https://doi.org/10.2352/EI.2026.38.10.HVEI-219

Utility of Color Symbology in Optical See-Through Head-Mounted Displays

Austin Erickson[1], Eric Seemiller[1], and Marc Winterbottom[2]
[1] KBR; Beavercreek, Ohio, United States;
[2] 711th Human Performance Wing; Wright-Patterson Air Force Base, Ohio, United States

Abstract

Imagery from optical see-through (OST) head-mounted displays (HMDs) is perceived as a blending of light emitted by the display added to the light from the user's physical environment, which can result in color distortions and desaturation of the virtual imagery. Due to these limitations, the user's ability to distinguish between colors shown on the display may be reduced compared to more traditional types of displays, which may impact the interpretation of the symbology, and potentially reduce performance. Further, individual variation in color perception may also impact the utility of color symbology in OST HMDs. In this paper, we present a user study that investigates the utility of color-coded symbology displayed on an OST Augmented Reality (AR) display within a flight simulator. We compare performance between participants with normal color vision and participants with color vision deficiencies in a dynamic flight simulator and investigate effects of symbology contrast and symbology color set on participant response times, accuracy, and eye behavior. Our results suggest that for the color sets tested, increasing the size of the set beyond a monochrome color results in reduced performance for both color normal and color deficient subjects. It's possible that custom color sets specific to OST displays are needed to achieve performance benefits.

Introduction

Modern optical see-through (OST) augmented reality (AR) displays allow the user to see virtual imagery superimposed over their view of their physical environment [1]. The main benefit of these displays is that the user gains access to additional context-relevant information within their field of view, while retaining a zero-latency view of their physical environment. For the specific application of aviation, these displays allow pilots to keep more information within their field of view at a time, potentially improving situational awareness and flight performance [2, 3].

OST displays are additive, where light emitted by the display is added to the light from within the user's environment and the virtual imagery tends to appear transparent and the perceived color shifts towards that of the user's environment [4, 5, 6]. Accurate color perception is important when interpreting color-coded information, and color blending in the OST display may cause users to misinterpret symbology or experience difficulty distinguishing between similar colors [5]. On top of this, color deficient users may experience additional difficulties in the use of these displays, limiting the accessibility of the technology.

Many of the existing OST displays employed in aviation contexts are monochrome displays limited to a single color channel, typically green. For such displays, visual cues such as shape,

Figure 1: A photo over a participants shoulder in the simulator. Five out-the-window (OTW) displays form a curve in front of the user, accompanied by a 27-inch monitor configured as a cockpit head-down display (HDD), and the AR HMD worn by the user.

brightness, and size must be employed to keep symbology distinguishable and interpretable. Hue has been shown to be particularly effective in information encoding, allowing users to search in parallel for a particular color, resulting in improved search performance compared to other visual cues [7, 8]. Blundell et al. demonstrated that adding additional color channels to simulated heads-up display (HUD) symbology can aid pilot performance in a flight simulation [9]. However, it remains to be seen if these benefits transfer when the appearance of the symbology is affected by the additive nature of an OST display, where the color of the symbology, the brightness of the user's environment, and the particular display in use each affect the contrast of the symbology, which limits the effectiveness of hue as an aid in visual search [10, 11, 6].

In this paper, we investigate the specific context of OST helmet-mounted displays (HMDs) for pilot use in aviation, and task participants to use color-coded symbology on an AR HMD to locate and select a target aircraft in the virtual scene. We investigate performance between three color sets with increasing number of colors in the set and across two color intensity levels. Additionally, we expand upon the literature by comparing performance between color normal users and users classified as color deficient by their results on a Konan cone contrast threshold (CCT) HD test (Konan Medical, Irvine, California).

Our results suggest that increasing the color set size in use on the OST HMD has significant detrimental effects on user performance, resulting in longer participant response times across both participant groups. These results highlight the need for careful selection of colors when conveying information on OST HMDs, and suggest that monochrome displays should be used until a suitable custom color set validated for use with OST HMDs is found.

Methods

We recruited a total of 16 participants (12 male, 4 female), ages 25-64 (mean 40, standard deviation 11). All participants completed the Konan Cone Contrast Threshold (CCT) HD color test, and were split into two groups: color normal and color deficient. The Konan CCT HD tests each cone type individually and provides output in the form of scores, contrast thresholds, and contrast sensitivity data for each cone type. Participants achieving contrast thresholds of less than 0.01 for the red, green, and blue subtest sections were classified into the color normal group, whereas participants scoring higher than a 0.01 on any of the three subtests were classified into the color deficient group. The color normal group consisted of 9 participants (5 male, 4 female), ages 29-49 (mean 38, standard deviation 6. The color deficient group consisted of 7 male participants, ages 25-64 (mean = 42, standard deviation = 16), five of which were found to be protanomolous and two of which were deuteranomalous. Prior to participant recruitment, our study design was approved by our organization's institutional review board (IRB). Participants provided written consent to participate after being provided with a detailed explanation of study procedures and protocols.

Aparatus

Our experiment was conducted on a research platform that synchronized imagery between an OST AR HMD conveying flight and targeting symbology, a 0.69 meter monitor configured as a cockpit head-down display (HDD) conveying a tactical situation display of other aircraft in the scene, and a set of five 2.08 meter OLED displays conveying out-the-window (OTW) imagery of the virtual environment (see figure 1).

The Vision Products SA62/S was used as the OST AR HMD (Vision Products LLC, Campbell, California). This device features two micro-OLED displays at a resolution of 1920x1200 per eye, a field of view of 55x33 degrees per eye. An HTC VIVE motion tracker was fixed to the HMD, which was used with two VIVE lighthouse units to provide positional and rotational head tracking during the simulation.

A Logitech x56 hands-on throttle and stick (HOTAS) controller was used to gather inputs allowing the user to control their aircraft and to select the designated target for each trial.

Study Design

Our study design consisted of a mixed factors design, with a between-subjects factor comparing participants grouped by score (pass/fail) on the Konan CCT HD, and 2 within-subjects factors: **color set (3 levels)** and **color intensity (2 levels)**.

The **color set** variable controlled the set of colors used to depict symbology on the AR HMD, and to a lesser extent the symbology shown on the tactical situation display of the HDD. This variable was varied between three different sets with an increasing number of colors in the set:

1. **Monochrome:** Green
2. **Primary:** Red, Green, and Blue.
3. **Full:** Red, Green, Blue, Cyan, Yellow, Magenta, and White.

During the simulation, the target plane was overlayed with a diamond-shaped symbol in a randomly-chosen color from the color set. Each distractor plane was depicted in a randomly chosen color from the remaining colors in the color set. In the monochrome condition, the user distinguished the symbology associated with the target aircraft by its unique diamond shape, whereas in the "primary" and "full" color sets, the target also had a unique color.

The **color intensity** variable was varied between two levels to vary the luminance and therefore the contrast and opacity of the symbology shown on the OST AR HMD:

1. **Full:** All symbology was rendered at full color intensity in the RGB color space (e.g. Red is [1 0 0]).
2. **Reduced:** All symbology was rendered at 0.75 color intensity in the RGB color space. (e.g. Red is [0.75 0 0]).

Operational Task

Each trial in the simulation started with the user's aircraft at an altitude of 1400 feet over an island. The HDD depicted a radar screen that showed the user's aircraft in the center surrounded by 10 distractor planes and one target plane. In each condition, the user was tasked to check the color of their own aircraft from the center of the screen and located the target aircraft on the HDD with the same color (see figure 2).

Target and distractor aircraft were randomly positioned to be within 100 meters in altitude, between 400 and 1000 meters away, and randomly within an angle of 50 degrees of the user's starting orientation. An angular dead zone of five degrees was implemented to prevent aircraft from spawning immediately ahead of the starting position. Each aircraft was randomly chosen with equal probability to be either a F35 fighter jet or a KC46 tanker plane. Upon starting a trial each of these aircraft moved radially outward from the starting position at a constant speed.

Upon locating the target aircraft on the HDD, participants used the joystick to bank toward the heading of the target (as shown in figure 3). Each distractor aircraft in the scene had either a virtual square or rectangular symbol superimposed over its position via the HMD in a randomly chosen color from the color set excluding the target color. The target had a unique diamond-shaped symbol superimposed over it via the HMD, which was depicted in green for the "monochrome" condition and otherwise in the target color for the "primary" and "full" conditions.

Once the target was located on the HMD using its shape and/or color, participants used the directional buttons on the joystick to position a cursor displayed on the HMD over the target aircraft and pulled the joystick trigger to select it. Upon selection, participants indicated the type of aircraft by pressing left/right using the hat buttons (left for F35 and right for KC46). Upon designation, a text-based label in the target color appeared above the selected aircraft via the HMD (see figure 3). This subtask was included to force users to switch attention between the HMD imagery and the OTW imagery.

Upon selection, the altitude of the target aircraft also appeared below the aircraft on the HMD in the target color (see figure 3). Users compared their altitude to that of the target aircraft. If they were within 50 units, they pressed down on the hat buttons

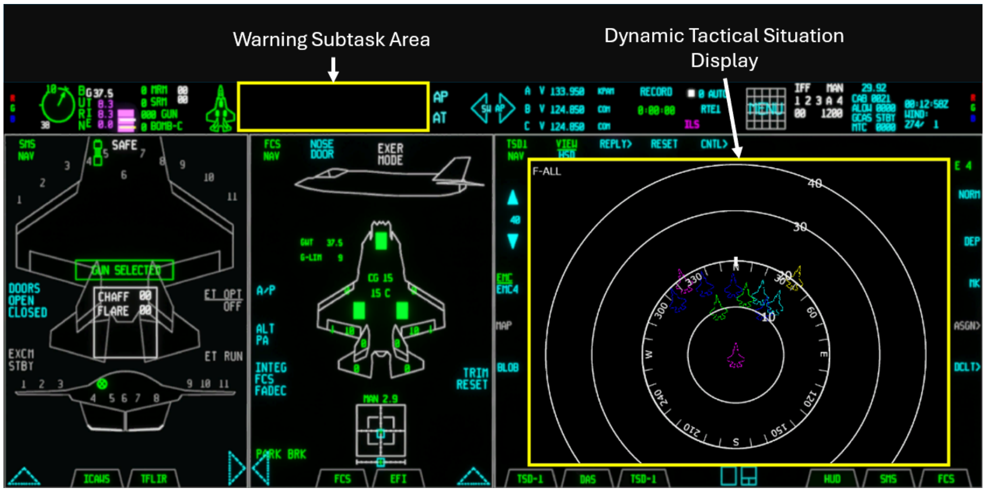


Figure 2: A screenshot of imagery rendered on the HDD. Annotations indicate the empty region of the screen where messages associated with the warning subtask would appear, as well as the dynamic tactical situation display. In this example, the target plane is magenta, as indicated by the color of the plane in the center of the radar screen.

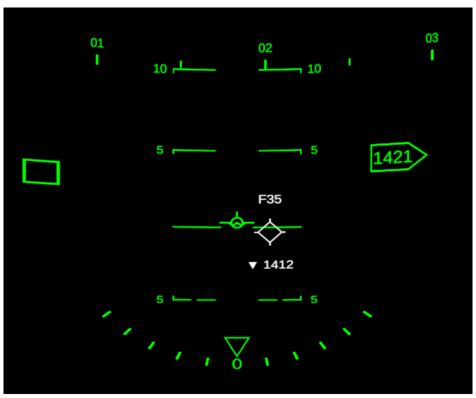


Figure 3: A screenshot of the symbology rendered to the OST AR HMD. The user's altitude appears to the right on the HMD, and a compass, pitch ladder, boresight, and bank indicator appear central on the HMD. In this example, the target plane symbol is drawn in white toward the center of the HMD. The user has already selected the plane and has designated it as an F35 within altitude range. Note that the black regions of the image will appear transparent to the user, such that only the green and white imagery is visible.

to indicate so, otherwise they pressed up. A triangle icon was shown next to the selected plane to indicate the user's response. This particular task was included to force users to have to read HMD symbology rendered in the target color.

Upon completion of the labeling and altitude tasks, users pressed the trigger to confirm their selections and moved on to the next trial.

Randomly throughout each trial, a warning sound occurred and a warning message appeared on the top of the HDD (see the warning subtask area annotation on figure 2). Based on the color of this warning message (yellow or red), the user was tasked to respond with one of two respective inputs on the joystick. This task was included to further increase the cognitive demand of the simulation.

Measures

Performance measures were recorded for each trial of the simulation. These included: trial length and altitude task accuracy. Trial lengths were averaged across all trials in a block to yield a single value for each condition of participant data. For our analysis, trial length measures were corrected to eliminate variation caused by the random angular offset of the user's aircraft to the target aircraft. For each trial length measure, we subtracted the minimum time needed to match the heading of the target aircraft. Thus, the trial length measure considers the time needed to identify the target on the HDD, time taken for unnecessary flight maneuvers, time needed to identify and select the target on the HMD, and time needed to perform the labeling and altitude subtasks.

Eye gaze position, origin, and pupil size were recorded at 200 Hz using the Pupil Labs Neon eye trackers (Pupil Labs, Berlin, Germany). The Neon's software provided additional output in the form of events for blinks, fixations, and saccades. All eye tracking data was segmented into subsets for each trial in the block, and the mean was taken across trials to create a set of aggregate eye tracking measures that could be compared between blocks/participants. For our analysis, we focused on fixation frequency.

Results

We examined participant performance and eye behavior between groups for the three colors sets and two color intensity levels. Corrected trial length and fixation frequency were analyzed using linear mixed-effects models and can be seen in figures 4 and 6 respectively. Participant accuracy in the altitude subtask was analyzed using a generalized linear mixed effects model with an aggregated binomial test family and logit link function, and can be seen in figure 5.

Corrected Trial Lengths

The estimated marginal means for corrected response times are plotted in figure 4 in log scale. A significant effect of color set on corrected trial length was found ($F(2, 26.22) = 6.444$,

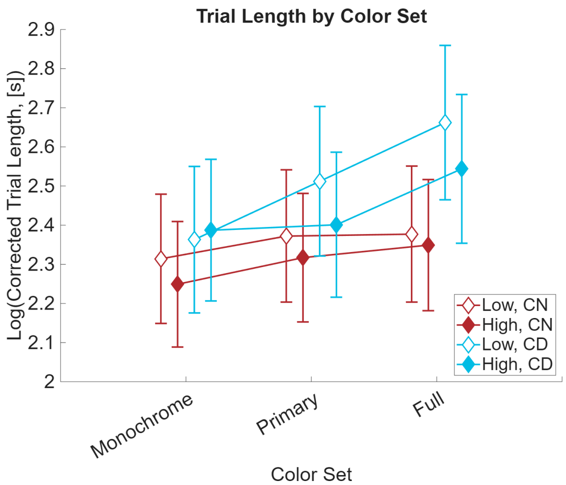


Figure 4: Estimated marginal means across color sets for corrected trial lengths in log scale. Error bars indicate ±1 standard error.

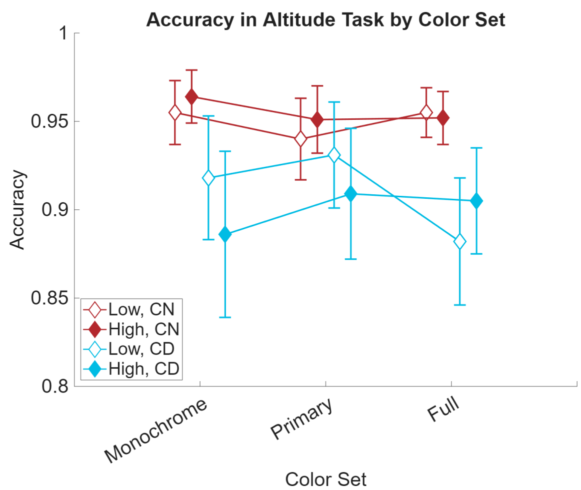


Figure 5: Estimated marginal means across color sets for accuracy in the altitude subtask. Error bars indicate ±1 standard error.

p=0.005), indicating increasing trial lengths with color set size (monochrome M=2.328, SE=0.119, primary M=2.40, SE=0.122, full M=2.483, SE=0.126). No significant effect was observed for color intensity on corrected trial length ($F(1,22.77) = 2.708$, p=0.114). No significant effect was observed for participant group on corrected trial length ($F(1,13.99) = 0.386$, p=0.544). No significant interaction effects were observed.

Altitude Task Accuracy

The estimated marginal means for participant accuracy in the altitude subtask can be seen in figure 5. No significant effect of color set on accuracy was observed ($\chi^2(2) = 1.060$, p=0.589). No significant effect of color intensity on accuracy was observed ($\chi^2(1) = 0.011$, p=0.918). No significant effect of participant group was observed ($\chi^2(1) = 1.821$, p=0.177). A significant 3-way interaction effect between color set, color intensity, and participant group was observed ($\chi^2(2) = 8.908$, **p=0.012**), indicating that the effect of color set depended on color intensity and was different between color normal and color deficient participant groups. All other interactions were non-significant.

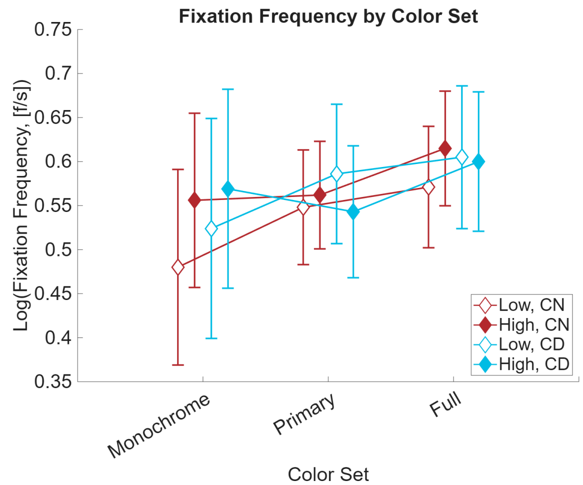


Figure 6: Estimated marginal means across color sets for fixation frequency in log scale. Error bars indicate ±1 standard error.

The significant 3-way interaction effect can be seen by comparing the color normal to color deficient plots in figure 5. From this, we see that color deficient subjects experienced a drop in accuracy in the full color set of the low color intensity condition, whereas this drop is not observed in the high color intensity conditions.

Fixation Frequency

The estimated marginal means of fixation frequency are plotted in log scale in figure 6. No significant effect of color set on fixation frequency was observed ($F(2,19.46) = 3.222$, p=0.062). No significant effect of color intensity on fixation frequency was observed ($F(1,17.72) = 1.694$ p=0.210), and no significant effect of participant group on fixation frequency was observed ($F(1,14.02 = 0.023$ p=0.883). No significant interaction effects were observed.

While no significant main effects were observed in the fixation frequency data, the estimated marginal means in figure 6 show an increasing trend of fixation frequency with color set set size (monochrome M=0.532 SE=0.079, primary M=0.560 SE=0.048, full M=0.598 SE=0.051).

Discussion

The results of the study are somewhat surprising, in that adding additional colors to the color set used to depict symbology on the OST HMD was actually detrimental to user performance, where subjects required more time to complete their trials in conditions with larger color sets compared to the monochrome control case. The lack of significant main effects on accuracy in the altitude subtask suggests that participants performed consistently across conditions. However, the 3-way interaction effect indicated that color deficient subjects had lower accuracy in the full color set compared to the others in the low color intensity conditions, again supporting the conclusion that more colors in the color set tends to be detrimental to user performance. The trend in the fixation data also suggests a similar conclusion, where participants needed to perform additional eye movements to find target symbology on the HMD or to perform the altitude subtask in conditions with larger color sets compared to those with smaller color sets.

One comment that was frequently mentioned by participants is that it was difficult to read the text for the altitude subtask depending on the target color for that trial, particularly so for the low color intensity conditions. The color sets in this study were chosen to test across the whole color gamut of the OST HMD at each primary color, secondary color, and the white point of the display, as well as across two different cross sections of the color space via modulation of the color intensity variable. Because of this, there is variation in luminance between colors in the primary and full color sets, where low luminance colors like blue or red have reduced luminance contrast and thus will be more difficult to see on the OST HMD compared to higher luminance colors like white, yellow, or green. Additionally, color deficient participants may perceive further deviation in luminance between colors in the color sets due to decreased sensitivity to either the red or green color channels of the display. In general, this could partially explain the increased trial lengths, and fixation frequency with color set size, where participants regardless of group are experiencing difficulty finding and interpreting symbology depicted in certain colors due to poor luminance contrast. This could also explain the 3-way interaction effect on accuracy, where color deficient observers struggled to accurately interpret the numbers shown on the OST HMD in the full color set for the low color intensity condition. In this particular condition, it seems likely that the perceived contrast of the symbology for color deficient participants decreased closer to (or below) participants' contrast thresholds, resulting in poor accuracy due to guesses being made in the altitude task.

In general, there were slight differences between color normal and color deficient participants in terms of trial length and accuracy. The lack of significant effects here is promising in terms of expanding the accessibility of the technology, suggesting that color deficient users are just as capable at using OST HMDs as color normal users, at least for the task investigated in this work. However, additional research should be performed across a more diverse range of tasks involving OST HMDs to better establish if and when performance differences exist between groups.

One implication of this research is that pilots may be better off using monochrome OST HMDs rather than color OST HMDs if the color sets employed are similar to the ones tested here. However, it is possible that there are color sets that could improve performance over the monochrome green condition if they are specifically tailored for use on OST HMDs. Future investigations into custom color sets for OST HMDs and/or color-deficient observers should aim to establish a small, well-defined palette that maintains high luminance contrast while ensuring reliable perceptual discriminability.

Conclusion

The additive nature of OST HMDs introduces variability in user perception of virtual imagery based on the lighting characteristics of the user's environment. While prior work has shown that hue can be an effective way to encode visual information, variation in hue investigated in this study was actually detrimental to user performance, resulting in longer trial lengths and higher fixation frequency across participant groups. Our results suggest that monochrome symbology sets should be employed until color sets can be designed specifically to avoid the performance limitations observed in this study. Future work should investigate color sets designed specifically for use with OST HMDs and establish a small set of high luminance colors that maintain perceptual discriminability.

Acknowledgments

This research was supported by USAF Contract FA8650-21-C-6277 to KBR, Inc. The views expressed are those of the authors and do not reflect the official views of the United States Air Force, nor the Department of Defense. Mention of trade names, commercial products, or organizations do not imply endorsement by the U.S. Government. This work is cleared for public release: AFRL-2026-1278.

References

[1] R. T. Azuma, "A survey of augmented reality," *Presence: teleoperators & virtual environments*, vol. 6, no. 4, pp. 355–385, 1997.

[2] W.-C. Li, J. Zhang, S. Court, P. Kearney, and G. Braithwaite, "The influence of augmented reality interaction design on pilot's perceived workload and situation awareness," *International Journal of Industrial Ergonomics*, vol. 92, p. 103382, 2022.

[3] A. Sarbach and T. R. Weber, "Next-generation navigation: Evaluating the impact of augmented reality on situation awareness in general aviation cockpits," in *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*, CHI '25, (New York, NY, USA), Association for Computing Machinery, 2025.

[4] J. L. Gabbard, J. E. Swan, and A. Zarger, "Color blending in outdoor optical see-through ar: The effect of real-world backgrounds on user interface color," in *2013 IEEE Virtual Reality (VR)*, pp. 157–158, 2013.

[5] J. L. Gabbard, M. Smith, C. Merenda, G. Burnett, and D. R. Large, "A perceptual color-matching method for examining color blending in augmented reality head-up display graphics," *IEEE Transactions on Visualization and Computer Graphics*, vol. 28, no. 8, pp. 2834–2851, 2022.

[6] A. Erickson, K. Kim, G. Bruder, and G. F. Welch, "Exploring the limitations of environment lighting on optical see-through head-mounted displays," in *Proceedings of the 2020 ACM Symposium on Spatial User Interaction*, SUI '20, (New York, NY, USA), Association for Computing Machinery, 2020.

[7] M. D'Zmura, "Color in visual search," *Vision Research*, vol. 31, no. 6, pp. 951–966, 1991.

[8] A. L. Nagy and R. R. Sanchez, "Critical color differences determined with a visual search task," *J. Opt. Soc. Am. A*, vol. 7, pp. 1209–1217, Jul 1990.

[9] J. Blundell, S. Scott, D. Harris, J. Huddlestone, and D. Richards, "With flying colours: Pilot performance with colour-coded head-up flight symbology," *Displays*, vol. 61, p. 101932, 2020.

[10] A. Erickson, G. Bruder, and G. F. Welch, "Analysis of the saliency of color-based dichoptic cues in optical see-through augmented reality," *IEEE Transactions on Visualization and Computer Graphics*, vol. 29, no. 12, pp. 4936–4950, 2023.

[11] A. Erickson, G. Bruder, and G. F. Welch, "Adapting michelson contrast for use with optical see-through displays," in

2022 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct), pp. 409–410, 2022.

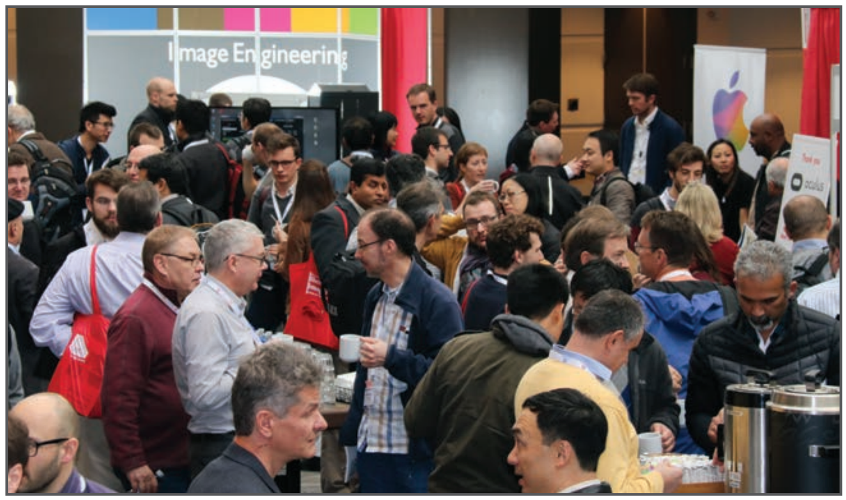
Author Biography

Austin Erickson is a Research Scientist for the Operational Based Vision Assessment (OBVA) Laboratory at Wright-Patterson AFB, OH. He received his BS in Computer Science from the University of South Florida, his MS in Computer Science from the University of Central Florida, and his PhD in Computer Science from the University of Central Florida. His work at the OBVA Lab spans perceptual, technological, and design considerations for augmented and virtual reality displays, and simulation platforms.

JOIN US AT THE NEXT EI!

electronic IMAGING

Imaging across applications . . . Where industry and academia meet!



• SHORT COURSES • EXHIBITS • DEMONSTRATION SESSION • PLENARY TALKS •
• INTERACTIVE PAPER SESSION • SPECIAL EVENTS • TECHNICAL SESSIONS •

www.electronicimaging.org

ADVANCING IMAGING SCIENCE 75 IS&T 1947 2022 imaging.org