

Novel Perception

Ariel M. Greenberg



ABSTRACT

The Novel Perception system enhances users' awareness of the world by providing them naturalistic access to sources of information beyond what is perceivable by the basic human senses. The system, conceived of and designed by a team at the Johns Hopkins University Applied Physics Laboratory (APL), collects signals from a variety of sensors, synchronizes them in real time, registers them in real space, and then overlays them onto the real world as imagery and holograms seen through a mixed reality (MR) headset. The concept includes "virtual lenses" of hyperspectral, radio frequency, social, physiological, thermal, and radiological/nuclear overlays, so that a user can select multiple virtual lenses to create on-the-fly custom compound lenses across these modalities. Because the volume and velocity of the data streaming from these sensor modalities may be overwhelming, the system is envisioned to leverage artificial intelligence and brain-computer interfaces to sculpt the deluge per the operational tasks at hand. This article describes the approach to developing the system as well as possible applications.

INTRODUCTION

Whereas microelectronic devices excel at sensing, humans excel at understanding what is sensed. In the spirit of centaur chess,¹ in which humans and computers work together for superior performance, the Novel Perception system, designed and developed at APL, augments the human user's sight with electronic sensors so the human can sense more. The system may then train on the human's visual search patterns and reaction to a scene given these novel percepts so the system can learn how to make sense of them.

The human eye is relatively limited: most humans see color in three bands (RGB, for red, green, blue),² but mantis shrimp can see in twelve,³ and mosquitoes⁴

and snakes⁵ can see in infrared (IR). The human visual system, however, is far more powerful than those of creatures with greater spectral coverage, thanks to humans' massive occipital lobes. Humans therefore can process relatively limited spectral information to rapidly determine the salience of visual features, interpreting those features and combining all that visual information together with prior knowledge to understand the meaning of a scene.⁶ This ability, called scene understanding, is essential in all aspects of interacting with the world but is critically important in high-stakes and often quickly unfolding tactical, forensic, and social situations.

An APL team began augmenting human scene understanding when developing IN:URfACE (Investigating Non-verbals: Using xReality for the Augmented Consideration of Emotion), a system that streams spatially registered sensor data to the Microsoft HoloLens mixed reality (MR) headset. The system draws a user's attention to social signals and expressions by overlaying markers on those features in MR. IN:URfACE focuses on social scenes and detection of deception, where existing research has identified features that are relevant in these situations. For more details on IN:URfACE, see "Mixed Reality Social Prosthetic System," in this issue.

For the Novel Perception project, the team sought to explore forensic and tactical scenarios, in which the relevance of features may be more open-ended. In these scenarios, important features may vary across situations, and contextual information is crucial to identify and interpret meaning across a wide range of possible objects, environments, people, and relationships.

PHILOSOPHICAL UNDERPINNING

In philosophy of science, the *umwelt* describes the sensory world of an agent, situated in the *umgebung*, or the universe of all that can be sensed. Different crea-

tures have different *umvelts* and different footprints in the *umgebung* landscape. Although mantis shrimp have a greater number of photoreceptors than humans, their insect brains are incapable of exploiting as much from their many as we are from our fewer. Bloodhound dogs have fewer photoreceptors than humans, but superior olfaction.

The Novel Perception project seeks to compress the *umgebung* into the human *umwelt* to endow humans with intimate access to exotic sensory modalities. As shown in Figure 1, this is achieved through sensing the greater electromagnetic spectrum, processing the information, and translating the information for our visual perception in the little slice of that spectrum we can sense from red through green to blue.

THE SYSTEM

The Novel Perception unit system contains the following components (shown in Figure 2): a sensor for provision to the user; a depth camera, eye-safe lidar, or motion capture system for registration; a computer to process sensor readings and place them within a virtual environment (Unity) registered to reality; a wireless access point; and an MR headset (Microsoft HoloLens).

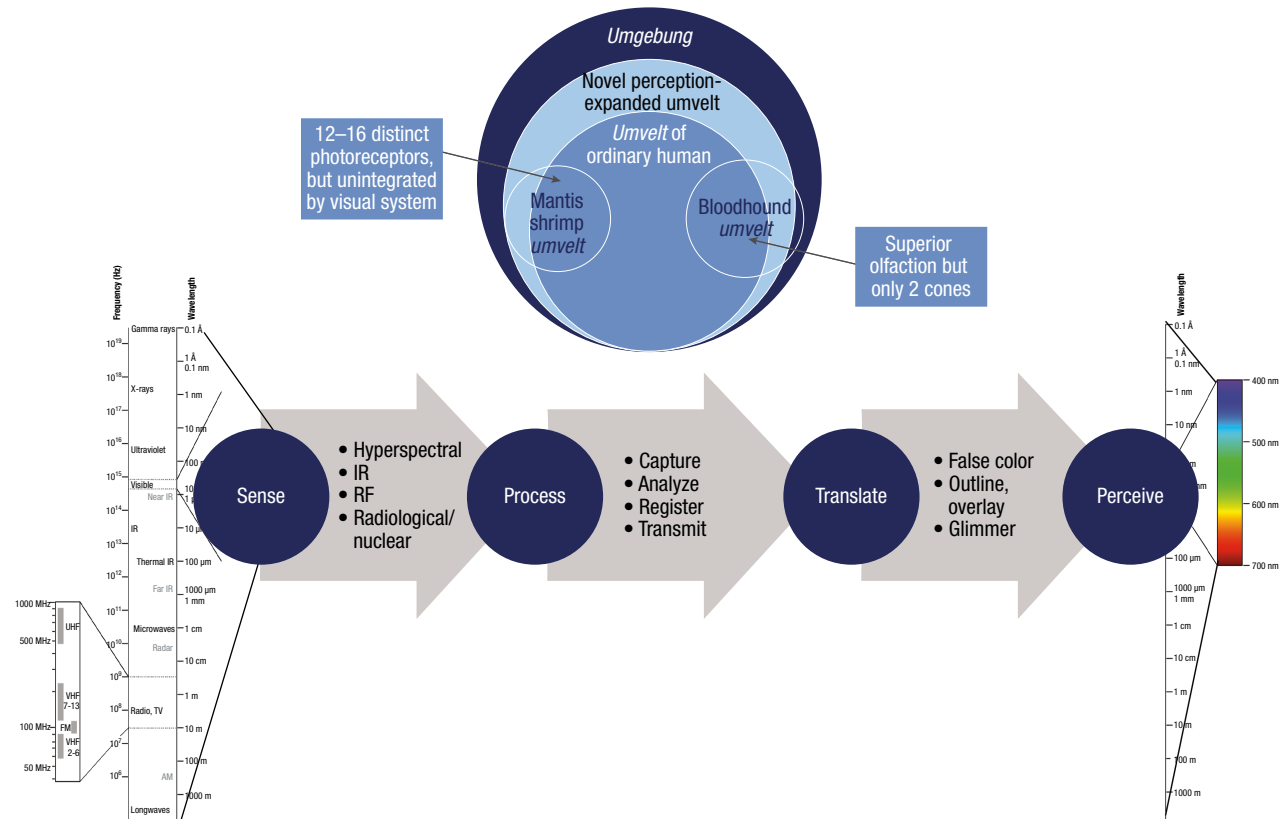


Figure 1. Philosophical underpinning and workflow of the Novel Perception project. The goal is to compress the *umgebung* into the human *umwelt* to endow humans with additional sensory modalities. This is achieved through sensing the greater electromagnetic spectrum, processing the information, and translating the information so that humans can perceive it in the RGB spectrum we can sense.

The experimental rig (shown in Figure 3) is composed of the various sensors affixed to a railing, each with motion capture markers and associated depth cameras, all visible from the motion capture camera mounted behind. This configuration allows experimentation with different

registration approaches (by depth camera, eye-safe lidar, or motion capture system). The railing also holds “the candelabra,” made up of various illumination sources ranging in spectral breadth. The railing assembly sits on a base mobile rack containing compute and networking. Processing and registering sensor data in real time is exceptionally intensive on the compute unit’s GPUs and input/output (I/O). Transmission to the MR headset is achieved with a dedicated wireless access point.

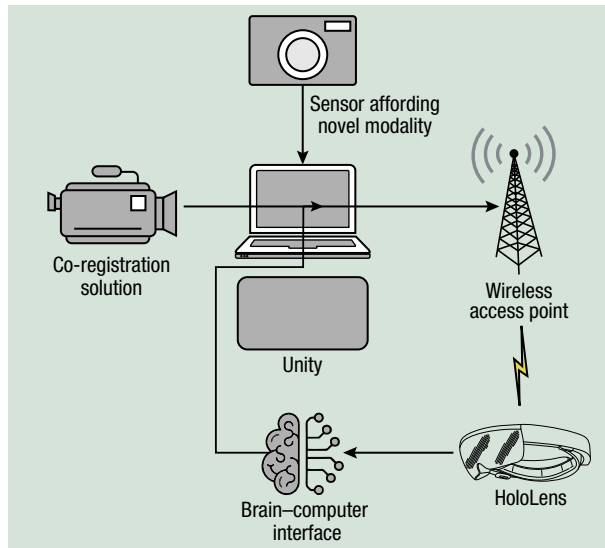


Figure 2. Novel Perception unit system diagram. The unit system includes a sensor for provision to the user; a depth camera, eye-safe lidar, or motion capture system for registration; a computer to process sensor readings and place them within a virtual environment (Unity) registered to reality; a wireless access point; and an MR headset (Microsoft HoloLens).

SENSING MODALITIES

Sensing modalities include hyperspectral, RF, social, physiological, thermal, and radiological/nuclear overlays. A user can select multiple overlays to create on-the-fly custom compound lenses across these modalities. Each modality is described in more detail below.

Hyperspectral Vision

Hyperspectral imaging (HSI) performs spectroscopy at distance. Exposed to broadband illumination, various materials reflect and absorb spectra differentially, even when equivalent in appearance to the human eye. HSI devices like Cubert’s Q285 produce data cubes equivalent to frames of dimensions $X \times Y$, with the added dimension of spectra.

The anticipated modes of operation include (1) matching to a database or index; (2) similarity and anomaly compared to other items in the scene; (3) a flip-

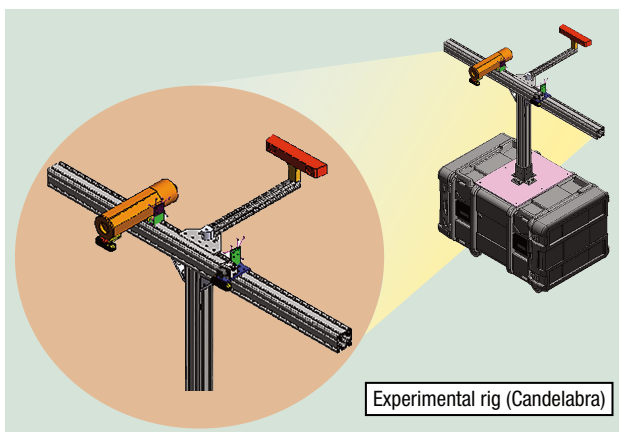
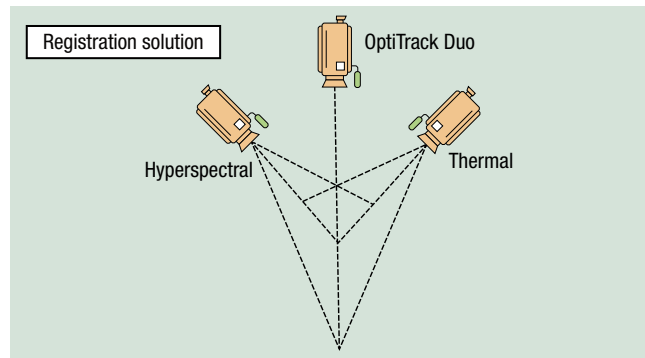
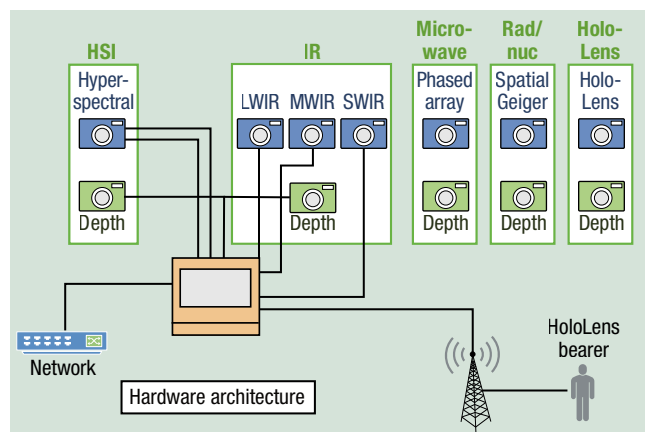


Figure 3. The experimental rig. The rig is composed of the various sensors affixed to a railing, each with motion capture markers and associated depth cameras, all visible from the motion capture camera mounted behind. This configuration allows experimentation with different registration approaches. The railing also holds “the candelabra,” made up of various illumination sources ranging in spectral breadth. The railing assembly sits on a base mobile rack containing compute and network. Transmission to the MR headset is achieved with a dedicated wireless access point.



book mode that cycles through spectral bands or material indices; and (4) a tracking mode that selects a target and identifies when it has reappeared in the scene. Demonstration of modes 1 and 3 compared natural and artificial plants. Cycling through wavelengths from 450 nm to 1000 nm with a copper-tone overlay indicating absorption revealed clear differentiation of the two materials. The bottom image at the left of Figure 4 shows the normalized difference vegetation index (NDVI),⁷ indicating the live plant in red. Demonstration of modes 2 and 4 identified all glass and all concrete in a scene and distinguished two white T-shirts laundered with different detergents (top image in Figure 4).

In many traditional applications, HSI signatures are computed offline and then presented through video playback with overlaid false color swatches indicating the presence of a signature. The Novel Perception system improves on this and similar capabilities by presenting signature information in real time through an MR headset, creating an experience that is equivalent to night vision goggles, but for hyperspectral signals. Such a naturalistic interface is intended to transform the operator's interaction with the information gleaned from hyperspectral imaging, compared with the same information provided on a computer screen.

In the concept of operations for virtual lenses as real-time indicators of HSI signatures, a user would be able to select a “lens” portraying an HSI signature algorithm. For example, to identify makeup⁸ or clothing⁹ in a scene, the lens would overlay the face paint or textile signature in false color on the wearer. The user of the MR headset would then be able to quickly orient their attention toward the falsely colored region. Combining the selection of multiple virtual lenses with set operations among lenses creates on-the-fly compound lenses. Thus, a user could select a compound lens to differentiate among individuals displaying makeup AND a particular piece

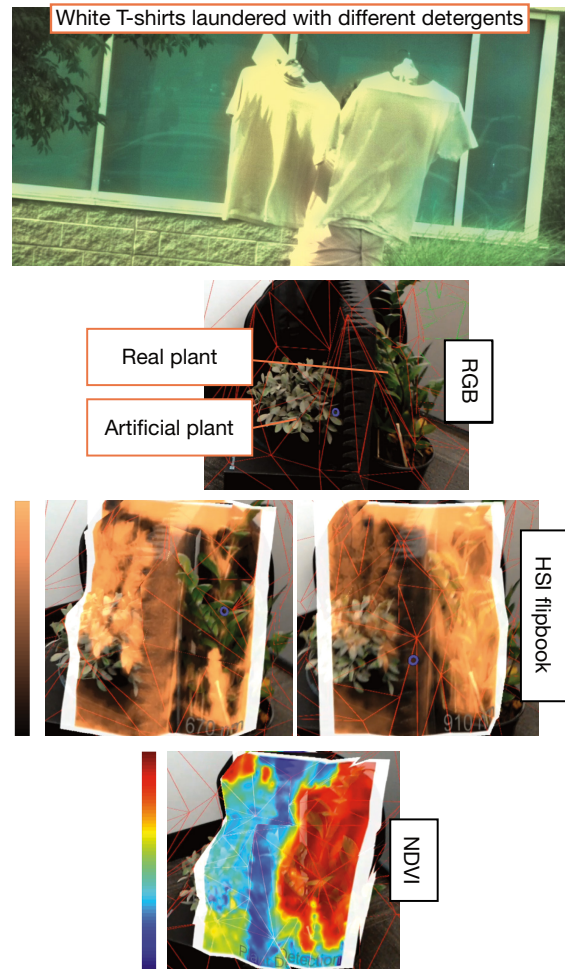


Figure 4. Hyperspectral vision. The image at the top demonstrates the ability to distinguish two white T-shirts laundered with different detergents. The other images compare natural and artificial plants. Cycling through wavelengths from 450 nm to 1000 nm with a copper-tone overlay indicating absorption shows clear differentiation of the two materials. The bottom image at the left shows the NDVI, indicating the live plant in red.

Table 1. Applications of HSI

	Range (nm)	Applications
Reflective		
Blue	400–500	Illuminates materials in shadows; penetrates water for bathymetry
Green	500–600	Penetrates water for bathymetry; discriminates oil on surface from water; identifies vegetation
Red	600–700	Penetrates water for bathymetry; differentiates vegetation
Near IR	700–1,100	Detects camouflage/netting; maps shorelines; identifies vegetation; detects watercraft on ocean; used in human-made object queuing
Emissive		
Shortwave IR	1,100–3,000	Discriminates oil from water; determines moisture content; detects plumes; discriminates camouflage/netting; detects explosions; identifies minerals
Midwave IR	3,000–5,000	Discriminates targets at night; differentiates ocean temperatures; detects smoke; identifies gases; used in thermometry
Longwave IR	5,000–14,000	Detects and identifies gases; supports thermal analysis; differentiates vegetation density and canopy cover; discriminates mineral and soil types

of clothing, allowing them to rapidly scan an area and identify that person matching both features. Other features of interest may include material differences and the presence of biofluids¹⁰ or explosives residue.

To ensure the most rapid and operationally relevant design, established signatures from related efforts in hyperspectral algorithm development would be selected to create the virtual lenses. Materials of operational utility include those identified in Table 1.

RF Vision

RF in the bands of microwave¹¹ and Wi-Fi^{12,13} has been demonstrated to reveal pose, gait, and periodic physiological phenomena even when there are obstructions to viewing the subject, like a wall for example. This capability supports urban search and rescue operations wherein victims are caught under rubble, or in a firefight when thermography is overwhelmed by heat of conflagration.

APL has developed Gunnplexers and transceivers for Doppler microwave. For proof of concept, the Novel Perception system uses simple proxies: a depth camera to perform and demonstrate visualization of pose measurement and a wearable device to perform and demonstrate visualization of heart rate measurement. In this setup, pose is transmitted and oriented to the MR headset, and heart rate is anchored to the skeleton of the user wearing the wearable device (the object of the viewer) (see Figure 5).

Social Vision

The IN:URfACE system measures head movement, facial expression, blink rate, pupillary response, and gaze direction (Figure 6; see “Mixed Reality Social Prosthetic System,” in this issue, for more details). The Novel Perception system includes the additional modality of facial thermography. Facial warming patterns correspond to affect.¹⁴ Fusing these modalities could reveal discrepancies between warming patterns and expressions (called hot spots in the literature).

Thermal Vision

Thermography, available through uncooled microbolometers sensitive to longwave infrared (LWIR), enables multipoint measurement of temperature at distance. We affixed an LWIR camera core to a HoloLens to simplify registration of the two perspectives. Of the several trials conducted, the most simple compares three covered objects: an ice pack, a hot sandwich, and a thermoplastic air bag used in packaging (see Figure 7). A more sophisticated use pulls residual heat from keypads to determine key codes from presses. The most technologically demanding application is facial thermography, used for example in the course of intelligence interviewing (for details, see “Mixed Reality Social Prosthetic System,” in this issue). We demonstrated this capability by having the viewed individual suck air through his teeth, cooling them relative to the surrounding skin.

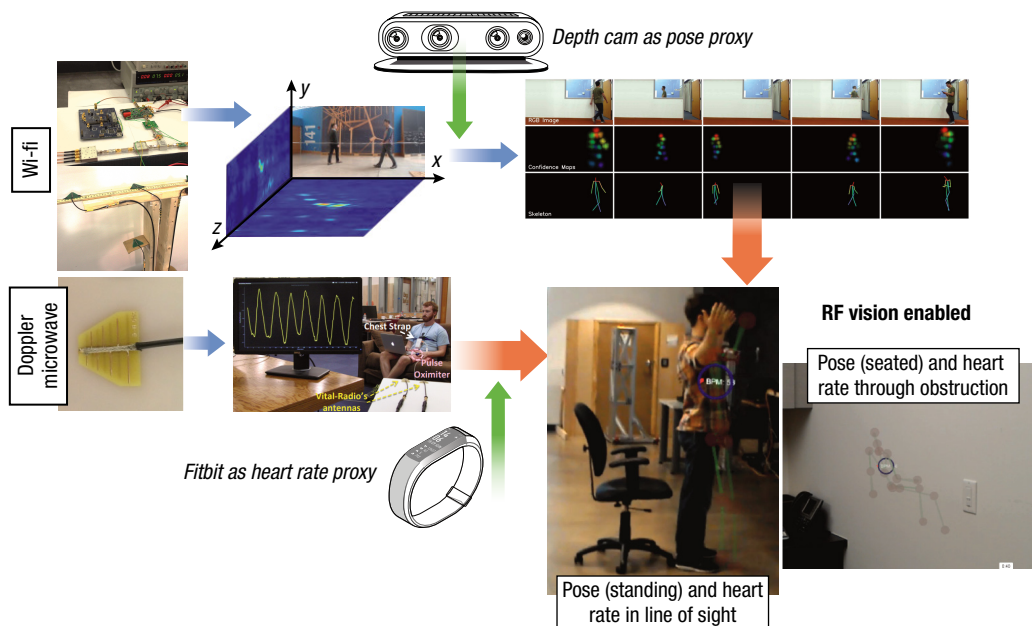


Figure 5. RF vision. RF in the microwave and Wi-Fi bands has been shown to reveal pose, gait, and periodic physiological phenomena even when there are obstructions to viewing the subject. For proof of concept, the system uses simple proxies: a depth camera to demonstrate visualization of pose measurement and a wearable device to demonstrate visualization of heart rate. In this setup, pose is transmitted and oriented to the MR headset, and heart rate is anchored to the skeleton of the user wearing the wearable device.



Figure 6. Social vision. This modality enables finer perception of nonverbal cues. The user wearing the headset is able to view data points overlaid on the interaction partner’s face, enabling the headset wearer to receive information in a way that is easily interpretable, in real time, and without interrupting the social interaction.

Radiological/Nuclear Vision

Directional radiation sensing is achieved by binding together multiple sensors and noting the differences among their readings. This sensor package is small enough to be carried in a backpack, mounted on an unmanned vehicle, or placed in a turnstile (see Figure 8). An agent equipped with the Novel Perception system would be able to visually scan the environment for indicators of radioactivity and quickly assess the location, extent, and concentration of emissions in real time as an augmented reality overlay. Such a capability is

especially useful for tracking moving targets in crowded mass transit.

INTEGRATION AND FUSION

Object recognition and semantic labeling traditionally operate on RGB data. As they are extended to multimodal sets including those sensors reviewed above, this approach to distill visual information to an annotated bounding box (see Figure 9) reduces transmission burden and enables those carrying the sensors to share

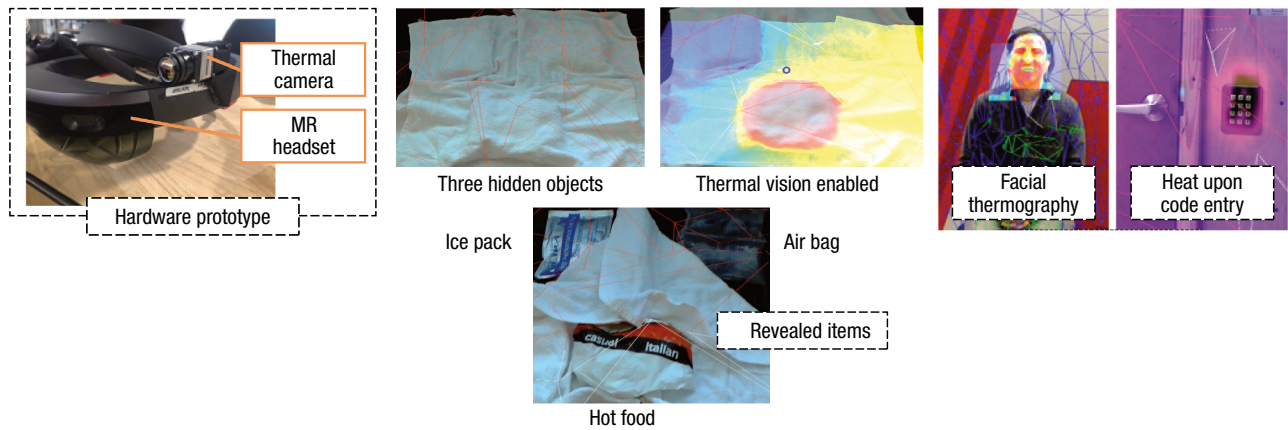


Figure 7. Thermal vision. The simplest trial compares three covered objects: an ice pack, a hot sandwich, and a thermoplastic air bag used in packaging. A more sophisticated use pulls residual heat from keypads to determine key codes from presses. The most technologically demanding application is facial thermography.



Figure 8. Directional radiation sensing is achieved by binding together multiple sensors and noting the differences among their readings. An agent equipped with the Novel Perception system, which is small enough to be carried in a backpack, mounted on an unmanned vehicle, or placed in a turnstile, would be able to visually scan the environment for hot spots indicating radioactivity and quickly assess the location, extent, and concentration of emissions in real time as an augmented reality overlay.

just the insights from them, as opposed to the full sensor readings, with the rest of the team.

CENTAUR VISION: BRAIN–COMPUTER INTERFACE OPPORTUNITIES

There are a number of potential research avenues and applications for this system. One is to explore how the human brain assimilates novel percepts, like how the conserved mammalian superior colliculus is applied to magnetoreception.¹⁵ Another research aim is to study what humans ascribe salience to when decomposing a scene

presents images at such speed that they are around the limit of conscious awareness but still induce recognition or anomaly response, it might be possible to exploit this signal and use the brain as a spectral processor to flag spiking stimuli for attention.

Combining these approaches that determine what humans attend to and how they react to what is attended to within the scene itself, with the actions taken in it, trains the system to understand a scene and the mission being conducted. The trained system is then able to assist users in identifying important features of complex, rapidly unfolding scenes, interpret those features in light of environmental and contextual information, realize how their own psychophysiological responses are influencing the scenario or their interpretation of the scenario, and ultimately understand the scene faster and more deeply.

The result is an encoding of expertise in scene decomposition that may be shared across missions and skill levels. This same means to capture the way in which experts attend to a scene may be transmitted across domains or to novices to speed learning or to recruit civilians on the fly to address an emergent purpose. This support is especially useful in scenarios that tap out humans' visual processing capabilities,

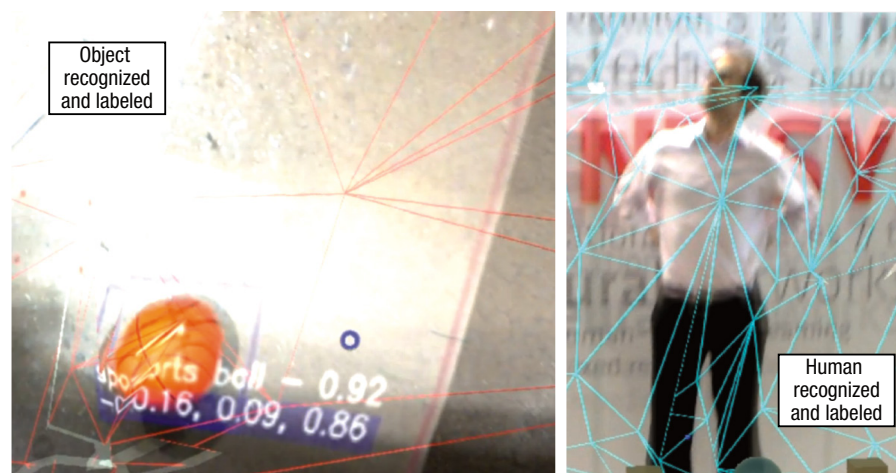


Figure 9. Integration and fusion of sensor information. Distilling visual information to an annotated bounding box reduces transmission burden and enables those carrying the sensors to easily share just the insights from them with the rest of the team.

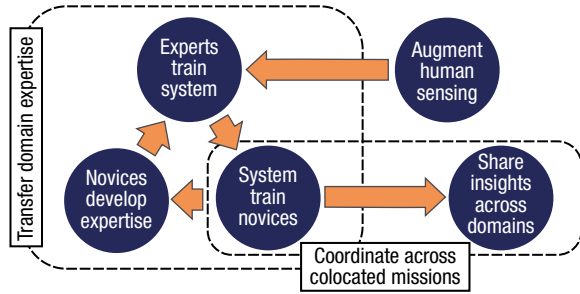


Figure 10. Novel Perception brain-computer interface opportunities. With human sensing augmented, the expert can train the system. The trained system is then able to train users to identify important features in scenes and interpret those features environmentally and contextually. This encoding of expertise in scene decomposition may then be shared across missions and skill levels.

such as during the chaos immediately following a catastrophe, or for nuanced scenarios such as those in which military peacekeepers or law enforcement seek to de-escalate (or simply understand) the situation. Figure 10 illustrates the process.

Concept of Operations for Defense Purpose

For defense and intelligence operations applications, hyperspectral technology has demonstrated utility in intelligence, surveillance, and reconnaissance and tagging, tracking, and locating.¹⁸ The centaur vision concept described in this article represents a potentially game-changing technology that could substantially speed up and democratize sensor processing in the field.

Imagine special forces operators deployed to an urban environment to handle imminent threats. They have tracked a target carrying a dirty bomb into a building (radiological/nuclear sensing detects the bomb). In that building is an individual kneeling by the window (RF sensing captures the pose), emitting a laser sight (shortwave IR sensing detects the beam). As the target's heart rate and respiration slow (detected by RF), it becomes clear that the individual is a sniper readying to take a shot. The construction material of the building (revealed by HSI) precludes use of ordinary rounds, so a local available unmanned vehicle loaded with munitions that can pierce the building exterior is directed to support.

Concept of Operations for Civilian Purpose

For homeland protection applications, the capability would support identification of individuals even when action has been taken to disguise features. Further, the system would build on the capability provided by IN:URfACE to improve ad hoc interviewing in naturalistic settings, across multiple individuals simultaneously. This has significant applications for law enforcement,

border patrol, and transportation security, among other intelligence and investigatory applications.¹⁹

Mass casualty incidents are an example application. These incidents are simultaneously social, medical, tactical, and forensic scenes. Among just some of the roles and activities involved in such an incident, first responders need to treat victims, crime scene investigators need to piece together what happened from physical evidence, detectives need to identify suspects in a crowd of victims by using behavioral evidence, and law enforcement officers need to track down and apprehend the perpetrators. If insights from expertise in all these domains could be transferred to all of the actors responding to the situation, a first responder might take the path to a victim that least disrupts forensic evidence and a crime scene investigator could converge on scene while the trail is fresh to direct law enforcement officers.

CONCLUSION

As compared to unintegrated, computer-mediated, post hoc provision of signatures, human vision augmented through MR is anticipated to dramatically speed response time, increase bandwidth, improve situational awareness, and enable discovery of previously unknown features of interest. Further, combining features from various modalities in a single display gives humans the opportunity to make new connections among the features they observe, improving their mission performance.

ACKNOWLEDGMENTS: Special thanks to the team, including Bruce Swett (now at Northrop Grumman), Kevin Chiou, Paul Finkel, Andrew Badger, Seth Billings, and Charlie Schuman. APL's Research and Exploratory Development Mission Area funded this experiment as part of its FY18 Strategic Independent Research and Development cycle.

REFERENCES

- ¹M. Cassidy, "Centaur chess shows power of teaming human and machine," *Huffington Post*, Dec. 30, 2014 (updated Dec. 6, 2017), https://www.huffpost.com/entry/centaur-chess-shows-power_b_6383606.
- ²G. Wyszecki and W. S. Stiles, *Color Science: Concepts and Methods, Quantitative Data and Formulae*, 2nd ed., New York: John Wiley & Sons, 1982.
- ³H. H. Thoen, M. J. How, T.-H. Chiou, J. Marshall, "A different form of color vision in mantis shrimp," *Science*, vol. 343, no. 6169, pp. 411–413, <https://doi.org/10.1126/science.1245824>.
- ⁴F. van Breugel, J. Riffell, A. Fairhall, and M. H. Dickinson, "Mosquitoes use vision to associate odor plumes with thermal targets," *Curr. Biol.*, vol. 25, no. 16, pp. 2123–2129, 2015, <https://doi.org/10.1016/j.cub.2015.06.046>.
- ⁵J. Fang, "Snake infrared detection unraveled," *Nature*, 2010, <https://doi.org/10.1038/news.2010.122>.
- ⁶G. J. Zelinsky, "Understanding scene understanding," *Front. Psychol.*, vol. 4, 2013, <https://doi.org/10.3389/fpsyg.2013.00954>.
- ⁷D. Haboudane, J. R. Miller, E. Pattey, P. J. Zarco-Tejada, and I. B. Strachan, "Hyperspectral vegetation indices and novel algorithms for predicting green LAI of crop canopies: Modeling and validation in the context of precision agriculture," *Remote Sens. Environ.*, vol. 90, no. 3, pp. 337–352, 2004, <https://doi.org/10.1016/j.rse.2003.12.013>.

- ⁸T. Nagaoka and Y. Kimura, "Quantitative cosmetic evaluation of long-lasting foundation using multispectral imaging," *Skin Res. Technol.*, vol. 25, no. 3, pp. 318–324, 2019, <https://doi.org/10.1111/srt.12651>.
- ⁹Y. Yan, W. Hua, Y. Zhang, Z. Cui, X. Wu, and X. Liu, "Hyperspectral camouflage target characteristic analysis," in *Proc. SPIE 10843, 9th International Symposium on Advanced Optical Manufacturing and Testing Technologies: Optoelectronic Materials and Devices for Sensing and Imaging*, 2019, pp. 108430G-1–108430G-6, <https://doi.org/10.1117/12.2505144>.
- ¹⁰"Hyperspectral imaging provides detection and high contrast imaging of biological stains," Philadelphia, ChemImage. White Paper, 2015. <https://cdn2.hubspot.net/hubfs/2753636/Life%20Sciences/biological-stains-white-paper.pdf>.
- ¹¹E. Landau, "FINDER search and rescue technology helped save lives in Nepal," NASA website, May 7, 2015, <https://www.nasa.gov/jpl/finder-search-and-rescue-technology-helped-save-lives-in-nepal>.
- ¹²M. Zhao, T. Li, M. Abu Alsheikh, Y. Tian, H. Zhao, et al., "Through-wall human pose estimation using radio signals," in *Proc. Comput. Vision Pattern Recognit. (CVPR)*, 2018, pp. 7356–7365, https://openaccess.thecvf.com/content_cvpr_2018/papers/Zhao_Through-Wall_Human_Pose_CVPR_2018_paper.pdf.
- ¹³F. Adib, H. Mao, Z. Kabelac, D. Katabi, and R. C. Miller, "Smart homes that monitor breathing and heart rate," in *Proc. CHI 2015*, Seoul, South Korea, Apr. 2015, <http://witrack.csail.mit.edu/vitalradio/>.
- ¹⁴S. Ioannou, P. Morris, H. Mercer, M. Baker, V. Gallese, and V. Reddy, "Proximity and gaze influences facial temperature: A thermal infrared imaging study" *Front. Psychol.*, vol. 5, article 845, pp. 178–189, 2014, <https://doi.org/10.3389/fpsyg.2014.00845>.
- ¹⁵P. Němec, J. Altmann, S. Marhold, H. Burda, and H. H. A. Oelschläger, "Neuroanatomy of magnetoreception: The superior colliculus involved in magnetic orientation in a mammal," *Science*, vol. 294, no. 5541, pp. 366–368, 2001, <https://doi.org/10.1126/science.1063351>.
- ¹⁶Y. Sugano and A. Bulling, "Seeing with humans: Gaze-assisted neural image captioning," 2016, arXiv:1608.05203.
- ¹⁷L. A. Farwell and E. Donchin, "Talking off the top of your head: Toward a mental prosthesis utilizing event-related brain potentials," *Electroencephal. Clin. Neurophysiol.*, vol. 70, no. 6, pp. 510–523, 1988, [https://doi.org/10.1016/0013-4694\(88\)90149-6](https://doi.org/10.1016/0013-4694(88)90149-6).
- ¹⁸M. Shimoni, R. Haelterman, and C. Perneel, "Hyperspectral imaging for military and security applications: Combining myriad processing and sensing techniques," *IEEE Geosci. Remote Sens. Mag.*, vol. 7, no. 2, pp. 101–117, 2019, <https://doi.org/10.1109/MGRS.2019.2902525>.
- ¹⁹J. Dawson and D. McLeod-Henning, "Hyperspectral imaging and the search for humans, dead or alive," Washington, DC, National Institute of Justice, <https://nij.ojp.gov/topics/articles/hyperspectral-imaging-and-search-humans-dead-or-alive>.



Ariel M. Greenberg, Research and Exploratory Development Department, Johns Hopkins University Applied Physics Laboratory, Laurel, MD

Ariel M. Greenberg is a senior research scientist in APL's Intelligent Systems Center; assistant editor-in-chief of the *Johns Hopkins APL Technical Digest*; steering committee co-chair of the International Conference on Social Computing, Behavioral-Cultural Modeling & Prediction and Behavior Representation in Modeling and Simulation (<http://sbc-brims.org/>); and an instructor in APL's Strategic Education Program. He received degrees in biology and electrical engineering from University of Maryland, College Park. In the spirit of cells-to-societies transdisciplinary research, Ariel's work addresses topics in psychophysiology and computational social science, toward applications in defense, intelligence, and health. His e-mail address is ariel.greenberg@jhuapl.edu.