimpose extreme reality scenarios for decision-making in real-time operations with "what-if" simulations and realistic, immersive training.

MULTIMODAL FPV VIDEO STREAMING

Perceiving the first-person-view (FPV) of the field is important to situation awareness and decision-making. This is especially true in incident response cases. In extreme situations such as dark, fire, and smoke, spectrum sensors such as far-infrared thermal cameras are necessary. In this study, we experimented with real-time video streaming over WiFi networks in four data

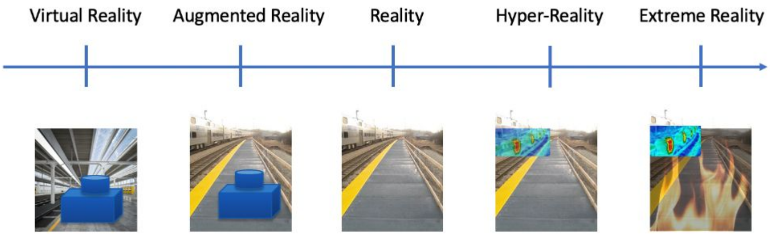

Figure 1: Extreme reality (EXR) versus hyper-reality, reality, AR, and VR.

Figure 2: The EXR helmet (left), streamed thermal video (middle), and edge overlay (right).

modalities: thermal, stereo, panoramic, and full-motion video (FMV) from a drone.

Live thermal video streaming from a helmet enables first responders and the incident command post to share the situation, coordinate, and evidence-based decision-making. The resolution of the thermal video is 160×120 pixels at 8.5 frames per second. The sensor provides thermal image detail and quality while the built-in radiometric can read temperatures up to 400 degrees Celsius. It enables the user to see-through smoke or darkness. The AR head-up display (HUD) provides a viewfinder for the user to aim to a target and see the thermal image. However, due to the low resolution of the thermal image, it is rather hard to perceive the objects and key elements in the scene. The Canny Edge detection algorithm is used to overlay the contour to the objects and background structure. This helps users to perceive the situation better. For the augmented reality design of the HUD, users can select thermal images, thermal images with edge overlay, or just edges. Edge-only display enables AR see-through effect better without obscuring the front view. However, it would lose details about the objects and cause confusion in interpretation. Experimental results show that a thermal image with a colored edge overlay is preferred.

In many teleoperations and remote technical support, we need to perceive depth from the live-streamed video. A stereo video streaming is a reasonable option. We use an off-the-shelf stereo video camera that merges two video channels into one and streams over a USB port. We stream the full-frame video at 30 frames per second over the WiFi network. From a video streaming point of view, there is no big difference between the mono channel and

Figure 3: The firefighter tested the live stereo streaming with a VR goggle at the Fire Station in Pittsburgh (left) and 3D navigation on the stereo projection screen (right).

two-channel videos, except the data traffic. The challenge is how to view the stereo videos in the real world. In this project, we provide two immersive interfaces: Virtual Reality (VR) goggles such as HTC and Google Cardboard, and stereo projection. VR google enables an immersive experience for a single user with head tracking or hand controller. Experiments show that the VR goggles show live stereo video remotely. However, it is not easy to share the view, annotate, and collaborate. Therefore, we developed an alternative to perceive the stereo video with a stereo projector. The stereo system consists of two video projectors with polarized filters. The users can use the polarized filter glasses to view the display within a normal office environment. In comparison, VR goggles are compact but weak in collaboration. Stereo projection or stereo monitors, on the other hand, are easy to share the view but not portable.

We have also developed an immersive interface to navigate through the captured 3D building data. We collected the 3D data from our office building from the parking garage to the fourth floor. We then developed our own navigation interface to walk through the floor with the First Person View (FPV). In order to move smoothly, we use the GamePad. We found that the raw 3D scan data is very noisy, especially the reflective surfaces such as windows and mirrors. We are currently working on cleaning the data, merging the multiple scans, and registering to floor plans. Furthermore, we will superimpose the real-time IoT data to the digital model.

Many teleoperations need to live stream a panoramic view of the scene in front of the driver in a moving vehicle. For teleoperation, field-of-view and transmission latency are two critical elements. For a driver's view, we typically need at least a 180-degree field of view, including the left window, front, and right window. We investigated three options: multiple-camera, 180-degree camera, and Omni-direction camera. We decided to use UDP as the transmission protocol, which was typical for video streaming. The three-camera method can generate 180 FOV without too much distortion, however, it increases the data traffic, size, complexity, and costs. Initially, we used three 720p cameras for capturing video. Each camera can produce a video stream with 1280x720p and 73° FOV. Every video stream was downsized

Table 1. Latency measurements.

	3-Cameras (combined)	3-Cameras (synced)	180° camera
Latency	600-700 ms	200-300 ms	< 100 ms

to 720x480p. Cameras at right and left were cropped to form approximately 180° with the center camera. The resulting resolution was adjusted to 1775x480p and sent out. The re-coding and processing of 3 video streams using FFmpeg brought too much pressure to the processing unit, which led to high latency.

We also tried to send 3 video streams separately and combined them at the receiving side, but the synchronization of 3 channels is susceptible to network speed and latency. Even with a timestamp, the smoothness and quality of streaming were not satisfactory. Video streams were buffered on the receiving side for 20 frames. The receiving side would check the timestamp of every frame of 3 video streams. If a frame from one stream was lost, the receiving side would use the previous frame of that stream from the buffer according to the timestamp of other successful streams. As a result, one part of the streaming video would occasionally freeze, which symbolizes a lost frame. On the other hand, 1775x480p is too long to be viewed on a typical 16:9 screen. The 180-degree camera uses fish-eye to obtain a wide-angle view. There are some optical distortions, but for many cases, it is acceptable. For example, to display the obstacles from the front, left, and right windows.

During the actual road experiment, remote viewers were able to identify the traffic light in the front and vehicles on the left and right sides. The FOV of human eyes is approximately 135°, so the 180° camera can even provide a better FOV for viewers than drivers. Besides, the part of the video that is streamed could be further optimized to exclude the panoramic roof and lower part of the dashboard so the size can be even smaller.

We don't consider the Omni-direction camera because of its significant optical distortion. It needs software to unwarp the video which might increase the processing latency. From our experiments, the transmission latencies for the three-camera and the 180-degree camera are

Our immediate application of this interface is to help the low-vision driver's training for using telescopic glasses in the vehicle. Some states require over 40 hours of in-vehicle training with a vision doctor onboard. The coach trains and tests the driver to look at the objects around the vehicle and see how the driver responds in real-time. This would be a problem during the pandemic. It would be desirable if the telemetric interface can replace the in-vehicle training. Our field tests showed that this approach is feasible. More improvements are necessary, including further reducing latency throughout cellular networks, adding a voice channel to the system, and improving the quality of the video and user interface.

EXTREME REALITY MODEL WITH LIVE IOT DATA

In this study, extreme reality models includes Incident Command Post (ICP, or Command Post) and First Responder's Head-Up Display (HUD). Both

Figure 4: The 180° camera mounted on the roof of a vehicle (left) and the remote view (right).

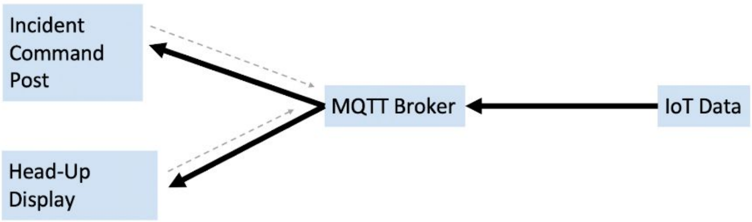

Figure 5: Streaming live IoT data to extreme reality models.

interfaces store 3D geographical models of the region of interest, from a building to a city. The digital model can be laid on a physical table, on a floor, or floating in the middle of the air. The real-time Internet-of-Things (IoT) data can be overlaid on the 3D model, creating a dynamic, interactive, and intuitive interface for communication, control, collaboration, and decision-making support. We use MQTT (Message Queue Telemetry Transport), a simple datagram protocol between devices. It is designed for connections with remote locations where resource constraints exist or the network bandwidth is limited. In our case, we only transmit text or numerical datagrams at a low frequency. In our system, the IoT data, such as flood level, speed, wildfire location, first responder's location, et al, are sent to the MQTT Broker, and then broadcasted to IPC and HUD. IPC and HUD can also send data inquiries to the MQTT broker and pull the data on-demand. Figure 5 shows the diagram of the data flow.

The visualization and interaction interfaces are implemented on AR headset Magic Leap with Unity and C#. There are a few innovative designs in the hologram of IPC: the system contains geographically accurate 3D terrain models captured from Google Earth, USGS, and OSM (Open Street Map). There is an adjustable spotlight above the model that lights up the hotspot. To avoid cognitive overflow, we display the critical IoT data on one screen and encapsulate data into groups, for example, when the operator clicks on the dots in the air, the CO_2 value will display. To accommodate the user's walk around the digital model in a physical space such as an office or a living room, the texts are oriented consistently toward the viewer. To avoid visual congestion, vertical stacks are used for co-locational data. The hologram also incorporates spectrum images such as thermal and sonar images for extreme scenarios such as wildfire, flood, and massic transit accidents.

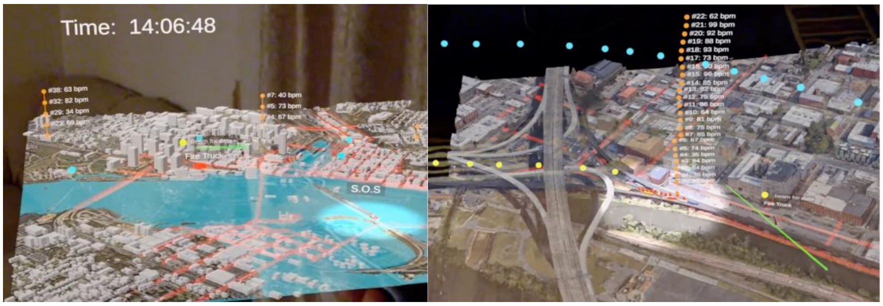

Figure 6: Flood (left) and transit accident (right) IoT data on the models of Richmond, VA.

Besides, soundscapes are used for enhancing immersive experience. The user can switch between the ICP hologram and HUD views within a scenario.

The design for the first responder's HUD includes the following innovations: spectrum images are overlaid to the real world to enable the see-through hologram, such as a sonogram overlay for revealing underwater obstacles, and thermal image overlay for searching victims in fire and smoke. Human contours are highlighted with edge detection to assist detection and identification. Some building models are overlaid to the real ones to help the user to see the interior structures before entering the building. Finally, the user can switch between the ICP and HUD views. Figure 6 shows the flood scenario and a massive transit accident scenario in the city of Richmond, VA. The models are registered with simulated live IoT data via an MQTT broker. The city model can be placed on a table, floor, or in the middle of the air. It can also be moved around and can be viewed around the model. The user can even walk into the model. At the current stage, the software has not been optimized. The model loading time needs to be reduced. Currently, we use minimal user interaction control interfaces such as the handheld control with buttons, triggers, touchpad, and a digital laser beam. In the future, we would incorporate more control interfaces such as hand gestures and tactile feedback, etc.

CONCLUSION

Extreme Reality (EXR) is an extended augmented reality that incorporates telemetry interfaces that can be operated in live or simulated extreme environments. In this study, we explore a few architectures of EXR, including real-time multimodal first-person view video streaming from the thermal, stereo, and panoramic sources, and incorporate live IoT (Internet of Things) data into extreme reality models via MQTT (Message Queue Telemetry Transport) on an AR headset. This technology can be applied to real-time operations and simulation training for incident command posts (ICP) and first responders with head-up display (HUD) helmets in extreme environments such as flood, fire, smoke, and shooting. It can also be applied to other applications such as the telemetry interface for assisting in training low-vision drivers.

ACKNOWLEDGEMENT

The study is sponsored by NIST PSCR and PSIA Programs and DoT-CMU Mobility21 Center. The author is grateful to Scott Ledgewood for his discussions and Brian Kikkola for his support on firefighter helmets and insightful comments. The authors are grateful for the support from Professor Raj Rajkumar, Stan Caldwell, Lisa Kay Schweyer, and Karen Lightman. The authors would also like to thank Charles Philips for his work on live video streaming and the firefighters from Pittsburgh Fire Station for their feedback.

REFERENCES

Cai, Y. (2020a), Hyper-reality helmet for public safety. *NIST PSCR*. Available at: https://www.nist.gov/ctl/pscr/hyper-reality-helmet-public-safety [Accessed February 15, 2022].

Cai, Y. and Arunachalam, U. (2021). Interactive floor mapping with depth sensors. *Lecture Notes in Networks and Systems*, pp. 19–25.

Cai, Y. and Vatura T. (2022) Head-Mounted Display for Remote Breath Pattern Detection, Proceedings of AHFE-2022, Taylor-Francis Group.

Cai, Y. et al. (2020b) Heads-Up LiDAR on Helmet. Proceedings of Imaging Science and Technology Conference.

Cai, Y.(2016). Instinctive Computing, Springer-London, 2016.

Cai, Y., Alber, F. and Hackett, S. (2020c). Human-Machine Sensory Fusion for Indoor Localization with a Hyper-Reality Helmet, AHFE 2020, July 16-20, San Diego.

Cai, Y., Kiefel, J. and Wu, Z. (2021) Gait Recognition from Drone Videos. In: Ahram T.Z., Falcão C.S. (eds) Advances in Usability, User Experience, Wearable and Assistive Technology. AHFE 2021. Lecture Notes in Networks and Systems, v.275. Springer, 2021.

Capaccio, T. (2022). Microsoft's $22 billion army goggles still aren't combat-ready. *Stars and Stripes*. Available at: https://www.stripes.com/branches/army/2022-01-25/microsoft-army-goggles-not-combat-ready-4412962.html [Accessed February 15, 2022].

Fink, C., (2020). Qwake technologies funded to fight fires with AR Tech. *Forbes*. Available at: https://www.forbes.com/sites/charliefink/2020/12/02/qwake-technologies-funded-to-fight-fires-with-ar-tech/?sh=4bd64283673c [Accessed February 15, 2022].

Hackett, S., Alber, and Cai, Y. (2020). Haptic Helmet for Emergency Responses in Virtual and Live Environments, in LNCS, HCII 2020, Springer-Verlag

Hackett, S., Cai, Y. and Alber, F. (2022). Remote Multi-Body Temperature Detection, Proceedings of AHFE-2022, Taylor-Francis Group

Hackett, S., Cai, Y. and Siegel, M. (2019). Sensor Fusion-Based Activity Recognition from Fireman's Helmet, ICSP-BMEI Conference, Huajiao. Won the best paper award.

Mira. https://www.mirareality.com/ [Accessed February 15, 2022]

NIST PSIAP (2019). Haptic Interfaces for Public Safety Challenge. Available at: https://www.nist.gov/ctl/pscr/open-innovation-prize-challenges/past-prize-challenges/2019-haptic-interfaces-public-safety [Accessed February 15, 2022].

NIST PSIAP (2020), Chariot Challenge: Advancing First Responder Communications. *NIST*. https://www.nist.gov/ctl/pscr/open-innovation-prize-challenges/past-prize-challenges/2020-chariot-challenge [Accessed February 15, 2022].

NIST PSIAP and University of Indiana. (2022). Indoor Tracking Challenge. https://www.nist.gov/ctl/pscr/funding-opportunities/past-funding-opportunities/psiap-fr3d-prize/indiana-first-responder. [Accessed February 15, 2022].