

PeriphAR: Fast and Accurate Real-World Object Selection with Peripheral Augmented Reality Displays

Yutong Ren
School of Information
University of Michigan
Ann Arbor, Michigan, USA
rentony@umich.edu

Arnav Reddy
School of Information
University of Michigan
Ann Arbor, Michigan, USA
arnavr@umich.edu

Michael Nebeling
School of Information
University of Michigan
Ann Arbor, Michigan, USA
nebeling@umich.edu

Abstract

Gaze-based selection in XR requires visual confirmation due to eye-tracking limitations and target ambiguity in 3D contexts. Current designs for wide-FOV displays use world-locked, central overlays, which is not conducive to always-on AR glasses. This paper introduces PERIPHAR (*/peh-ree-faar/*), a visualization technique that leverages *peripheral vision* for feedback during gaze-based selection on a monocular AR display. In a first user study, we isolated text, color, and shape properties of target objects to compare peripheral selection cues. Peripheral vision was more sensitive to color than shape, but this sensitivity rapidly declined at lower contrast. To preserve preattentive processing of color, we developed two strategies to enhance color in users' peripheral vision. In a second user study, our strategy that maximized contrast of the target to the neighboring object with the *most similar color* was subjectively preferred. As proof of concept, we implemented PERIPHAR in an end-to-end system to test performance with real-world object detection.

CCS Concepts

• **Human-centered computing** → **Interface design prototyping; Mixed / augmented reality.**

Keywords

Monocular AR; peripheral displays; gaze-based interaction.

ACM Reference Format:

Yutong Ren, Arnav Reddy, and Michael Nebeling. 2026. PeriphAR: Fast and Accurate Real-World Object Selection with Peripheral Augmented Reality Displays. In *Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems (CHI '26)*, April 13–17, 2026, Barcelona, Spain. ACM, New York, NY, USA, 16 pages. <https://doi.org/10.1145/3772318.3791637>

1 Introduction

Future AR glasses will require strategic compromises in display and sensing functionality to achieve optimal form factors conducive to always-on, pervasive AR [14]. For example, the new Meta Ray-Ban Display smart glasses set the emphasis on wearable *AI*, rather than *AR*, featuring only a limited, monocular display compared to AR headsets such as HoloLens. While these smart glasses come with advanced input channels using ultra-wide cameras, directional microphones, and EMG wristbands, they essentially function as a

heads-up display. The lack of world-locked rendering makes real-world object selection, a fundamental interaction step in any AR experience [4], particularly challenging. A related problem with a near-eye display that projects an image slightly off to the side of one eye, thus not directly in the user's line of sight, is that it essentially creates a floating window in the user's periphery. Interacting with such a display requires frequent gaze shifts and focal length adjustments that can be distracting from real-world interaction [42] as well as increasing digital eye strain [19, 21].

In this paper, we introduce PERIPHAR (*/peh-ree-faar/*), a visualization technique that leverages *peripheral vision* for feedback during gaze-based selection on monocular AR displays. Being well established in XR [12], gaze-based selection was previously shown to naturally support intuitive hands-free interaction [40, 43, 51, 52]. However, existing techniques typically rely on visual cues, such as eye cursors, gaze rays, or object highlights [5, 43], to be directly overlaid in the user's central (or foveal) vision. Although prior work has shown that near-peripheral vision can present secondary information without disrupting tasks that require foveal attention [22], it has not been previously studied as a feedback channel for object selection. With peripheral vision being less sensitive to detail [33], the main challenge is deriving a visual representation from real-world targets that encodes the most salient visual properties of the target so that it can be clearly perceived in peripheral vision.

With PERIPHAR, we make two contributions: (1) we present controlled experiments that provide empirical insights into effective peripheral cues for real-world object selection that enhance selection accuracy, reduce cognitive load, and increase user confidence; (2) we detail the design of a pipeline implementing two strategies to enhance color for peripheral vision, demonstrating that our strategy that maximized contrast of a target to the neighboring object with the *most similar color* was preferred and yielded higher confidence.

2 Background and Related Work

This research establishes principles of limited, peripherally glanceable AR displays that allow users to keep their visual attention in the real world and simultaneously perceive content in the display without having to focus on it.

Glanceable AR Displays. There is an existing concept of glanceable AR presented by Lu *et al.* [30, 31]. In [30], they studied benefits and drawbacks of glanceable AR applications for different use cases at work, home, or on-the-go based on video and functional prototypes. In [31], they evaluate interaction techniques designed to bring AR content from the peripheral to central vision. Finally, in [32], they studied adaptable and adaptive AR content transition



This work is licensed under a Creative Commons Attribution 4.0 International License. *CHI '26, Barcelona, Spain*

© 2026 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-2278-3/26/04
<https://doi.org/10.1145/3772318.3791637>

mechanisms for moving interface widgets to different locations in the user’s physical context. Their concept is related but different from ours. One important difference is that these studies assume more capable AR displays (modeled after HoloLens and Magic Leap, *i.e.*, with binocular views, world-locked rendering, and wide FOV). Importantly, they distinguish the periphery relative to the eye and to the head, and reserve the peripheral vision to anchor virtual objects with physical locations to present secondary information in collapsed views that will expand as the user brings them into their central vision through either head or eye movements. In contrast, we aim to avoid the need for foveating towards the display, instead providing visual, peripheral cues in the monocular display so that users can perceive them while keeping their attention real-world.

Prior work has investigated lightweight, glanceable designs for AR glasses, highlighting challenges of limited fidelity and brightness [26, 44]. Researchers proposed context-aware and adaptive interfaces to improve glanceability [9, 29]. One particularly promising direction explored dynamic color enhancement to maximize perceptual clarity under resource constraints. For example, Chroma tailored AR palettes for color-blind users [56], Chen *et al.* [6] applied imperceptible modulation for power saving, and Duinkharjav *et al.* [11] optimized rendering with perception-guided recoloring. These works illustrate how adaptive color feedback can reduce clutter or energy use. Our work builds on this literature but applies color enhancement specifically to peripheral cues to improve selection confirmation.

Designing for Peripheral Vision. Most existing designs for head-worn AR target central and foveal vision, largely neglecting peripheral vision. Yet, peripheral vision excels at detecting motion and high-contrast changes, providing users with ambient awareness without requiring direct gaze shifts. Strasburger *et al.* [53] characterize peripheral vision to be of lower spatial resolution but with a higher sensitivity to contrast, making it ideal for subtle but salient notifications. Building on this, Sun and Varshney [54] quantified the perception time in the far peripheral field for VR/AR, showing that users can register peripheral stimuli within milliseconds, even under complex visual loads.

The HCI literature contains few studies on peripheral vision, which can be classified as building custom hardware to create new types of displays [8, 34], studying display placement [2, 18], and the level of detail of the information shown [10, 15, 16]. Of particular interest to this research is the work by Luyten *et al.* [34] who created a display setup involving two peripheral displays mounted within the 20-degree extremes of the periphery to study the usable peripheral display area in terms of position and size relative to the eye, and the impact of shape, orientation, color, and animation on users’ perception. They recommended (1) using a limited set of simple, prominent shapes (*e.g.*, rectangles and circles in 2D), (2) using primary, strongly contrasting colors (their participants correctly perceived red, green, and blue, but had challenges with yellow, purple, and orange), and (3) using animation intentionally to communicate more complex information (*e.g.*, motion in the direction an arrow is pointing).

A new stream of research has leveraged peripheral channels for subtle, multisensory cues. Ku *et al.* [27] used rapid serial visual presentation (RSVP) to deliver textual overlays that are readable

using peripheral vision. Trepkowski *et al.* [57] combined visual, auditory, and haptic signals to improve awareness under a narrow field-of-view (FOV). Parmar & Silpasuwanchai [41] studied attentional tunneling while walking, Janaka *et al.* [22] explored paracentral visualization strategies, and Syiem *et al.* [55] showed reduced cognitive load during multitasking. This stream suggests peripheral vision is an effective communication channel; however its use to interact with AR content remains underexplored.

Feedforward and Feedback in Visual Guidance. Our framing of PERIPHAR as a visual *feedback* mechanism builds on prior work that distinguishes feedforward from feedback in interaction design. Vermeulen *et al.* [58] clarify feedforward as information presented *before* an action is carried out to reduce Norman’s Gulf of Execution, complementing but differing from feedback about the outcomes of completed actions. In pointing interfaces, Guillon *et al.* [17] introduce a design space of visual feedforward for target expansion, emphasizing that users must see how targets will expand *prior* to movement to benefit from the technique. Sadasivan *et al.* [47] similarly use recorded eye movements as feedforward training cues that preview how experts would inspect complex aircraft structures, again focusing on anticipatory guidance rather than on-line correction.

Recent XR work further sharpens this distinction. Yu *et al.* [61] propose a design space for visual feedforward and *corrective feedback* in XR motion guidance, defining feedforward as cues that specify desired future motion and feedback as cues that respond to a user’s ongoing performance and highlight errors. In parallel, research on peripheral visual feedback shows that unobtrusive cues in the visual periphery can efficiently signal system state and guide attention without disrupting foveal tasks [38]. AR studies have leveraged such feedback to support continuous control and search: Wang *et al.* [59] compare central and peripheral visual feedback (and other modalities) for trajectory tracking in AR, and Richards *et al.* [46] combine peripheral visual and vibrotactile feedback to help users locate dynamic virtual entities in augmented reality. Gaze-based targeting work likewise uses visual highlights and cursor changes as feedback about the system’s current interpretation of the user’s gaze selection [13]. Following Yu *et al.*’s distinction between feedforward and *corrective feedback* in XR motion guidance [61], these cues are presented in response to the user’s or system’s *current* state, rather than predicting future actions.

We explicitly position PERIPHAR as providing *peripheral feedback* instead of feedforward: our peripheral proxies appear only after the system has interpreted a fixation on a real-world object, and they visualize “what the system currently thinks you are selecting” in the monocular display. Users can then confirm or reject this interpretation without foveating toward the display.

Preattentiveness to Color and the Link to Peripheral Vision. Classic work on preattentive vision shows that color is a particularly powerful feature for guiding visual search. In dense arrays, color singletons can be detected in parallel and “pop out” even when observers are not deliberately attending to them [39]. Later work refined this view by demonstrating that individual hues differ in how strongly they guide attention. For example, Andersen *et al.* found that red tends to be detected the fastest, followed by green, blue, and yellow, whereas orange and purple are less effective

at improving search performance [1]. Beyond low-level contrast, preattentive responses are also shaped by color categories themselves. Using peripheral oddball paradigms and visual mismatch negativity, Clifford *et al.* showed that across-category differences (e.g., blue vs. green) elicit stronger preattentive responses than equally large within-category differences [7]. These results suggest that ergonomic design guidelines for color should consider not only contrast and luminance, but also the choice and categorical perception of specific hues.

Preattentive color processing is closely linked to peripheral vision. Visual search theories argue that feature maps for color, orientation, and motion are computed across the visual field, providing a coarse salience landscape that can guide eye movements toward promising locations [60]. Empirical studies show that color signals support rapid detection even when targets are presented outside the fovea. For example, Khomeriki and Lomashvili reported that under peripheral viewing, red stimuli are detected faster than green or blue, and that this advantage interacts with task complexity and spatial location in the visual field [24]. Together with classic findings on preattentive filling-in and texture segmentation [39], this work indicates that color-based guidance is available in peripheral vision, albeit at lower spatial resolution than in central vision.

Our work builds on this literature by exploiting peripheral color sensitivity in an AR context. PERIPHAR presents proxy objects in the user’s peripheral rather than foveal vision, and our studies suggest a predisposition to color and similar ordering of color effectiveness (with red targets noticed fastest, and green/yellow slower), consistent with preattentive color findings in both central and peripheral vision [1, 7, 24, 39]. We leverage these insights when reasoning about how color-enhanced peripheral proxies can support rapid confirmation of gaze-based selections on low-FOV AR glasses, and we return to them in our discussion of future extensions for multi-colored, real-world targets (Sec. 8).

3 Peripheral AR Display Simulation

To simulate a peripheral AR display and support the design of our PERIPHAR feedback technique, we developed a system based on Quest Pro, a current MR headset with built-in support for video passthrough and eye tracking. As illustrated in Fig. 1, the four main components of our system are: (1) **peripheral AR display emulation** to gracefully degrade Quest Pro’s display and rendering capabilities and simulate monocular AR glasses, (2) **real-world object simulation** to generate virtual objects as targets and control their number, size, and distance in our experiments, (3) **gaze-driven object selection** to enable object selection using eye tracking for both virtual objects generated by our simulation and physical targets detected by our system, and (4) **peripheral proxy generation** to create a visual representation of a targeted virtual or real-world object as feedback in the simulated peripheral display.

Display Emulation. Simulating AR glasses is complex. For our experiments, we needed to control key parameters such as display position and size relative to a user’s visual field on Quest Pro, but not accurately model all the optics (e.g., luminance and brightness under different environmental conditions). We scoped our display emulation to meet two requirements: (i) **control the simulated display’s field of view and rendering capabilities** separately

from the Quest Pro MR display with video passthrough rendering; and (ii) **emulate a monocular display context** that renders an image of the selected target in the simulated display only for the right eye.

Our system provides visual controls to adjust the position and size of the display area. We added means to simulate lower FPS and adjust the display image in terms of resolution and color. As part of the calibration in our experiments, we placed the display area in each user’s periphery so that they could still see all four corners of the simulated display when foveating towards it. We consistently chose a 20° horizontal field of view and simulated the display at VGA resolution with RGB color at 30 FPS for most parts of the experiment (the text and shape conditions used grayscale). This yields a limited AR display that falls within the spectrum of current smart glasses. For example, the recently released Meta Ray-Ban Display glasses feature a 600x600 pixel, full-color RGB, monocular display in the right lens with a 20-degree field of view, and a 90Hz refresh rate (content refreshes at 30 FPS).

Target Simulation. For our experiments, we wanted to be able to simulate real-world targets using virtual 3D shapes and models registered in world space, serving as stand-ins for real objects in the physical environment. For example, our first study used real-world targets in the form of tetromino blocks (Fig. 3) (with seven different colors and shapes), and our second study used 3D fruit models (with similarly looking colors and shapes). This made it possible to observe gaze dynamics with respect to virtual targets whose real-world appearance and in-display representation we could fully control. It also meant that we could run experiments without being impacted by object recognition and segmentation errors.

To test our PERIPHAR visualization technique on real objects in the physical environment, we later developed an end-to-end system using YOLO11n [23] object detection and segmentation (Sec. 7). This served as a proof of concept that that our pipeline could be extended with computer vision techniques and used outside the lab, but was not meant as a generalization from our controlled studies.

Gaze-Driven Selection. Our system implements gaze classification via velocity threshold identification (I-VT) for saccades and dispersion threshold identification (I-DT) for fixations [48], with fixed thresholds established in prior work [49]. For durations, we used a 30ms maximum for saccades and a 50ms minimum for fixations in Study 1. In Study 2, we used 45ms minimum fixation duration to compensate for the delay in peripheral proxy generation.

Our system further classifies gaze samples to determine display peripherality as a metric for comparison and proxy for glanceability. The key idea is that a display condition is better if it effectively reduces eye movements, since it would make it easier to process the visual feedback rendered in the peripheral display and match it to the real-world target *without* foveating towards the display. Thus, we were interested in detecting where fixations and saccades happened relative to the display.

As illustrated in Fig. 2, our system structures the user’s visual field into a gradient around the display, leading to clear-cut *real-world vs. in-display* attention zones, and a *world/display* transition zone in between. Our display peripherality metric, then, consists of the combined durations of fixations on objects in the world, as opposed to fixations on the display, as well as the frequency of

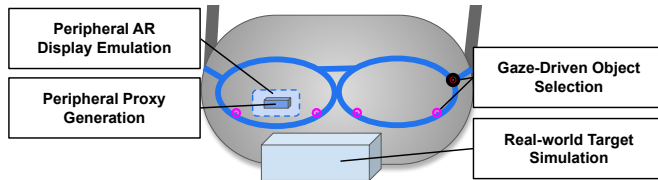


Figure 1: Our system consists of four components: (1) *peripheral AR display emulation* simulates monocular AR glasses with limited FOV on Quest Pro, while (2) *gaze-driven object selection* uses eye tracking and video passthrough for object selection; (3) *real-world target simulation* enables simulation of target objects using virtual 3D shapes and models in world space, while (4) *peripheral proxy generation* creates a visual representation of the real-world target on the simulated display screen. The system was used in two studies to determine key object properties for visually salient target representations.

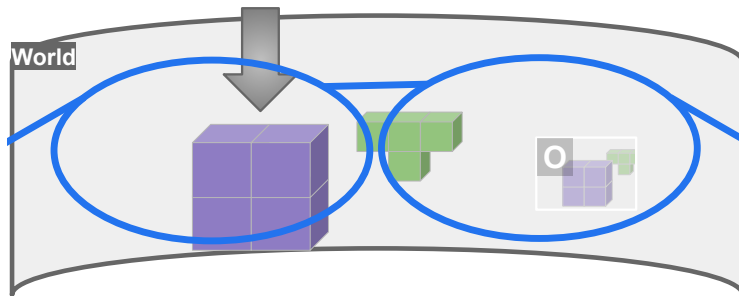


Figure 3: The tetris game in Study 1 simulated real-world targets using virtual objects in the shape of tetrominoes in four display conditions: the *snapshot* condition combined the text label (O), square shape in color (blue O), and surrounding objects (green T). The *text*, *color*, and *shape* conditions isolated these attributes as shown on the right.

world/display transitions measured in terms of saccades between real-world and in-display fixations. We define an AR display to be of higher peripherality the higher the time real-world (tW), the lower the time in-display (tD), and the shorter and the fewer the transitions. We note that, when comparing different display conditions, key performance indicators depend on the complexity of the content shown in the display. That is, when comparing more complex displays, we can also expect higher values for the in-display metrics, leading to lower real-world to in-display ratios, hence decreased glanceability.

To determine intersections of the gaze with real-world targets, our system implemented a version of the flashlight technique often used in 3D contexts [28, 42]. The system would lock onto the closest intersecting target and render a visual representation of it in the simulated display. In addition to visual cues rendered in the simulated display, we also added audio cues by playing distinct sounds each time the target has changed and when selection was confirmed via controller button press. We used a controller-based confirmation rather than hand tracking to circumvent hand tracking errors and allow participation of users who were less familiar with XR technology.

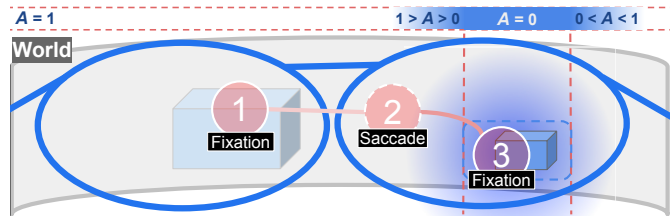


Figure 2: Our system measures *display peripherality* based on gaze dynamics: a design that requires less time to look at, and fewer gaze transitions into, the display means it provides better peripherality. By structuring the visual field into a gradient (blue) around the simulated display, we compute the ambient value (A) of each gaze sample as: (1) $A = 0$, when the vector points *in-world* and away from the display, (2) $0 < A < 1$, when in the blue zone modulated within 10° in each direction around the display, or (3) $A = 1$, when *in-display* and intersecting with the simulated display's screen area (dotted blue box).

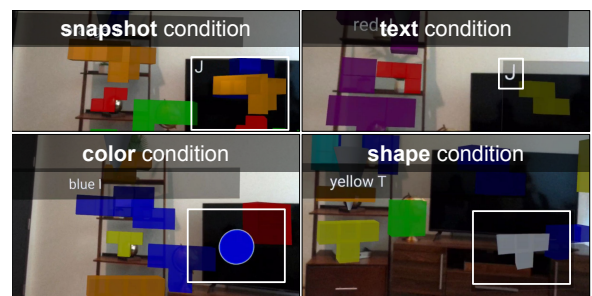


Figure 4: The four display conditions used in the tetris game: *snapshot* with the text label and image of the target with surrounding shapes; *text* with only the text label; *color* with only a circle filled in the target's color; *shape* with only the target's shape isolated in gray.

Peripheral Proxy Generation. We developed techniques to generate visual representations of real-world objects targeted by the user and used them to create four display conditions in Study 1 and three display conditions in Study 2, detailed in the next sections. The studies had different, yet complementary goals. In Study 1 (Sec. 4), we wanted to isolate key object properties with respect to display peripherality based on controlled experiments with virtual objects inspired from the popular tetris game. In Study 2 (Sec. 6), we wanted to compare different strategies for generating the peripheral proxy and assess if it can be interpreted peripherally without gaze shifts to the simulated display. Finally, we developed an end-to-end system as a proof of concept (Section 7), employing real-time object detection and peripheral proxy generation from Quest Pro's video passthrough, and produced results for realistic environments.

4 Study 1: Peripheral Sensitivity

The goal of our first study was to condition the AR display design in terms of visual features—text, color, and shape—and study how each individually contributes to display peripherality. To maximize control and balance visual features, we simulated targets using virtual objects in a game we developed inspired by Tetris. For the experiment, we wanted virtual objects that can approximate the

shape of real-world objects while being able to control and measure the differences. Cuboids are commonly used in studies that segment real-world scenes (e.g., [36]), which motivated us to develop the Tetris game (Figure 3). Although the chosen colors in the game are less representative of the appearance of real-world objects, the selection of tones and shades in the color spectrum led to smaller and larger perceptual differences, which we considered realistic.

Task. Each trial involved selecting one of ten tetrominoes which were generated by our system using a combination of seven shapes (I or ‘straight,’ O or ‘square,’ T, L, J, S, and Z) and seven colors (cyan, blue, purple, green, yellow, orange, red) common to Tetris. Generated tetrominoes were unique in that the same color *or* shape were possible, but no more than one tetromino of any color was present in a task.

All tetrominoes were placed in fixed physical locations in front of participants who remained seated for control. The system randomly selected each tetromino exactly once and announced it verbally (e.g., “red T”), marking it with an arrow. We used this as an attention guiding technique [45] with the goal of minimizing search time and reducing user effort to targeting, similar to when someone knows what object they are looking for and roughly where it is in the environment. We opted for a continuous task design in that the next tetromino was marked and announced immediately after the correct one was selected. As recommended in [3], we considered a discrete task design, but pilots found it too disruptive to realistically capture world/display transitions and make task completion time a factor. It would also have decreased the performance aspect of the game, which challenged participants to be accurate, yet efficient, by selecting the correct tetrominoes quickly.

Procedure. Participants conducted three blocks of 10 trials per block under four conditions (randomized and counterbalanced): (1) *snapshot* showed a colored image of the target shape with surrounding shapes as well as a text label with the shape’s letter; (2) *text* only showed the text label without color or shape, (3) *color* just showed a circle in the color of the target, and (4) *shape* only showed the target’s shape without color (Figure 4). The *snapshot* condition functioned as a baseline with all features; the others were control conditions that isolated the individual characteristics to see how much each feature in isolation contributed to user performance.

Participants. In total, 32 participants completed the study (17 female, 15 male; ages 18–65, $M=38.28$, $SD=12.04$). Participants’ AR/VR experience varied from no prior experience (8 participants) to ten years ($M=1.72$ years, $SD=2.51$ year), using AR/VR on average 0.79 days per week ($min=0$, $max=5$). Screening confirmed that the participants were at least 18 years old and had not previously experienced motion sickness when using VR. Clear vision was required, wearing contact lenses rather than glasses, if needed, not to interfere with the eye tracking of Quest Pro. Each study session lasted between 60 and 90 minutes. Participants were compensated \$75 USD per hour.

Calibration & Pre-Study Questionnaires. Each participant performed individual calibration of Quest Pro using suggested fit adjustments and IPD values, then went through the eye tracking calibration procedure to build an optimized gaze model for each user. After successfully testing the calibrated model, participants

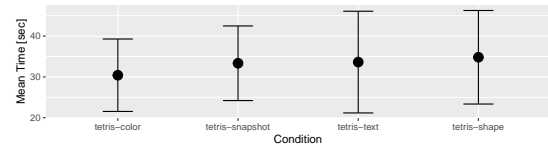


Figure 5: Mean task completion times in ascending order from left to right. Error bars show standard deviation.

completed a pre-study questionnaire asking for their perception of the most common symptoms of motion sickness (SSQ) and digital eye strain (DES) per Hirzle *et al.* [20]. After each condition, we administered an abbreviated questionnaire to monitor any symptoms. Throughout the study, we followed best practices and exercised caution to be sensitive to any issues noted by the participants. Although no participants experienced motion sickness, one participant felt mild symptoms of digital eye strain after calibration, but the symptoms disappeared soon after and they completed the study. Before each condition, we calibrated our system to ensure that the simulated display was positioned correctly in the participants’ visual field and the gaze cursor was tracked accurately.

Data Collection & Analysis. We used our system to record user sessions and produce real-time analytics based on our peripherality metrics. The record/replay system allowed us to collect and analyze detailed logs of participants’ eye movements per task, block, and condition. We securely stored participants’ anonymous experiment data and developed scripts in JavaScript and R to pre-process data from successful trials and perform statistical analysis. In post-hoc analysis, we compared the display conditions in terms of logged statistics and subjective ratings collected in post-task questionnaires.

Table 1 presents key findings in terms of task performance recorded by the system and subjective performance based on post-task questionnaires on preference for condition, perceived task performance and display peripherality.

Objective Performance. We discarded the first block in all conditions that was used as training. For statistical analysis, we ran Mauchly’s test of sphericity against condition, then one-way repeated-measures ANOVA (with Huyn-Feldt correction if the sphericity could not be assumed) and, if significant, post-hoc pairwise t-tests with Bonferroni correction. Significant results mean $p < 0.05$. The condition had a significant effect on task completion time, with *color* being significantly faster than the *shape* condition. Overall, the fastest conditions per participant were *color* (17×), *snapshot* (6×), *shape* (5×), and *text* (4×). The *color* condition also outperformed the other conditions on average spending the least amount of time foveating towards the display, with significant differences between *color* and all other conditions. The $tW:tD$ ratios were the best for the *color* condition, followed by *snapshot*, then *shape*, and finally *text*. The *color* condition had the fewest world/display transitions on average, with significant differences between *color* vs. *shape* and *color* vs. *text*. It also had the best *real-world* vs. *in-display* gaze ratio and was rated the best in terms of display comfort, timing, feeling supported, and not being in the way. It also had the best ratings for mostly looking at the world vs. the display.

In terms of the total number of selection errors, the *shape* condition had the highest number of errors for all participants ($total =$

condition	t [s]	tD [s]	tW:tD	transitions	errors	satisfied	efficient	confident	comfortable	at right time	felt supported	felt in the way
snapshot	33.3	4.9	4.1:1	10	1.31	5.69	5.03	5.44	4.81	5.47	4.97	3.38
text	33.6	5.4	3.6:1	11	1.06	5.91	5.56	5.91	5.19	5.41	5.25	2.66
color	30.4	3.0	6.0:1	7	1.31	6.06	5.69	5.88	6.13	6.09	5.91	2.09
shape	34.8	5.8	3.7:1	11	1.64	5.97	5.34	5.63	5.22	5.31	5.00	3.13

Table 1: Peripherality metrics and post-task ratings for our Tetris game (best results in bold): mean task completion times (t) and time spent *in-display* (tD), time *real-world vs. in-display* ratio ($tW:tD$), average number of world/display transitions and errors, and 7-point Likert scale ratings of task performance (satisfied, efficient, confident) and display quality (comfortable, at the right time, felt supported, felt in the way). The *color* condition was the fastest on average and achieved the best *real-world vs. in-display* gaze ratio, fewest transitions, and the best ratings in terms of display comfort, timing, feeling supported, and not feeling in the way.

105, $mean = 1.64$). The *snapshot* and *color* conditions performed similarly (both $total = 84$, $mean = 1.31$), while *text* had the fewest errors ($total = 68$, $mean = 1.06$). For further analysis, we replayed study sessions to analyze user behavior when errors occurred. There was a fairly even spread of wrong color vs. shape selections for *shape* (23× wrong color, 22× wrong shape, 60× both wrong), suggesting that shape alone may be insufficient. The *color* condition more often led to selecting the wrong shape (9× wrong color, 33× wrong shape, and 42× both wrong). The *snapshot* condition reversed the wrong color vs. shape ratio (21× wrong color, 11× wrong shape, 51× both wrong). The *text* condition had the fewest errors (16× wrong color, 6× wrong shape, 46× both wrong). The top-3 wrong colors were selecting orange instead of yellow (10×), purple instead of teal (6×), green instead of blue (5×), and yellow instead of green (5×). The top-3 wrong shapes were selecting J instead of L (12×), L instead of J (11×), and S instead of Z (4×). Participants made the fewest errors with red (only once selecting teal instead) but confused all shapes at least once.

Subjective Performance. For statistical analysis, we used Friedman rank sum test and, if significant, post-hoc pairwise Wilcoxon signed rank test with Bonferroni correction. Participants tended to rate their task performance the best for the *color* and *text* conditions, but the differences were not significant. For the questions on the display, the *color* condition was significantly rated the highest. We found significant differences between *color* and all other conditions in terms of whether it was easy to understand the information shown on the display and whether it felt comfortable to look at the display. It was also rated significantly better than *snapshot* and *shape* in terms of whether the information on the display was displayed at the right time and whether the display felt like it was in the way. When asked whether the participants felt they mostly looked *real-world vs. in-display*, there was a trend towards *real-world*, with the *color* and *shape* conditions rated the highest, but the differences were not significant.

Reflection & Takeaways. From our experiment observations and participant feedback, two main themes emerged: (1) *Perceptual efficiency of color*: Participants favored the color condition for its efficient communication of information. Particularly when perceived peripherally, color was the most potent feature and it provided a manageable amount of information, making it easier for participants to interpret. (2) *Limitations of color alone*: Some participants encountered difficulties when nearby tetrominoes had similar colors, prompting reflections on the limitations of relying solely on color cues on a peripheral display. Participants suggested

combining color with other features such as shape. Some articulated benefits of the snapshot condition, as it showed more context, while others maintained that it was confusing.

5 Peripheral Proxy Generation

Given the findings from Study 1—that participants favored the color condition for its perceptual efficiency, but struggled when nearby objects had similar colors—we devised a technique that generates a peripheral proxy from a real-world object targeted by the user’s eye gaze and makes differences to similar objects nearby more visible. Our technique is based on a color enhancement algorithm that shifts key color attributes—luminance, saturation, and chromaticity—to make the proxy more noticeable in peripheral vision while remaining consistent with the real-world object.

Color Enhancement Algorithm. As shown in Fig. 6, our algorithm takes two inputs: a *target* image for enhancement and a *reference* image against which the target is contrasted. The images are transformed through a pipeline in three steps: (1) quantization, (2) palette-level color distance analysis, and (3) masked enhancement of the target. We detail the algorithm in Appendix A but give an intuitive explanation below.

Consider the green pear (target) and the green apple (reference) shown in Fig. 6. The green pear has a lighter and more yellowish green than the green apple. The differences are noticeable in foveal vision but much harder to detect in peripheral vision. The algorithm first simplifies the green pear’s texture into a small set of dominant colors (quantization), then enhances lightness, saturation, and chromaticity to maximize contrast with the green apple (the quantization step and the full pipeline will be explained in detail below). The resulting green pear proxy appears brighter and more distinct, making it easier to recognize peripherally. If the target is darker than the reference, the algorithm performs only quantization and skips enhancement, since further darkening would reduce visibility in the periphery.

Technical Pipeline. Building on the gaze classification and display emulation described in Section 3, we implemented a pipeline that connects real-world fixation to peripheral proxy generation and subsequent rendering. Once a valid fixation is detected on a real-world object, our system designates it as the *target* and extracts its texture as the target image from a decimated version of the object model. Decimation reduces geometric detail while retaining the outline shape, ensuring that proxies remain lightweight for peripheral rendering. Next, a *reference* image is obtained using one of two strategies described below. Both images are color enhanced

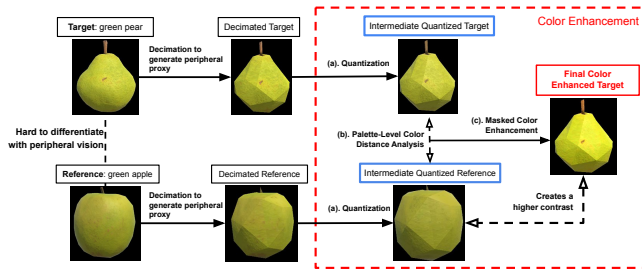


Figure 6: Color enhancement pipeline. (a) The target and reference undergo color quantization ($k=7$). (b) Palette-level color distances are computed between the intermediate quantized target and reference. (c) The intermediate quantized target is color-enhanced based on the computed palette-level distances, producing the final texture for the peripheral proxy.

by our algorithm, which adjusts the target’s appearance relative to the reference to increase salience of the peripheral proxy while preserving visual consistency with the target. The final color enhanced target is then displayed in the peripheral view, completing the feedback loop from gaze fixation to proxy confirmation.

Strategy 1: Snapshot as Reference. The first strategy assumes that users rely on the surrounding context of the real-world object. When a valid fixation is detected, PERIPHAR registers the real-world object as the target and extracts a localized screenshot centered on it, including adjacent objects and background, similar to the *screenshot* condition in Study 1. The screenshot serves as the reference image. By analyzing the screenshot’s color distribution, the algorithm adjusts the target’s luminance, saturation, and chromaticity to increase contrast with its surrounding objects captured in the screenshot. We refer to this as the *screenshot* strategy.

Strategy 2: Most Similar Object as Reference. The second strategy goes beyond the first by identifying the one nearby object that is most likely to be confused with the target. Upon fixation, PERIPHAR compares the target’s color to its neighbors in CIELAB space using the CIEDE2000 color-difference formula [50], selecting as reference the adjacent object with the smallest color distance—i.e., the object most similar in color and therefore most likely to be confused with the target. The algorithm then enhances the target relative to this reference, adjusting luminance, saturation, and chromaticity to maximize perceptual contrast to this reference. We refer to this as the *most similar color* (MSC) strategy.

Screenshot optimizes contrast against the average context, whereas *MSC* optimizes contrast against the worst-case distractor. In our second study, we tested both strategies (average vs. worst-case) to generate a peripheral proxy. We wanted to see which one was more salient and allowed users to distinguish the target rendered on the peripheral display with greater confidence. But first, we aimed to empirically tune three parameters of the color enhancement algorithm so that the peripheral proxy was both salient in peripheral vision and visually consistent with the real-world target.

Color Enhancement Parameter Calibration. As illustrated in Fig. 7, each trial displayed a pair of fruits of similar color on a virtual shelf, with the left fruit serving as a reference and the

right fruit as the target. During a series of parameter adjustments, participants evaluated whether the peripheral proxy matched the target fruit while remaining noticeable with peripheral vision and distinguishable from the reference fruit only rendered real-world.

At the start of each trial, the peripheral proxy of the target fruit was rendered with default parameter settings. Similar to Study 1, participants first completed IPD fit and eye tracking calibration on Quest Pro to ensure accurate gaze tracking. We then fine-tuned the parameters sequentially using a dichotomous (binary search) procedure, i.e., for each parameter, the current color enhancement setting was compared against an alternative at the opposite bound of its range (e.g., $\text{max_luminance}=1$ vs. 8). Participants indicated which rendering viewed better for the peripheral proxy, defined as more salient and more differentiable from the reference fruit while still clearly matching the target fruit. The preferred value became the new baseline, and the parameter was then compared against the midpoint of the range. This halving procedure continued until a stable setting was reached, with at most three comparisons per parameter. The final value was then fixed before moving to the next parameter. Three parameters were tuned (Fig. 8): max_luminance (upper bound on luminance increase; default 1; range 1–8), max_sat_boost (upper bound on saturation boost; default 1; range 1–16), and ab_push (Lab chromaticity shift magnitude; default 0; range 0–60).

In total, 8 participants (1 female, 7 male; ages 19–29, $M=24$, $SD=2.74$) with normal or corrected vision and no color-vision deficiencies completed the calibration. All provided their informed consent and were compensated \$20 USD for their time. We aggregated parameter settings across participants and reported the 75th percentile to derive conservative, perceptually effective bounds: $\text{max_luminance} = 2.125$; $\text{max_sat_boost} = 9.75$; $\text{ab_push} = 30.0$. These group-level values reflect settings that ensure sufficient salience without over-enhancement. We also report the grouped 75-percentile parameter values based on colors in Table 2.

Color	Max Luminance	Max Saturation Boost	AB Push
Green	2.00	6.875	15.00
Yellow	3.00	10.50	38.00
Red	3.00	10.50	28.75
Blue	3.00	6.500	38.00

Table 2: Group-level 75th percentiles for each parameter by shelf color.

6 Study 2: Peripheral Proxies

Our final study applied the color enhancement calibration values to evaluate the *screenshot* and *Most Similar Color* (MSC) strategies for PERIPHAR against a *baseline* condition. The *baseline* used a peripheral proxy with only color quantization—i.e., without color enhancement—and applied the same decimation process. To maximize control and balance visual features, we simulated real-world targets using virtual objects with a video passthrough environment similar to Study 1; however, this time, the real-world objects were chosen as representations of real-world objects in a virtual fruit market scene.

Task. Each trial presented a virtual shelf with more than 40 similarly shaped and colored virtual fruits. Participants selected two designated fruit types in a fixed alternating sequence (six of

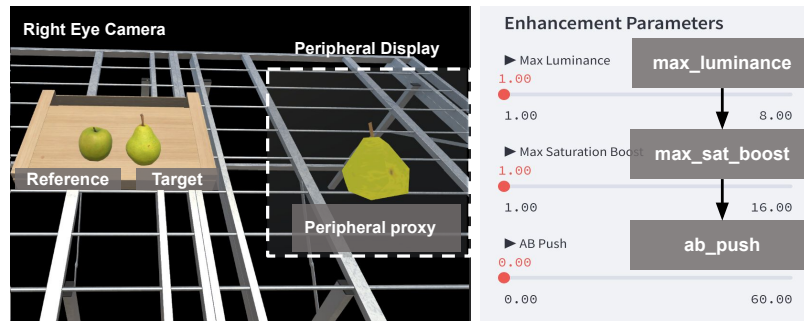


Figure 7: Calibration workflow. Left: Participants viewed a fruit shelf through left and right eye cameras; the peripheral proxy of the target fruit was rendered in the right-eye periphery. Participants assessed whether the proxy was visually consistent with the real-world target and distinguishable from the reference object. Right: We adjusted color enhancement parameters (max_luminance , max_sat_boost , ab_push) via a web interface not visible to participants. This participant–researcher loop iterated until the proxy appeared to be both salient in peripheral vision and consistent with the target, yielding empirically tuned parameter values.

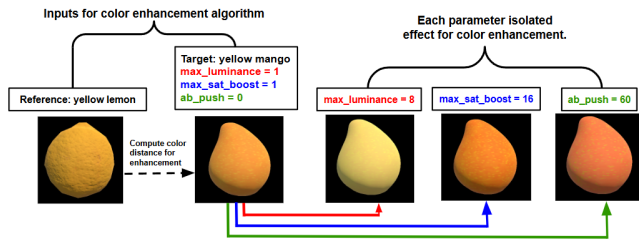


Figure 8: Each parameter’s isolated effect for color enhancement. The target fruit (yellow mango) is enhanced relative to the reference fruit (yellow lemon). From left to right: baseline target, luminance boost, saturation boost, and chroma push. Each output shows the effect of a single parameter while the others remain fixed at their default values.

each; 12 correct selections total). They were instructed to make selections as accurately as possible without foveating toward the peripheral proxy, maintaining gaze on the shelf and using peripheral vision for confirmation.

Procedure. The study used a within-subjects design with three conditions (*baseline*, *screenshot*, *MSC*) and three color-themed shelves (green, yellow, red fruits), yielding nine trials per participant. The set of fruit types associated with each shelf color (green, yellow, red), the positions of the designated fruits within each shelf, and the order of shelves were all fixed across participants. Shelf layouts were identical across conditions to maximize comparability. Only condition order was randomized, counterbalanced via a 3×3 Latin square (six groups; two participants per group). As in Study 1, participants first completed IPD fit and eye-tracking calibration to ensure accurate gaze tracking.

At the start of each trial, the two target fruit types were announced (e.g., *green apple* → *guava* → *green apple* → *guava*, etc.). For each selection, the participant (1) recalled the next fruit in sequence, (2) fixated on the corresponding fruit on the shelf, (3) received audio feedback as PERIPHAR rendered the peripheral proxy for confirmation, and (4) pressed a controller button to confirm the selection. Correct selections removed the target from the shelf

with a confirmation sound; wrong selections left it visible with an error audio cue. After completing all three shelves for a condition, participants filled out a questionnaire and participated in a brief semi-structured interview before proceeding to the next condition. See Fig. 9 for the study setup.

Participants. In total, 12 participants (3 female, 9 male; ages 21–28, $M=24.7$, $SD=1.8$) with normal or corrected vision and no color-vision deficiencies took part. All provided informed consent and were compensated \$20 USD.

Data Collection & Analysis. We logged gaze traces, gaze events, and selection events during the study session, together with post-questionnaire and interview feedback, providing objective and subjective measures. **Objective metrics** included: (1) foveal vs. peripheral gaze distribution (percentage of gaze time spent looking directly at the peripheral proxy, lower is better), (2) selection accuracy (proportion of correct selections, higher is better), and (3) completion time (average time to complete one trial, lower is better). **Subjective metrics** (7-point Likert scale, from post-condition questionnaire) included: confidence in selection, ease of noticing the peripheral proxy, perceived speed of confirmation, level of distraction (lower is better), perceived effectiveness, cognitive load (lower is better), reliance on color/shape cues of the peripheral proxy when selection, and gaze distribution (shelf vs. proxy, higher means more time on shelf). In addition, participants reported their *overall preference* by ranking the three conditions.

We securely stored anonymized data and performed scripted preprocessing and statistical analysis. Similar to Study 1, ANOVA or Friedman significance tests, with Holm or Bonferroni corrections as appropriate, were applied for objective and subjective data analysis.

Objective Performance. Table 3 shows the objective metrics (gaze distribution, selection accuracy, completion time). We found no significant effect of condition per these metrics (all $p > 0.05$).

To examine how the color context influenced errors, we aggregated error rates by shelf across all conditions (Table 4). Because yellow shelves proved to be more error-prone than green and red shelves, we focused on the yellow shelf (Table 5). Interestingly, on the yellow shelf, the *screenshot* condition produced a higher error

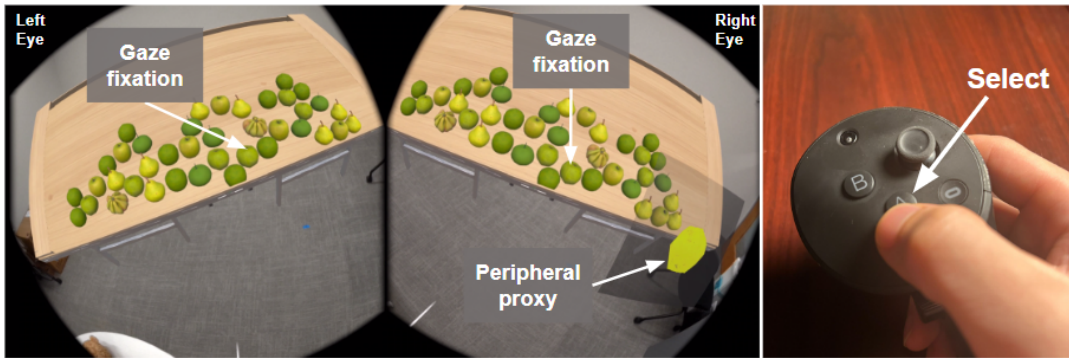


Figure 9: Study 2 setup. Left: Left/right eye views of the shelf; upon fixation on a target (e.g., green apple), a peripheral proxy is rendered for confirmation. Right: The participant commits the selection with the “A” button; correct selections remove the gazed fruit.

Metric	Base	SS	MSC
Gaze (%)	0.95	0.82	0.79
Error (%)	17.1	16.0	10.6
Time (s/trial)	115.59	91.90	97.06

Table 3: Objective metrics.

Shelf	Wrong	Total	Err.%
Green	72	504	14.3
Yellow	123	555	22.2
Red	27	459	5.9

Table 4: Error rates by shelf.

Yellow	Wrong	Total	Err.%
Baseline	41	185	22.2
Screenshot	51	195	26.2
MSC	31	175	17.7

Table 5: Yellow shelf error rates.

rate (26.2%) than *baseline* (22.2%), whereas *MSC* reduced errors to 17.7%. We then analyzed the main distractors that the participants mistakenly selected (Table 6). In both *baseline* and *MSC*, errors were spread across yellow persimmon and Asian pear with similar frequency. In contrast, errors in the *screenshot* condition were dominated by confusions with Asian pear (19 times), in the majority of cases when the intended target was *yellow apple* (Table 7).

This pattern directly reflects how *screenshot* constructs its reference image from a snapshot of the target on the yellow shelf with highly similar color distributions. The algorithm tries to enhance both the target and distractor but, in this case, against a very similar average, thus pushes in similar hue directions. This essentially reduces their perceptual distinctiveness in the peripheral display and results in insufficient enhancement. Participant feedback aligned with this interpretation. For example, P10 noted: “At the periphery, the Asian pear and yellow apple looked like uniform blocks of color, making them very difficult to differentiate [in screenshot].”

In contrast, *MSC* explicitly selects the neighbor with the most similar color as reference. For the yellow shelf, *yellow apple* and *Asian pear* were the closest pair in CIELAB space. When *Asian pear* was the target, *yellow apple* became the reference, and the algorithm enhanced the pear proxy to increase contrast. Conversely, when *yellow apple* was the target and *Asian pear* the reference, the apple was darker and therefore left unenhanced by the algorithm. This asymmetry in the algorithm produced a visible difference in the periphery. In post-condition interviews, five participants reported deliberately alternating their gaze between the two fruits to exploit this difference, noting that *MSC* made such comparisons more reliable. Across the study, most correct selections did *not* involve such deliberate comparisons: depending on shelf and condition, roughly half to two thirds of correct trials were completed with zero gaze switches between target and distractors (e.g., 57% for *baseline*, 61% for *screenshot*, and 64% for *MSC* across all shelves). This matches our design goal that PeriphAR should normally support “one-look”

confirmation. Gaze switching therefore mainly occurred in a subset of more difficult trials where participants chose to double-check ambiguous proxies. To understand these harder cases, we examined gaze switches between *yellow apple* (target) and *Asian pear* (primary distractor) on the yellow shelf, where errors were most concentrated. In the *screenshot* condition, participants made 13 mis-selections. When they incorrectly selected the *Asian pear*, they performed on average 3.64 gaze switches ($SD=4.96$), compared to 7.11 ($SD=11.45$) for correct selections. This difference was not significant (Mann–Whitney U , $p=0.097$), suggesting that additional comparisons between the two peripheral proxies did not reliably resolve their ambiguity. In the *MSC* condition, participants made only 7 mis-selections, and they were typically made with almost no comparison ($M=0.57$, $SD=0.90$), whereas correct selections involved some back-and-forth gaze switches ($M=4.64$, $SD=9.24$). These differences were significant (Mann–Whitney U , $p=0.016$). Together with the lower error rate on the yellow shelf for *MSC*, this pattern indicates that when participants chose to compare *yellow apple* and *Asian pear*, the asymmetric enhancement strategy in *MSC* produced more diagnostic peripheral differences, whereas the scene-wide reference in *screenshot* often left the two compared fruits (target and primary distractor) perceptually similar even after additional gaze switches to compare the intended target and the shown peripheral proxy (Fig. 10). Importantly, many trials in all conditions were completed without any gaze switching, so PERIPHAR is not designed to rely on deliberate back-and-forth inspection; rather, *MSC* appears to offer more useful information in those difficult comparisons.

We also examined whether target color affected completion time. The red fruit shelf were completed fastest ($M=67.78$ s, $SD=36.64$), followed by green ($M=103.90$ s, $SD=61.29$) and yellow ($M=132.98$ s, $SD=100.20$). A Friedman test revealed a significant main effect of color, $\chi^2(2)=12.67$, $p=0.0018$. Post-hoc Wilcoxon signed-rank tests with Holm correction showed that red trials were significantly faster than both yellow ($p_{\text{holm}}=0.0044$) and green ($p_{\text{holm}}=0.0098$),

Fruit Type	baseline	screenshot	MSC
Top 1	yellow persimmon (13)	Asian pear (19)	yellow persimmon (12)
Top 2	Asian pear (12)	yellow mango (11)	Asian pear (11)
Top 3	apricot (6)	yellow persimmon (10)	yellow mango (5)

Table 6: Top wrong fruit selections (yellow shelf).

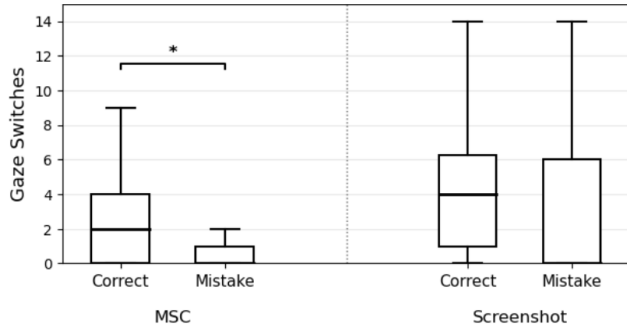


Figure 10: Gaze switches between *yellow apple* (target) and *Asian pear* (distractor) on the yellow shelf for the *MSC* and *screenshot* conditions. Each box shows the distribution of gaze switches per trial grouped by condition and outcome (correct vs. mistake). The asterisk marks a significant difference between correct and mistaken selections in the *MSC* condition (Mann–Whitney U test, $p=0.016$); no significant difference was found for *screenshot*.

whereas time differences for green and yellow were not significant (Fig. 11). The significant differences for red align with prior work on preattentive color processing, which reports that red tends to guide attention more efficiently than other hues [1, 39].

Subjective Performance. Table 8 shows subjective ratings from the post-condition questionnaires. Overall, the scores tended to favor conditions with color enhancement, particularly using the *MSC* strategy. We found significant differences for two key measures (Fig. 12): *confidence in selection* (Friedman $\chi^2(2)=9.15$, $p=0.010$; *MSC* > *baseline* $p_{\text{holm}}=0.022$) and *ease of noticing the proxy* (Friedman $\chi^2(2)=11.89$, $p=0.003$; *MSC* > *baseline*, $p_{\text{holm}}=0.020$). Participants also reported relying slightly more on shape information in the peripheral proxy under *MSC* ($M=5.42$) than under *baseline* ($M=5.08$) or *screenshot* ($M=4.58$; Table 8), but these differences were not statistically significant. This trend can be explained by considering that enhancing color relative to the most similar colored neighbor can increase local chromatic contrast along the target’s edges. As a result, *MSC*’s enhancement may sometimes have made object boundaries in the peripheral proxy easier to interpret. This is consistent with work showing that preattentive processing of color can facilitate segmentation and grouping [1, 39]. Qualitative interview feedback echoed this interpretation: P10 noted, “*The shape was more noticeable, and I could discern the fruit’s stamp direction in the peripheral cue, which helped me differentiate between similar fruits.*”, and P2 felt that “*the shape was clearer in this [MSC] condition, which helped [them] make decisions beyond just color.*”

Intended fruit	Mis-selected as Asian pear
Yellow apple	13
Yellow lemon	6
Total	19

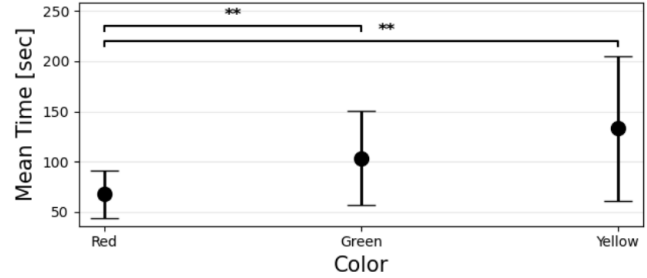
Table 7: Breakdown of Asian pear errors in the *Screenshot* condition.

Figure 11: Mean completion times across target colors in Study 2. Error bars show \pm SD. Asterisks indicate significant pairwise differences based on a Friedman test with Wilcoxon signed-rank post-hoc comparisons (Holm-corrected).

The *MSC* condition was chosen by 8 of 12 participants, *screenshot* by 3, and *baseline* by 1. A chi-square test against uniform preference was significant, $\chi^2(2)=6.18$, $p=0.045$. Pairwise two-proportion z -tests with Holm correction revealed a significant preference for *MSC* over *baseline* ($z=2.951$, $p_{\text{holm}}=0.009$).

Reflection & Takeaways. In summary, the *MSC* condition improved the subjective experience, leading to higher preference for the condition, selection confidence, and peripheral noticeability over the baseline. While there was no significant effect of condition per the objective metrics, deeper analysis showed trends in favor of *MSC*. Our error selection analysis revealed that color context mattered: on yellow shelves, *screenshot* amplified confusion with similarly colored fruits (Asian pear), whereas *MSC* was most robust across shelves. The *screenshot*’s scene-wide reference tended to homogenize color enhancement across similar objects, inadvertently amplifying confusion in cluttered contexts. Conversely, the *MSC* strategy leveraged the most similar neighbor to accentuate subtle but critical differences, yielding more reliable peripheral confirmation in visually challenging scenarios, notably the yellow shelf trials. Participant feedback suggests that color enhancement relative to the most similar object in color goes beyond improving color contrast alone, potentially strengthening peripheral shape perception as well. Finally, our analysis of completion times by target color suggests that PeriphAR inherits classic preattentive color patterns: red targets were noticed fastest, with green and yellow slower, mirroring established findings on color-guided attention [1, 39].

Metric	baseline	screenshot	MSC
Color cue reliance	5.667	5.250	5.500
Shape cue reliance	5.083	4.583	5.417
Confidence*	4.833	5.167	6.167
Ease of noticing*	5.000	5.667	6.250
Speed of confirmation	5.000	5.250	5.417
Distraction (lower, better)	1.917	1.750	1.583
Effectiveness	4.833	5.083	5.500
Gaze distribution	4.833	5.250	5.833
Cognitive load (lower, better)	3.000	2.500	2.167

*Significant difference ($p < 0.05$).

Table 8: Mean subjective ratings (1–7). Higher is better unless noted otherwise.

7 End-to-end System

As a final step, we wanted to see how PERIPHAR might perform in less controlled environments compared to our lab studies. However, we don't think of this step as a generalization attempt; rather, we simply wanted to learn if and how it could be made more effective when part of an end-to-end system and tested in potentially challenging settings with many similarly looking, cluttered physical objects (vending machines, bookshelves, etc.). To enable this exploration, we extended the system with real-time support for (1) *processing live camera video passthrough* and (2) *object detection and segmentation*, then profiled each step in the pipeline to identify where most computing resources might be needed.

Switching to Real-World Objects. First, accessing the externally facing cameras on Quest Pro is an experimental feature. The recently released Passthrough API is only available for Quest 3. We were able to exploit an Android media projection workaround to screen-grab the live video feed. Next, we needed to know where the user is looking within the passthrough to isolate the target object. However, converting the eye gaze cursor from 3D coordinates to the corresponding point in the 2D passthrough image was challenging because of barrel distortion and the eye gaze originating from the left eye alone. As a workaround, we added a red dot to indicate where the user is looking, so that the resulting 2D image also contains this red dot in the accurate location. The system can then parse the image for the center of the reddest pixels to find the 2D location of the user's gaze.

Second, when transitioning from virtual fruits to real-world objects, we needed object segmentation to isolate individual objects and grab their corresponding masks. The YOLO11n segmentation model [23] was used to maximize accuracy while simultaneously limiting latency. Once the objects are segmented, the system determines if the decoded gaze pixel resides within any of the detected object masks and if so, that object is treated as the target. The bounding box for that target is expanded in all directions to create a 2D box collider. If any other object bounding boxes intersect this expanded box, then it is treated as a neighboring object. Once the target mask and neighbor masks are gathered, the system runs the color similarity algorithm to determine which neighbor mask is most similar in color to the target mask. Once the target and reference masks are determined, a color-enhanced and quantized version of the target object is created. The enhanced image is then finally displayed in the simulated peripheral display.

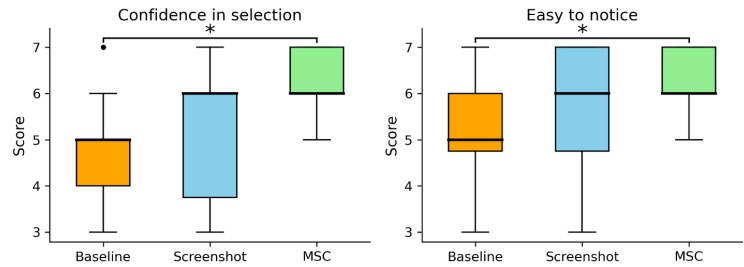


Figure 12: Box plots for Confidence in selection and Ease of noticing. Asterisks mark significant pairwise differences (Holm-corrected).

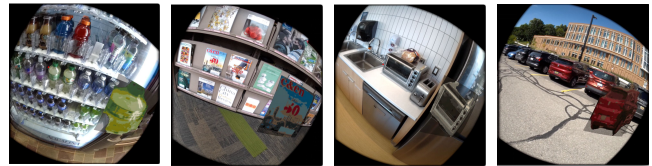


Figure 13: Examples from our end-to-end system tests showing only the right eye view with peripheral proxy on simulated peripheral display.

Testing in Realistic Settings. We then used the resulting end-to-end system for a variety of real-world tasks (Figure 13): buying a drink at a vending machine, selecting magazines and books in a library, operating a kitchen appliance, and finding a car. We treat these deployments as proof-of-concept demonstrations, intended to show that the full PeriphAR pipeline can run on current XR hardware and to obtain failure cases, rather than as a comprehensive evaluation of deployment-ready performance. We chose to not incorporate a fixation threshold on the end-to-end system, unlike the previous two studies, as applying object segmentation to the real world already takes a significant amount of time and acts as a threshold (min. time of 113ms) ensuring the user is indeed viewing a relevant object. Accompanied with the fact that per-frame fixation is observed on the server-side, as opposed to the headset, we decided that a fixation threshold would unnecessarily increase the latency between viewing a physical object and receiving the enhanced mask. Similarly, request rates between the headset and server were limited as to not overload the network.

In these examples, going through the entire process from acquiring the passthrough image with the real-world target to rendering the final color enhanced target took on average between 0.72s (parking lot, where enhancement was often skipped due to no other objects being detected) and 1.55s (kitchen appliance, with an average of 2.6 neighbors including refrigerator, sink, chair, bowl, table, oven). During profiling the various steps of the pipeline, we found that the object segmentation takes on average about 25% of the time, while the mean time to find the most similar neighbor object ranged from 17.8% (vending machine, avg. neighbors 15.6 including refrigerator, bottle) to 28.5% of the time (parking lot, avg. neighbors 2.3–3.9 including bus, car, truck, person) depending on the complexity of the scene and size of neighbors. Lastly, the quantization

and color enhancement of the image typically took a majority of the time ranging from 50.1% (vending machine) to 61.9% (kitchen).

Beyond latency, two practical aspects are important for interpreting these deployments: occlusion safety and lighting conditions. First, in the headset, the monocular peripheral display occupies only a small $\sim 20^\circ$ horizontal FOV and is rendered on a semi-transparent plane, leaving the central view unobstructed and limiting occlusion in the periphery.¹ Although the proxy appears larger in the captured video frames in Fig. 13, its apparent size in the headset is quite modest. In our end-to-end system use cases, participants used PERIPHAR while standing still or moving slowly, and the proxy was shown only in brief bursts (2 seconds) to confirm a selection. Deploying such overlays in safety-critical situations (e.g., approaching traffic or navigating uneven terrain) would require additional safeguards, such as more conservative opacity and size, or temporarily suppressing the peripheral proxy during rapid locomotion.

Second, because the peripheral proxy is rendered directly into the passthrough view, its visibility depends on both the headset display and the captured real-world lighting condition. In our tests, the end-to-end system was used mainly in indoor environments and in typical daylight outdoors (sunny but not high-glare conditions), where the Quest Pro display provided sufficient dynamic range for the proxy to remain clearly visible. We did not systematically vary ambient lighting, and in very bright sunlight or strong glare the same color enhancement would have less headroom on the physical display, reducing perceived contrast between proxy and background. A more lighting-aware version of PERIPHAR that adapts enhancement strength to ambient conditions is therefore an important direction for future work.

Limitations in Real-World Scenarios. While our algorithm performed reliably in controlled conditions, its behavior with real-world objects that contain multiple prominent color regions reveals a notable limitation. Consider the example of two *BelVita* cookie packages, Cranberry Orange (target) and Cinnamon Brown Sugar (reference), shown in Fig. 14. Both share a large yellow header region at the top of the package, while their lower halves differ in dominant hue (orange vs. red). When the color enhancement algorithm processes the full packages, it identifies the shared yellow header as the dominant overlapping region (see Fig. 14a). Because this region is nearly identical across both packages, the similarity measure falls below the skip enhancement threshold. This threshold serves as a safeguard: when two textures are already visually close, additional enhancement would either have little perceptual benefit or risk distorting the target’s appearance. By skipping enhancement in such cases, the algorithm avoids unnecessary adjustments and preserves alignment with the real-world object. As a result, the peripheral proxies remain visually similar.

This mismatch highlights an important limitation: the algorithm assumes that the most shared dominant region is the critical region for differentiation. One way to address this is to redirect the color enhancement algorithm toward the subdominant but visually distinguishing regions when dominant regions are nearly identical. As an experiment, extracting and processing only the lower halves of the packages led the algorithm to enhance successfully (see Fig. 14b), underscoring the gap between algorithmic focus and

human perceptual strategies. Our exploratory analysis of Study 2 completion times (Fig. 11) also suggests that not all colors contribute equally to fast peripheral detection: red targets were noticed significantly faster than green and yellow, echoing preattentive color findings [1, 7]. Taken together with the *BelVita* example, this points to a perceptually informed extension of our MSC strategy from object-level to patch-level enhancement. Rather than treating each object as a single texture, the system could segment it into chromatic subregions and, for each subregion, estimate (i) how strongly it differentiates the target from its most similar color neighbor (i.e., the color difference between the corresponding areas in target and neighbor, discounting large regions that are nearly identical such as the shared yellow header), and (ii) its predicted preattentive salience based on hue category (e.g., reddish/orangish accents tend to guide attention more strongly). Color enhancement could then be concentrated on subregions that both show strong target–neighbor contrast and are high-salience, instead of boosting the entire object uniformly. Such a patch-wise extension would preserve the current MSC mechanism of choosing the most similar color neighbor as reference, while making *PeriphAR* more robust for multi-colored, cluttered scenes (e.g., packaged goods), where users often rely on small distinctive regions in the periphery to disambiguate similar items. We view our *BelVita* case as highlighting where the present proof-of-concept starts to break down and as motivating these perceptually informed refinements as concrete future work toward improved generalizability in real-world deployments.

More broadly, our prototype should be interpreted as a proof-of-concept for peripheral confirmation on low-FOV AR glasses rather than a deployment-ready system. Our end-to-end system focused on relatively low-risk scenarios (e.g., standing or slow movement), not fast locomotion or safety-critical tasks, so any real-world deployment would need to incorporate explicit safety constraints on when and where peripheral proxies may appear.

8 Discussion

Opportunities & Tradeoffs with Display Simulation. The system we created to design our PERIPHAR technique provided a *quasi-realistic* simulation of future AR glasses on Quest Pro, enabling peripheral vision experiments in the absence of wide-FOV AR displays. Simulation is an important tool in HCI research to experience advanced technology [37]. We posit that future AR glasses will have to be limited in critical ways and, even if more capable with larger FOVs, users may not always want full-on AR. So, our idea of adopting *graceful degradation* as a design principle could provide a way forward.

While we believe in the practical value and applicability of our work, findings are based on simulation. We modeled display characteristics like position and FOV digitally. Without the hardware we cannot verify accuracy. Another critical aspect is that our controlled setting was free from peripheral distractions in the real world. With optical see-through displays, digital eye strain [19] will likely be more pronounced. In our studies, eye movements when switching from real-world targets to the simulated display required relatively little accommodation. Using a unifocal video passthrough display meant that the eyes’ focal length did not need to adjust as much.

¹See Section 3 for details on display emulation.



Figure 14: Effect of the color-enhancement algorithm on real-world objects with multiple color regions, illustrated using two *BelVita* cookie packages. (a) With full packages, the algorithm selects the shared yellow header as the dominant region. Since the color distance of the shared dominant region between target and reference is below the enhancement threshold, enhancement is skipped, and the target proxies remain visually similar to the reference. (b) With only the lower halves, the dominant regions (orange vs. red) differ more, exceeding the threshold and triggering enhancement, making the target proxy more distinguishable to the reference.

Context-Aware Parameter Tuning. While we selected a single global setting for our color-enhancement algorithm parameters based on group-level 75th-percentile values, the calibration results indicated distinct optimal values for different colors. For instance, participants generally required less aggressive enhancement for green fruits compared to yellow or red fruits (see Table 2). Consequently, participants with higher peripheral sensitivity occasionally found certain peripheral proxies to be overly enhanced. For example, the green apple proxy sometimes made alignment with the real-world object challenging. Future implementations should adopt context-aware parameter tuning, dynamically adjusting enhancement parameter values based on color context to ensure consistent peripheral proxy effectiveness across diverse visual scenarios.

Neighbor-Aware Enhancement Bounds. In the current *MSC* implementation, enhancement is based solely on the single adjacent object with the closest color. For example, in a green shelf trial, when participants fixated on a lighter green apple, the algorithm selected a darker green guava as the reference. This shifted the peripheral proxy of the apple brighter toward yellow-green, creating a stronger contrast with the green guava. However, this also risked enhancing the proxy to the point that it no longer aligned well with the green apple, instead shifting toward a lighter, more yellowish green that resembled the nearby green pear. This over-enhancement made it harder for participants to match the proxy to the intended target in peripheral vision. To address this limitation, future work should extend the *MSC* strategy with a neighbor-aware constraint mechanism. After selecting the most similar neighbor,

PERIPHAR could check all other adjacent objects and clamp enhancement so that the target never exceeds any neighbor. In this way, enhancement would still increase target–neighbor contrast but remain bounded by the broader visual context. Such safeguards would reduce the risk of overly bright or misleading proxies and better preserve perceptual alignment in complex, multi-object environments. These refinements highlight promising directions for improving the robustness and generalizability of peripheral color enhancement in AR-based selection tasks.

Loss of Fine-Grained Details. Although color enhancement increases overall salience, it can inadvertently obscure subtle features that aid recognition. For example, P12—the only individual who preferred the *baseline* condition—reported difficulty aligning the peripheral proxy with the real-world target on the fruit shelf. In the *baseline* condition, the proxy is only quantized, which preserves the dark striping on the tomato’s surface. After enhancement in other two conditions, however, those stripes blended into a uniform red, making the peripheral proxy appear “too flat” and harder to match to the real fruit. Future refinements should balance global color boosts with the preservation of local contrast and edge details. This would help maintain both visibility and recognizable texture cues.

Simpler Shape Representations. As our selection error analysis in Study 1 showed, targets close in hue and shapes with only minor differences led to the most errors. The decimation step prior to color enhancement in Study 2 mitigated these issues; however, color enhancement was based on the full textures of the virtual

fruits rather than only on the viewable perspective. Our end-to-end system used object masks from the segmentation based on low-res passthrough images, skipping 3D representation altogether making visual similarity dependent on viewing angle and size of region. While modern segmentation and 3D reconstruction techniques such as *Segment Anything Model* (SAM) [25] or *Neural Radiance Fields* (NeRF) [35] could perhaps be added one day, such fine-grained segmentation and photo-realistic reconstruction would be overkill. Our work calls for new techniques that abstract real-world objects by compressing their texture and shape into simple yet distinguishable visual patterns. For example, finding rough visual contours approximating an object to its primitive shape yet exaggerating certain features in the context of nearby objects, or semantically shifting colors based on keys that humans associate with particular objects, such as green for apples or yellow for bananas despite not being plain in color, could further improve glanceability.

9 Conclusion

This paper systematically developed PERIPHAR, a gaze-based feedback technique that recreates real-world targets as peripheral proxies optimized for glanceability. To do this, we first developed the key technical components to simulate future AR glasses with peripheral display modes operated in monocular fashion on Quest Pro using graceful degradation. We demonstrated that fundamental tasks like object selection can be performed efficiently and accurately with our AR display designs that used color enhancement to become peripherally glanceable. We discussed the limitations of our display simulation approach and identified opportunities for improvement of our technique, pointing toward a roadmap for advancing PERIPHAR toward real deployment in future always-on AR glasses. By shifting confirmation feedback to the user's peripheral field, our work opens a new design space for lightweight, low-powered AR glasses that align with both hardware constraints and human perceptual strengths.

References

- [1] Emil Andersen and Anja Maier. 2019. The attentional guidance of individual colours in increasingly complex displays. *Applied Ergonomics* 81 (2019), 102885. <https://doi.org/10.1016/j.apergo.2019.102885>
- [2] Parth Arora, Ethan I Kimmel, Katherine Huang, Tyler Kwok, Yukun Song, Sofia Anandi Vempala, Georgianna Lin, Ozan Cakmakci, and Thad Starner. 2025. Positioning Monocular Optical See Through Head Worn Displays in Glasses for Everyday Wear. *CoRR* abs/2505.09047 (2025). <https://doi.org/10.48550/ARXIV.2505.09047> arXiv:2505.09047
- [3] Joanna Bergström, Tor-Salve Dalsgaard, Jason Alexander, and Kasper Hornbæk. 2021. How to Evaluate Object Selection and Manipulation in VR? Guidelines from 20 Years of Studies. In *Proc. CHI*, Yoshifumi Kitamura, Aaron Quigley, Katherine Isbister, Takeo Igarashi, Pernille Bjørn, and Steven Mark Drucker (Eds.). ACM, 533:1–533:20. <https://doi.org/10.1145/3411764.3445193>
- [4] Doug A. Bowman, Ernst Kruijff, Joseph J. LaViola Jr., and Ivan Poupyrev. 2021. *3D User Interfaces: Theory and Practice* (2nd ed.). Addison-Wesley, Boston, MA, USA.
- [5] Di Laura Chen, Yuzhen Chen, Michelle Annett, and Daniel Vogel. 2023. GazeRayCursor: Facilitating Virtual Reality Target Selection by Blending Gaze and Controller Raycasting. In *Proceedings of the 29th ACM Symposium on Virtual Reality Software and Technology (VRST '23)*. Association for Computing Machinery, New York, NY, USA, 1–11. <https://doi.org/10.1145/3611659.3615696>
- [6] Kaizhan Chen, Batchuluun Duinkharjav, Nilay Ujjainkar, Ensar Shahan, Akanksha Tyagi, Jing He, and Qian Sun. 2023. Imperceptible Color Modulation for Power Saving in VR/AR. In *ACM SIGGRAPH 2023 Emerging Technologies*. 1–2.
- [7] Alexandra Clifford, Branka Spehar, and Anna Franklin. 2010. Color categories affect pre-attentive color perception. *Biological Psychology* 85, 2 (2010), 275–282. <https://doi.org/10.1016/j.biopsycho.2010.07.009>
- [8] Enrico Costanza, Samuel A. Inverso, Elan Pavlov, Rebecca Allen, and Pattie Maes. 2006. eye-q: eyeglass peripheral display for subtle intimate notifications. In *Proc. Mobile HCI*, Marko Nieminen and Mika Rökkee (Eds.). ACM, 211–218. <https://doi.org/10.1145/1152215.1152261>
- [9] Seyed Davari, Feiyu Lu, and Doug A. Bowman. 2022. Validating the Benefits of Glanceable and Context-Aware Augmented Reality for Everyday Information Access Tasks. In *2022 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, 436–444.
- [10] Stephen DiVerdi, Tobias Höllerer, and Richard Schreyer. 2004. Level of Detail Interfaces. In *Proc. ISMAR*. IEEE Computer Society, 300–301. <https://doi.org/10.1109/ISMAR.2004.38>
- [11] Batchuluun Duinkharjav, Kaizhan Chen, Akanksha Tyagi, Jing He, Yixin Zhu, and Qian Sun. 2022. Color-Perception-Guided Display Power Reduction for Virtual Reality. *ACM Transactions on Graphics (TOG)* 41, 6 (2022), 1–16.
- [12] Anna Maria Feit, Shane Williams, Arturo Toledo, Ann Paradiso, Harish Kulkarni, Shaun K. Kane, and Meredith Ringel Morris. 2017. Toward Everyday Gaze Input: Accuracy and Precision of Eye Tracking and Implications for Design. In *Proc. CHI*, Gloria Mark, Susan R. Fussell, Cliff Lampe, m. c. schraefel, Juan Pablo Hourcade, Caroline Appert, and Daniel Wigdor (Eds.). ACM, 1118–1130. <https://doi.org/10.1145/3025453.3025599>
- [13] Ajay S. Fernandes, Ken Pfeuffer, and I. Scott MacKenzie. 2025. Gaze Inputs for Targeting: The Eyes Have It, Not With a Cursor. *International Journal of Human-Computer Interaction* (2025), 1–19. <https://doi.org/10.1080/10447318.2024.2419477>
- [14] Jens Grubert, Tobias Langlotz, Stefanie Zollmann, and Holger Regenbrecht. 2017. Towards Pervasive Augmented Reality: Context-Awareness in Augmented Reality. *IEEE Trans. Vis. Comput. Graph.* 23, 6 (2017), 1706–1724. <https://doi.org/10.1109/TVCG.2016.2543720>
- [15] Uwe Gruenefeld, Tim Claudius Stratmann, Jinki Jung, Hyeopwoo Lee, Jeehye Choi, Abhilasha Nanda, and Wilko Heuten. 2018. Guiding Smombies: Augmenting Peripheral Vision with Low-Cost Glasses to Shift the Attention of Smartphone Users. In *Proc. ISMAR Adjunct*. IEEE, 127–131. <https://doi.org/10.1109/ISMAR-Adjunct.2018.00050>
- [16] Uwe Gruenefeld, Tim Claudius Stratmann, Lars Prädell, and Wilko Heuten. 2018. MonocularAR: a radial light display to point towards out-of-view objects on augmented reality devices. In *Proc. MobileHCI Adjunct*. ACM, 16–22. <https://doi.org/10.1145/3236112.3236115>
- [17] Maxime Guillon, François Leitner, and Laurence Nigay. 2015. Investigating Visual Feedforward for Target Expansion Techniques. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15)*. ACM, New York, NY, USA, 2777–2786. <https://doi.org/10.1145/2702123.2702375>
- [18] Malcolm Gibran Haynes and Thad Starner. 2017. Effects of Lateral Eye Displacement on Comfort While Reading from a Video Display Terminal. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 1, 4 (2017), 138:1–138:17. <https://doi.org/10.1145/3161177>
- [19] Teresa Hirzle, Maurice Cordts, Enrico Rukzio, and Andreas Bulling. 2020. A Survey of Digital Eye Strain in Gaze-Based Interactive Systems. In *Proc. ETRA*, Andreas Bulling, Anke Huckauf, Eakta Jain, Ralph Radach, and Daniel Weiskopf (Eds.). ACM, 9:1–9:12. <https://doi.org/10.1145/3379155.3391313>
- [20] Teresa Hirzle, Maurice Cordts, Enrico Rukzio, Jan Gugenheimer, and Andreas Bulling. 2021. A Critical Assessment of the Use of SSQ as a Measure of General Discomfort in VR Head-Mounted Displays. In *Proc. CHI*, Yoshifumi Kitamura, Aaron Quigley, Katherine Isbister, Takeo Igarashi, Pernille Bjørn, and Steven Mark Drucker (Eds.). ACM, 530:1–530:14. <https://doi.org/10.1145/3411764.3445361>
- [21] Teresa Hirzle, Jan Gugenheimer, Florian Geiselhart, Andreas Bulling, and Enrico Rukzio. 2019. A Design Space for Gaze Interaction on Head-mounted Displays. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, CHI 2019, Glasgow, Scotland, UK, May 04–09, 2019*, Stephen A. Brewster, Geraldine Fitzpatrick, Anna L. Cox, and Vassilis Kostakos (Eds.). ACM, 625. <https://doi.org/10.1145/3290605.3300855>
- [22] Nuwan Janaka, Paul Strohmeier, Dongwook Yoon, Yutaka Tokuda, Wei Sun, and Kai Kunze. 2022. Paracentral and Near-Peripheral Visualizations: Towards Attention-Maintaining Secondary Information Presentation on OHMDs during In-Person Social Interactions. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 1–18. <https://doi.org/10.1145/3491102.3517681>
- [23] Rahima Khanam and Muhammad Hussain. 2024. Yolov11: An Overview of the Key Architectural Enhancements. *arXiv preprint arXiv:2410.17725* (2024).
- [24] Manana Khomeriki and Natela Lomashvili. 2025. Visual Attention Distribution According to Size, Color, and Spatial Location of Stimuli under Foveal and Peripheral Vision Conditions. *ESI Preprints (European Scientific Journal, ESJ)* 21, 15 (2025), 22–22.
- [25] Alexander Kirillov, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tete Xiao, Spencer Whitehead, Alexander C. Berg, Wan-Yen Lo, Piotr Dollár, and Ross Girshick. 2023. Segment Anything. arXiv:2304.02643 [cs.CV]
- [26] Georgios A. Koulieris, Kemal Akşit, Michael Stengel, Rafal K. Mantiuk, Katarzyna Mania, and Christian Richardt. 2019. Near-Eye Display and Tracking Technologies for Virtual and Augmented Reality. In *Computer Graphics Forum*, Vol. 38. 493–519.

- [27] Pei-Shan Ku, Yu-Chuan Lin, Yu-Hao Peng, and Ming-Yu Chen. 2019. PeriText: Utilizing Peripheral Vision for Reading Text on Augmented Reality Smart Glasses. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, 630–635.
- [28] Jiandong Liang and Mark Green. 1994. JDCAD: A highly interactive 3D modeling system. *Computers & Graphics* 18, 4 (1994), 499–506. [https://doi.org/10.1016/0097-8493\(94\)90062-0](https://doi.org/10.1016/0097-8493(94)90062-0)
- [29] David Lindlbauer, Anna Maria Feit, and Otmar Hilliges. 2019. Context-Aware Online Adaptation of Mixed Reality Interfaces. In *Proc. UIST*. ACM, 147–160. <https://doi.org/10.1145/3332165.3347945>
- [30] Feiyu Lu and Doug A. Bowman. 2021. Evaluating the Potential of Glanceable AR Interfaces for Authentic Everyday Uses. In *Proc. VR*. IEEE, 768–777. <https://doi.org/10.1109/VR50410.2021.00104>
- [31] Feiyu Lu, Shakiba Davari, Lee Lisle, Yuan Li, and Doug A. Bowman. 2020. Glanceable AR: Evaluating Information Access Methods for Head-Worn Augmented Reality. In *Proc. VR*. IEEE, 930–939. <https://doi.org/10.1109/VR46266.2020.1581100361198>
- [32] Feiyu Lu and Yan Xu. 2022. Exploring Spatial UI Transition Mechanisms with Head-Worn Augmented Reality. In *Proc. CHI*, Simone D. J. Barbosa, Cliff Lampe, Caroline Appert, David A. Shamma, Steven Mark Drucker, Julie R. Williamson, and Koji Yatani (Eds.). ACM, 550:1–550:16. <https://doi.org/10.1145/3491102.3517723>
- [33] Kris Luyten, Dries Degraen, Gustavo Rovelo Ruiz, Sven Coppers, and David Vanacken. 2016. Hidden in Plain Sight: An Exploration of a Visual Language for Near-Eye Out-of-Focus Displays in the Peripheral View. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. 487–497.
- [34] Kris Luyten, Donald Degraen, Gustavo Alberto Rovelo Ruiz, Sven Coppers, and Davy Vanacken. 2016. Hidden in Plain Sight: an Exploration of a Visual Language for Near-Eye Out-of-Focus Displays in the Peripheral View. In *Proc. CHI*, Jofish Kaye, Allison Druin, Cliff Lampe, Dan Morris, and Juan Pablo Hourcade (Eds.). ACM, 487–497. <https://doi.org/10.1145/2858036.2858339>
- [35] B Mildenhall, PP Srinivasan, M Tancik, JT Barron, R Ramamoorthi, and R Ng. 2020. NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis. In *Proc. ECCV*.
- [36] Arsalan Mousavian, Dragomir Anguelov, John Flynn, and Jana Kosecka. 2017. 3D Bounding Box Estimation using Deep Learning and Geometry. In *Proc. CVPR*. 7074–7082.
- [37] Roderick Murray-Smith, Antti Oulasvirta, Andrew Howes, Jörg Müller, Aleks Ikkala, Miroslav Bachinski, Arthur Fleig, Florian Fischer, and Markus Klar. 2022. What simulation can do for HCI research. *Interactions* 29, 6 (2022), 48–53. <https://doi.org/10.1145/3564038>
- [38] Mark I. Nikolich and Nadine B. Sarter. 2001. Peripheral Visual Feedback: A Powerful Means of Supporting Effective Attention Allocation in Event-Driven, Data-Rich Environments. *Human Factors* 43, 1 (2001), 30–38. <https://doi.org/10.1518/001872001775992525>
- [39] Hans-Christoph Nothdurft. 1993. The role of features in preattentive vision: Comparison of orientation, motion and color cues. *Vision Research* 33, 14 (1993), 1937–1958. [https://doi.org/10.1016/0042-6989\(93\)90020-W](https://doi.org/10.1016/0042-6989(93)90020-W)
- [40] Yun Suen Pai, Tilman Dingler, and Kai Kunze. 2019. Assessing Hands-Free Interactions for VR Using Eye Gaze and Electromyography. *Virtual Reality* 23, 2 (2019), 119–131. <https://doi.org/10.1007/s10055-018-0341-3>
- [41] Mihir Parmar and Chai Silpasuwanchai. 2023. Impact of User Mobility on Attentional Tunneling in Handheld AR. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–6.
- [42] Ken Pfeuffer, Yasmeen Abdrabou, Augusto Esteves, Radiah Rivu, Yomna Abdelrahman, Stefanie Meitner, Amr Saadi, and Florian Alt. 2021. ARtention: A design space for gaze-adaptive user interfaces in augmented reality. *Computers & Graphics* 95 (2021), 1–12.
- [43] Ken Pfeuffer, Benedikt Mayer, Diako Mardanbegi, and Hans Gellersen. 2017. Gaze+ Pinch Interaction in Virtual Reality. In *Proceedings of the 5th Symposium on Spatial User Interaction (SUI '17)*. Association for Computing Machinery, New York, NY, USA, 99–108. <https://doi.org/10.1145/3131277.3132180>
- [44] Philipp A. Rauschnabel, Alexander Brem, and Young Ro. 2015. Augmented Reality Smart Glasses: Definition, Conceptual Insights, and Managerial Importance. Unpublished manuscript. Working paper, College of Business, University of Michigan–Dearborn.
- [45] Patrick Renner and Thies Pfeiffer. 2017. Attention guiding techniques using peripheral vision and eye tracking for feedback in augmented-reality-based assistance systems. In *Proc. 3DUI*, Maud Marchal, Robert J. Teather, and Bruce H. Thomas (Eds.). IEEE Computer Society, 186–194. <https://doi.org/10.1109/3DUI.2017.7893338>
- [46] Kendra Richards, Danilo Gasques, Jay Peters, and Bruce H. Thomas. 2019. Analysis of Peripheral Vision and Vibrotactile Feedback During Proximal Search Tasks with Dynamic Virtual Entities in Augmented Reality. In *Proceedings of the Symposium on Spatial User Interaction (SUI '19)*. ACM, New York, NY, USA, 1–9. <https://doi.org/10.1145/3357251.3357581>
- [47] Sajay Sadasivan, Randall L. Morrison, Michael J. Dorneich, and J. Gregory Trafton. 2005. Use of Eye Movements as Feedforward Training for a Synthetic Aircraft Inspection Task. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '05)*. ACM, New York, NY, USA, 141–150. <https://doi.org/10.1145/1054972.1054993>
- [48] Dario D. Salvucci and Joseph H. Goldberg. 2000. Identifying fixations and saccades in eye-tracking protocols. In *Proceedings of the Eye Tracking Research & Application Symposium, ETRA 2000, Palm Beach Gardens, Florida, USA, November 6-8, 2000*, Andrew T. Duchowski (Ed.). ACM, 71–78. <https://doi.org/10.1145/355017.355028>
- [49] Naveen Sendhilnathan, Ting Zhang, Ben Lafreniere, Tov Grossman, and Tanya R. Jonker. 2022. Detecting Input Recognition Errors and User Errors using Gaze Dynamics in Virtual Reality. In *Proc. UIST*, Maneesh Agrawala, Jacob O. Wobbrock, Eytan Adar, and Vidya Setlur (Eds.). ACM, 38:1–38:19. <https://doi.org/10.1145/3526113.3545628>
- [50] Gaurav Sharma, Wencheng Wu, and Edul N. Dalal. 2005. The CIEDE2000 color-difference formula: Implementation notes, supplementary test data, and mathematical observations. *Color Research & Application* 30, 1 (2005), 21–30.
- [51] Ludwig Sidenmark, Christopher Clarke, Xuesong Zhang, Jenny Phu, and Hans Gellersen. 2020. Outline Pursuits: Gaze-Assisted Selection of Occluded Objects in Virtual Reality. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–13. <https://doi.org/10.1145/3313831.3376873>
- [52] Ludwig Sidenmark and Hans Gellersen. 2019. Eye&Head: Synergetic Eye and Head Movement for Gaze Pointing and Selection. In *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology (UIST '19)*. Association for Computing Machinery, New York, NY, USA, 1161–1174. <https://doi.org/10.1145/3332165.3347945>
- [53] Hans Strasburger, Ingo Rentschler, and Martin Jüttner. 2011. Peripheral Vision and Pattern Recognition: A Review. *Journal of Vision* 11, 5 (2011), 13–13.
- [54] Xinran Sun and Amitabh Varshney. 2018. Investigating Perception Time in the Far Peripheral Vision for Virtual and Augmented Reality. In *SAP*. 13–1.
- [55] Bryan V. Syiem, Rebecca M. Kelly, Jose Goncalves, Eduardo Velloso, and Tobias Dingler. 2021. Impact of Task on Attentional Tunneling in Handheld Augmented Reality. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–14.
- [56] Erin Tanuwidjaja, Duong Huynh, Khoa Koa, Cuong Nguyen, Cheng Shao, Patrick Torbett, and Nico Weibel. 2014. Chroma: A Wearable Augmented-Reality Solution for Color Blindness. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. 799–810.
- [57] Christoph Trepkowski, Andreas Marquardt, Thies D. Eibich, Yusuke Shikanai, Jens Maiero, Kiyoshi Kiyokawa, and Peter König. 2021. Multisensory Proximity and Transition Cues for Improving Target Awareness in Narrow Field of View Augmented Reality Displays. In *IEEE Transactions on Visualization and Computer Graphics*, Vol. 28. 1342–1362.
- [58] Jo Vermeulen, Kris Luyten, Elise van den Hoven, and Karin Coninx. 2013. Crossing the Bridge over Norman’s Gulf of Execution: Revealing Feedforward’s True Identity. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '13)*. ACM, New York, NY, USA, 1931–1940. <https://doi.org/10.1145/2470654.2466255>
- [59] I-Jan Wang, Hao-Cheng Kuo, Cheng-Kai Huang, Yi-Liang Chuang, Ting-Han Fan, Yu-Ting Liu, Hao-Ting Chen, and Liwei Chan. 2024. Assistive Sensory Feedback for Trajectory Tracking in Augmented Reality. *Journal of Computing and Information Science in Engineering* 24, 3 (2024), 031001. <https://doi.org/10.1115/1.4065218>
- [60] Jeremy M. Wolfe. 2018. Visual search. In *Stevens' Handbook of Experimental Psychology and Cognitive Neuroscience*, John T. Wixted (Ed.). Vol. 2. John Wiley & Sons, Chapter 13, 1–55. <https://doi.org/10.1002/9781119170174.epcn213>
- [61] Xingyao Yu, Benjamin Lee, and Michael Sedlmair. 2024. Design Space of Visual Feedforward and Corrective Feedback in XR-Based Motion Guidance Systems. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems (CHI '24)*. ACM, New York, NY, USA, 723:1–723:15. <https://doi.org/10.1145/3613904.3642143>

A Color Enhancement Algorithm

Our color enhancement algorithm entails three steps: (1) quantization, (2) palette-level color distance analysis, and (3) masked color enhancement. Fig. 6

Quantization. In the first stage, both target and reference textures are reduced to a compact set of dominant colors. Using Mini-BatchKMeans with $K = 7$, each pixel is reassigned to its nearest centroid, yielding a quantized image and a histogram of cluster coverages. This step highlights perceptually salient palettes consistent with peripheral vision’s reduced sensitivity to detail, while also lowering computational load.

Palette-Level Color Distance Analysis. Next, we compare the palettes of the target and reference textures to determine the extent of enhancement needed for the target. We compute weighted perceptual distances across all pairs of palette centroids:

$$\Delta E_{\text{total}} = \sum_{i=1}^7 \sum_{j=1}^7 h_T(i) h_R(j) \Delta E_{00}(c_i^T, c_j^R),$$

where c_i^T and c_j^R are the i -th and j -th centroids of the target and reference palettes, $h_T(i)$ and $h_R(j)$ are their normalized cluster weights, and $\Delta E_{00}(c_i^T, c_j^R)$ is the CIEDE2000 color difference [50] between the two centroids in CIE Lab space. We use CIEDE2000 because it captures perceptual non-uniformities, providing differences that better reflect how humans perceive changes in lightness, chroma, and hue compared to simpler metrics such as Euclidean distance. Analogous computations yield ΔL for lightness and ΔC for chroma differences. Together, these metrics quantify how distinct the two textures are in overall appearance.

For enhancement preparation, we also identify a *shared dominant color*, chosen from the reference palette but also prominent in the target. This color acts as an anchor, marking the overlap where confusion is most likely in peripheral vision. By centering enhancement around this shared color, the algorithm selectively increases contrast where target and reference would otherwise appear too similar. The global distance metrics (ΔE_{total} , ΔL , ΔC) together with the shared dominant color form the basis for constructing the boost map in the next stage.

Masked Color Enhancement. In the final stage, enhancement is applied selectively to regions of the target associated with the shared dominant color. A similarity mask is computed by measuring the CIEDE2000 distance from each target pixel to this color. Scaled by the global perceptual distance ΔE_{total} , the mask produces a *boost map* $B(x, y)$ that localizes enhancement to areas where target and reference overlap most strongly. This ensures adjustments increase perceptual contrast precisely where peripheral confusion is most likely. While ΔE_{total} controls overall enhancement strength, the relative contributions of ΔL and ΔC determine whether emphasis falls on lightness or chroma.

We capture this weighting by

$$\alpha = \frac{\Delta C}{\Delta L + \Delta C},$$

where α biases enhancement toward chroma when ΔC dominates and toward luminance when ΔL dominates. Saturation is applied as

a general vividness boost, independent of α . With this weighting, the three channels are adjusted as follows:

- **Luminance:** brightness is lifted by scaling Lab lightness:

$$L'(x, y) = L(x, y) [1 + B(x, y)(\text{max_luminance} - 1)(1 - \alpha)],$$

where max_luminance caps the maximum increase. Gamma correction and CLAHE (Contrast Limited Adaptive Histogram Equalization) preserve highlight detail and local contrast.

- **Saturation:** HSV saturation is boosted by

$$S'(x, y) = S(x, y) [1 + B(x, y)(\text{max_sat_boost} - 1)\alpha],$$

with max_sat_boost controlling the maximum gain. Saturation is weighted by α so that it contributes more when chromatic differences (ΔC) dominate, and less when luminance differences (ΔL) are the primary factor. This ensures that saturation lifting is emphasized only when it is most perceptually useful.

- **Chroma push:** Lab chroma values (a, b) are shifted outward from the reference mean μ_{ab} :

$$(a', b')(x, y) = (a, b)(x, y) + B(x, y) \cdot \text{ab_push} \cdot \frac{(a, b)(x, y) - \mu_{ab}}{\|(a, b)(x, y) - \mu_{ab}\|_2},$$

where ab_push sets the maximum directional shift. Unlike saturation, chroma push operates independently of α , because its role is to explicitly separate the target’s hue from the reference. This guarantees local color contrast even when overall luminance differences are dominant.

The result is a color-enhanced target image that is brighter, more vivid, and chromatically distinct from the reference image, allowing participants to associate the peripheral proxy with the correct real-world target at a glance.