

What to Distinguish and How? Opportunities and Challenges of Augmenting Multiple, Cluttered Objects in Complex Scenes for People with Low Vision

Yuheng Wu

Department of Computer
Sciences
University of
Wisconsin-Madison
Madison, WI, USA
yuheng.wu@wisc.edu

Ruijia Chen

Department of Computer
Sciences
University of
Wisconsin-Madison
Madison, WI, USA
ruijia.chen@wisc.edu

Jaewook Lee

Paul G. Allen School of
Computer Science &
Engineering
University of Washington
Seattle, WA, USA
jaewook4@cs.washington.edu

Jia Li

University of Texas at
Dallas
Richardson, TX, USA
jia.li@utdallas.edu

Kexin Zhang

Department of Computer
Sciences
University of
Wisconsin-Madison
Madison, WI, USA
kzhang284@wisc.edu

Meng Fong Lio

Department of Computer
Sciences
University of
Wisconsin-Madison
Madison, WI, USA
mengfong@wisc.edu

Weibing Wang

Department of Computer
Sciences
University of
Wisconsin-Madison
Madison, WI, USA
wwang652@wisc.edu

Sanbrita Mondal

School of Medicine and
Public Health
University of
Wisconsin-Madison
Madison, WI, USA
smondal4@wisc.edu

Jon E. Froehlich

Paul G. Allen School of
Computer Science &
Engineering
University of Washington
Seattle, WA, USA
jonf@cs.uw.edu

Yapeng Tian

Department of Computer
Science
University of Texas at
Dallas
Richardson, TX, USA
yapeng.tian@utdallas.edu

Yuhang Zhao

Department of Computer
Sciences
University of
Wisconsin-Madison
Madison, WI, USA
yuhang.zhao@cs.wisc.edu

Abstract

People with low vision (PLV) struggle to perceive complex scenes like busy kitchens and crowded streets, which contain many objects, visual clutter, and dynamic elements. Prior AR systems for low vision either enhance low-level visual features or augment task-relevant objects for single tasks in simple settings, leaving multi-object augmentation in complex scenes underexplored. Informed by a formative study characterizing important objects and their perceived importance for PLV, we built SceneGlance, a wearable AR system that recognizes important objects and visually distinguishes them by importance level. Through a controlled lab study with 12 PLV in a mock-up kitchen scene and a free-form think-aloud study with 13 PLV navigating an outdoor route, we found that AR distinction on object importance shifted PLV's attention toward objects of higher importance, and supported perception strategies such as building mental snapshots from the augmentation distribution and hierarchical scanning by importance. However, this attention shift came with a tradeoff of reduced overall scene recall. The studies also surfaced challenges posed by AR augmentations in complex

scenes, such as adjacent augmentations blending or interfering with each other, yielding design implications for more practical AR vision enhancement systems in the complex real world.

CCS Concepts

• **Human-centered computing** → **Accessibility systems and tools**; **Accessibility technologies**; **Mixed / augmented reality**.

Keywords

Accessibility, Augmented Reality, Low Vision, Vision Enhancement

ACM Reference Format:

Yuheng Wu, Ruijia Chen, Jaewook Lee, Jia Li, Kexin Zhang, Meng Fong Lio, Weibing Wang, Sanbrita Mondal, Jon E. Froehlich, Yapeng Tian, and Yuhang Zhao. 2026. What to Distinguish and How? Opportunities and Challenges of Augmenting Multiple, Cluttered Objects in Complex Scenes for People with Low Vision. In *The 28th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS '26)*, October 25–28, 2026, Vila Nova de Gaia, Portugal. ACM, New York, NY, USA, 25 pages. <https://doi.org/10.1145/3797867.3828974>

1 Introduction

Complex scenes—characterized by a high quantity and variety of objects, high visual clutter, complex spatial layout, and the presence of dynamic, unpredictable objects [10, 54, 59, 76, 82, 92]—are ubiquitous in everyday settings like a busy kitchen or crowded street.



This work is licensed under a Creative Commons Attribution 4.0 International License. *ASSETS '26, Vila Nova de Gaia, Portugal*
© 2026 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-2521-0/2026/10
<https://doi.org/10.1145/3797867.3828974>

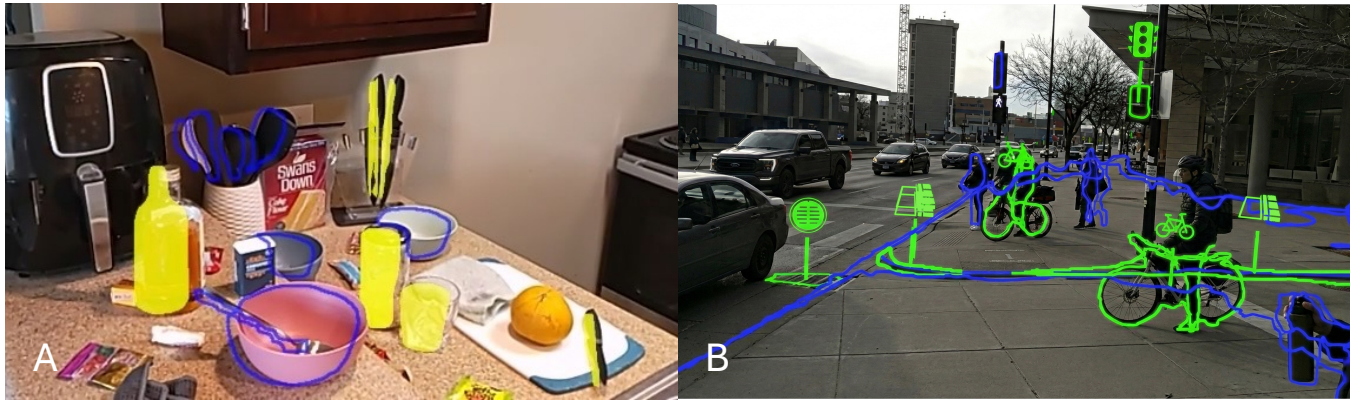


Figure 1: We explore the opportunities and design challenges of augmenting and distinguishing multiple objects in complex scenes to support scene perception for people with low vision, using SceneGlance as a technology probe. (A) SceneGlance distinguishes multiple objects in a complex kitchen, with primary-important objects (glasses, knives) overlaid in solid yellow, and secondary ones (bowls, spoons, ladles) in a blue outline. (B) SceneGlance distinguishes multiple objects on a crowded street, with primary-important objects (curb, sewer drain, pedestrian signal, bicycles) in a green outline with floating icon labels, and secondary ones (sidewalk, pedestrian) in a blue outline only.

For people with low vision (PLV) who experience uncorrectable vision loss [79], these scenes pose significant challenges in daily tasks, such as finding products on crowded store shelves [48, 98] or avoiding hazards on busy streets [78, 118].

Prior research has explored using augmented reality (AR) systems to enhance PLV's visual abilities [123]. Conventional AR systems enhance scene visibility by augmenting low-level visual features such as edges and contrast [41, 125]. However, they also amplify details unnecessary for scene interpretation, which may create visual clutter and cognitive overload in complex environments [1, 25]. To address this, recent research has developed AR systems that assist PLV with specific tasks by augmenting task-relevant objects, such as a target product for grocery shopping [126], or graspable and hazardous parts of kitchen tools for cooking [58]. However, these systems were only evaluated in relatively simple settings with few augmented objects and low visual clutter.

In contrast, perceiving complex scenes requires quickly scanning and identifying multiple important objects simultaneously while filtering out distractors [37, 120]. To cope with this challenge, the human visual system selectively allocates attention to a subset of important objects in the scene for more efficient perception [37, 70, 110, 117]. However, this process can be challenging for PLV as their vision loss hinders fast identification of important objects. We seek to support this selective attention by automatically detecting and augmenting important objects. However, in complex scenes with many important objects, simply augmenting all of them equally can cause visual clutter and fail to provide guidance on where to look. Therefore, we propose **AR distinction**, a method that visually distinguishes objects of different importance through different AR augmentations, thus enhancing important objects while reducing visual overwhelm in complex scenes.

To understand what objects are important for PLV in complex scenes and how to augment and distinguish them, we conducted a formative study with six PLV in two complex scenarios: *cooking in a busy kitchen* and *navigating a crowded street* [7, 60, 78, 98]. The study

characterized three categories of important objects (safety-related, visually challenging, frequently used) and two factors determining their importance levels, and derived guidelines for AR augmentation and distinction. Informed by these findings, we developed **SceneGlance**, a head-mounted AR system that combines RGB and depth information to recognize, locate, and augment multiple important objects in 3D space with AR distinction. To understand how SceneGlance supports PLV in complex scene perception and what challenges arise, we investigate:

RQ1: How do AR augmentation and distinction support and change PLV's perception of complex scenes?

RQ2: What are the design challenges and implications when augmenting visually complex scenes with AR for PLV?

To answer these questions, we evaluated SceneGlance in two scenarios: an indoor mock-up kitchen representing a complex, cluttered, static scene, and an outdoor scenario that supplements it with dynamic objects, reflected in two studies (Figure 2). We first conducted a well-controlled lab study with 12 PLV perceiving a mock-up kitchen countertop to understand the impact of AR distinction on PLV's viewing strategy and mental map development. The study revealed that, compared to perceiving without augmentation, SceneGlance effectively shifted participants' attention toward objects of higher importance. Beyond attention shift, AR distinction helped PLV perceive complex scenes by enabling them to build a scene snapshot from augmentation distribution and providing visual guidance for scanning. However, augmenting many objects reduced overall object recall compared to perceiving without augmentation, revealing an attention-recall tradeoff from multi-object augmentation in dense, cluttered scenes. Our study also revealed new perception challenges of applying AR augmentations to complex, cluttered scenes, such as partially occluded and visually similar objects, adjacent augmentations blending in, and spatial misalignment between objects and augmentations. We further conducted a free-form think-aloud study with 13 PLV navigating outdoor streets using SceneGlance, revealing unique challenges and needs, such

as augmenting continuous ground surfaces (e.g., sidewalks) and dynamic objects (e.g., pedestrians, cyclists). Our studies also uncovered participants' changing preferences on AR distinction designs between indoor and outdoor environments due to lighting conditions. We finally derive design implications for future AR vision enhancement systems in the complex real world.

2 Related Work

Our work builds on prior research on the challenges of perceiving complex scenes for people with blindness or low vision, as well as AR-based visual augmentation systems for low vision.

2.1 Challenges of Perceiving Complex Scenes for People with Blindness or Low Vision

Many key daily activities require the ability to perceive complex scenes, such as locating ingredients in a crowded kitchen or walking down a bustling street. A complex scene refers to an environment containing numerous objects with diverse visual features that increase the difficulty of perception and interpretation [31, 82]. Prior work has characterized scene complexity as a multidimensional concept, influenced by the number and variety of objects [10, 82], the degree of visual clutter [54, 76, 92], the presence and temporal unpredictability of dynamic objects [59], and the global scene layout, including symmetry, openness, and the spatial distribution of scene elements [54, 82]. These dimensions pose challenges for everyday visual activities even for sighted individuals, such as slowing visual search [36, 80, 92, 100], impeding object recognition [87, 112, 116], and reducing scene recall [3].

Perceiving complex scenes requires recognizing multiple key objects and understanding their spatial relationships [37, 70, 110, 117]. However, this is particularly challenging for PLV as scene complexity compounds their visual challenges. For example, visual clutter disproportionately impairs object recognition for people with central vision loss, as it is difficult for the peripheral visual system to clearly separate the target object's visual features from the surrounding distractors [103, 112]. Moreover, for people with peripheral vision loss, navigating complex environments with many objects requires high mental effort, as they must intentionally decide where to direct their gaze without the visual preview afforded by peripheral vision [106]. These challenges obstruct everyday activities, such as navigating crowded store aisles or locating products among cluttered shelves during shopping [44, 48, 98, 105], locating and distinguishing visually similar kitchen items in a complex kitchen [7, 60, 115], and noticing moving hazards (e.g., pedestrians, trolleys, cars) [43, 98, 118] or identifying wayfinding cues (e.g., doors, intersections) [43, 78, 121] in busy navigation spaces. These challenges significantly impact the independence, quality of life, and well-being of PLV [18, 35, 44, 77, 102].

To address these challenges, prior work has developed assistive technologies that support scene understanding for people with blindness or low vision through non-visual cues. Some systems leverage object recognition models on users' camera feeds to recognize nearby objects and announce them via spatial audio [17, 21, 49, 50]. However, in complex scenes with many objects, announcing each individual object can cause information overload. Instead of simply enumerating objects, more recent work uses

multi-modal large language models (LLMs) to interpret users' real-time camera streams and generate high-level scene descriptions [11, 14, 33, 34, 51, 71, 72, 74, 88, 90]. To further reduce cognitive overload in complex environments, VIPTour [63] distills object information based on aesthetics, freshness, and basic needs into a simplified hierarchical representation that users can selectively explore on smartphones via gestures. NaviNote [14] similarly lets users access location-based audio annotations of their surroundings on demand to avoid information overload.

Despite the potential of supporting complex scene perception, existing tools focus on non-visual feedback for blind users, overlooking the needs of PLV who still rely heavily on their residual vision for daily activities and prefer visual enhancement tools over audio cues [98, 99]. It remains unexplored how to enhance PLV's visual perception of complex environments with a large number of cluttered objects (sometimes visually similar or even moving) while minimizing visual and cognitive load. To fill this gap, we investigate the unique design opportunities and challenges of AR-based visual augmentations on complex scenes for people with low vision.

2.2 AR-based Visual Augmentation Systems for Low Vision

AR systems can support PLV by directly enhancing users' vision through real-time visual enhancements [123]. Conventional AR systems enhance scene visibility by augmenting the user's full field of view using image processing techniques, such as magnification [97, 125, 127], edge enhancement [41, 53, 125], contrast enhancement [125], scene recoloring based on depth information [5, 39, 109], and pixel remapping for field of view loss [67, 93, 94, 122]. For example, Hwang et al. [41] designed an edge enhancement system on Google Glass that increases edge contrast to help people with age-related macular degeneration view their environments. Zhao et al. developed *ForeSee* [125], a head-mounted system rendering magnification, edge enhancement, and contrast enhancement to support PLV with diverse visual conditions. While promising, these systems augment all low-level visual features equally, including details irrelevant to the user's task, potentially increasing visual clutter and distraction in complex scenes [1, 25]. For example, in a complex outdoor scene, an edge enhancement system may improve the visibility of cars by enhancing their outer edges, but it also highlights irrelevant details like cloud texture and road surface, causing visual overload [25].

More recent AR systems support PLV in specific daily activities by selectively augmenting task-relevant objects with visual cues [13, 26, 27, 40, 56–58, 124, 126]. For example, Zhao et al.'s *CueSee* [126] facilitates visual search by enhancing the target object with five types of visual augmentations (e.g., a flashing outline on the object, a guideline pointing at the object); Lee et al.'s *CookAR* [58] supports kitchen tool interaction by rendering green overlays on graspable areas (e.g., knife handle) and red overlays on hazardous areas (e.g., blade) to indicate tool affordances; Lee et al.'s *ARSports* [57] assists PLV in playing tennis and basketball by rendering colored overlays on sport-related objects (e.g., basketball, tennis racket, player) to help detect and locate them faster; and Chen et al.'s *VisiMark* [13] supports indoor navigation by rendering icon labels at landmarks along the route and signboard overviews at intersections to help

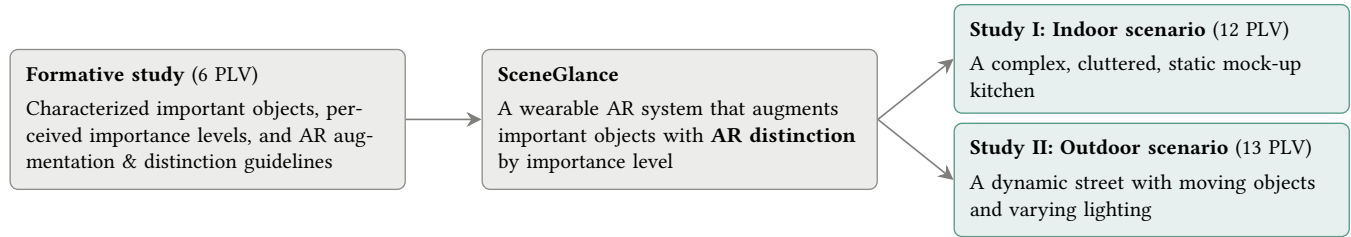


Figure 2: Research method overview. A formative study with six PLV characterized important objects, their perceived importance levels, and AR augmentation and distinction guidelines. The results informed the design of SceneGlance, a wearable AR system that augments important objects with AR distinction by importance level. We evaluated SceneGlance in two complementary scenarios: Study I, a complex, cluttered, static indoor kitchen, and Study II, a dynamic outdoor street, each examining how AR distinction supports complex scene perception and the design challenges arising in this context.

PLV identify landmarks and preview hallway structure. Beyond ocular low vision, Gamage et al. [27] extended AR augmentation to cerebral visual impairment (CVI), co-designing augmentations with two CVI adults for daily difficulties such as locating a target in cluttered scenes and guiding attention toward salient objects.

While these systems augment only task-relevant objects instead of the entire scene to reduce visual distraction, systems like *CookAR* [58] and *ARSports* [57] were evaluated in relatively simple settings with a small number of augmented objects and low visual clutter. *CueSee* [126] was similarly evaluated on a mock grocery store shelf with a structured layout (e.g., products sequentially placed with no occlusion or clutter) and only one augmented target product at a time. *VisiMark* [13] augmented more than one landmark along indoor hallways, but it was evaluated using simple routes with only 4–5 landmarks per route, and the landmarks were far from each other without introducing visual crowdedness.

In complex scenes, new challenges may arise for AR systems, such as visual clutter from augmentations [101] and occlusion between augmentations and real-world objects [85], which may pose additional challenges for PLV due to their visual impairments. However, to our knowledge, no research has thoroughly investigated how AR-based visual augmentations can support complex scene interpretation. Our paper seeks to address this problem by exploring different AR object augmentation and distinction techniques in visually complex, cluttered scenes for PLV, revealing unique design challenges and implications for this context.

3 Formative Study

To inform the design of SceneGlance, we first conducted a formative study with six PLV to understand: (1) objects that are important for PLV to perceive in complex scenes; (2) the importance level of these objects (i.e., whether all objects are equally important, and which objects are more important); and (3) PLV’s design preferences for augmenting and distinguishing different important objects with AR in complex scenes.

3.1 Method

3.1.1 Participants. We recruited six participants (P1–P6; age: $M = 66$, $SD = 13.83$; two male, four female) through contact lists from local low vision clinics and communities. Participants were eligible

if they were at least 18 years old, had low vision, and used functional vision in daily tasks (e.g., using magnifying glasses to read). Table 5 in Appendix B shows their demographics, visual conditions, and prior AR experience. Participants covered a broad range of low vision conditions, such as central vision loss (P1, P6), peripheral vision loss (P2–P5), and severe low visual acuity (P5). Three participants (P2, P3, P5) were legally blind (i.e., could not read the 20/100 line after best correction, or had field of view narrower than 20 degrees) [4]. Two participants (P2, P5) had prior AR experience but did not use it regularly. Participants were compensated \$25 per hour. This study was approved by our university IRB.

3.1.2 Procedure. We conducted a single-session two-hour interview and observational study at each participant’s residence to understand their daily experience. We focused on two representative scenarios, *cooking in a busy kitchen* and *navigating a crowded street*, as complex, challenging indoor and outdoor tasks, respectively [91]. We began the study by asking about participants’ demographic information, visual conditions, and prior AR experience. Then, we demonstrated a HoloLens-based AR prototype that recognized a water bottle and augmented it with bright yellow outlines to help participants fully understand the concept of *object augmentation*.

Understanding Important Objects and Importance Level. With a better understanding of AR augmentations, we discussed important objects and their importance levels in kitchen and street environments. In the kitchen scenario, participants first showed us their kitchens and described all objects they considered important to augment and explained their rationales. We further probed them by showing an egocentric video¹ of various cooking tasks (e.g., cutting, transferring food) in complex kitchen environments [7, 46, 60], extracted from the Epic-Kitchens-100 dataset [16]. While watching the video, participants were asked to imagine themselves cooking and “think aloud,” describing what objects they considered important to augment and why. The video was played on a large display, and participants could pause and zoom in to examine certain frames as needed. Finally, participants rated the importance level of each important object they mentioned on a five-point scale (1 means least important for augmentation; 5 means most important for augmentation) and explained their rationales.

¹<https://drive.google.com/file/d/15MMxi4EwVKo-6bSzYB-ExHGP042ZR5T9/view?usp=sharing>

For street navigation, we shadowed participants walking along two blocks in their neighborhoods. The routes were pre-selected to ensure they included sufficient potentially important or challenging objects (e.g., curb, pole, traffic sign) [96, 121]. During the navigation, participants “thought aloud” to describe important objects they encountered and why. After the walk, we further probed them with an egocentric video of navigating on a busy street² and asked them to point out objects they preferred to augment. Participants then reviewed their list of important objects for street navigation and rated importance levels for each object on the same five-point scale.

AR Augmentation and Distinction Co-Design. Finally, we conducted a probe-based co-design to understand participants’ preferences for augmenting and distinguishing objects of different importance. We probed participants using slides by presenting four base augmentations for individual objects, adapted from prior research on low vision augmentation [13, 58, 126]: (1) a static outline around each object, (2) a solid overlay, (3) a flashing outline, and (4) a text label floating above each object with its name (Figure 10 in Appendix A). We illustrated each design on two kitchen example images and two street view example images to provide participants with sufficient context. Participants commented on the effectiveness, preferences, and suitable scenarios for each design and brainstormed other design ideas.

We then investigated how participants prefer to distinguish objects of different importance levels. We probed with four AR distinction methods (i.e., using different AR augmentation designs to distinguish different objects): (1) by form—solid overlay for high-importance objects vs. static outline for low-importance objects; (2) by color—bright yellow outline vs. dark blue outline; (3) by visual effect—flashing outline vs. static outline; and (4) by additional visual information—static outline and text label combined vs. static outline only (Figure 11 in Appendix A). Participants viewed these AR distinction designs on our example images, critiqued each method, and brainstormed preferred design ideas.

Following thematic analysis [8], two researchers independently open-coded two transcripts (33% of the data) to develop an initial codebook upon agreement. One researcher coded the remaining transcripts and derived themes on the characteristics of important objects, importance levels, and preferred AR augmentation and distinction designs.

3.2 Findings

Our formative study characterized important objects, factors influencing their importance levels, and design guidelines for augmenting and distinguishing multiple objects in complex scenes, extending prior work that identified important objects for PLV without distinguishing their perceived importance [42].

3.2.1 Characterizing Important Objects. We identified three categories of important objects: *safety-related*, *visually challenging*, and *frequently used* objects.

Safety-related objects. All participants considered safety-related objects as important, including both hazards and risk indicators. Hazards included *tripping hazards* such as curbs and sewer drains (P1–P6), *cutting hazards* such as knives and scissors (P1, P2, P5, P6),

and *burning hazards* such as stovetop surfaces and carafes (P1–P6). Beyond hazards, all participants found objects that indicated potential risks or safety boundaries to be important, such as the indicator lights on stoves (P1), crosswalks (P1, P2, P4–P6), pedestrian signals (P1–P6), and sidewalk boundaries (P1, P2, P4–P6).

Visually challenging objects. All participants considered visually challenging objects as important to augment, including low-contrast objects (P1–P6; e.g., a black spatula on a dark counter, unclear crosswalk lines), transparent objects (P1–P6; e.g., glasses), cluttered objects (P1–P4, P6; e.g., a stack of dishes), objects of similar shape (P2, P3, P5, P6; e.g., spoons and forks), and small objects (P1, P3, P5; e.g., small clips, pedestrian signals).

Frequently used objects. Five participants (P1–P3, P5, P6) considered frequently used objects to be important to augment, such as light switches (P3), utensils in the kitchen (P2), bus stops (P1, P3, P5), and traffic signs (P1–P3, P5).

3.2.2 Factors that Impact Object Importance Levels. We identified two factors influencing the perceived importance level of objects, including risk severity and visual difficulty.

Risk Severity. All participants considered objects that posed higher safety risks to be more important to augment. Cutting and burning hazards were typically rated as high-importance due to the potential for severe injuries (P1–P6). For tripping hazards, lower objects that people may trip over tended to be of higher importance, while larger obstacles that people may bump into tended to have lower importance (P1, P3–P6). Beyond hazards, risk indicator objects also had different importance levels based on their associated risk severity (P1–P6). For example, the stove indicator lights had higher importance as they indicated severe burning risks (P1). Similarly, crosswalks, pedestrian signals, tactile domes on curb cuts, traffic signs, and construction cones were rated as high-importance as they all indicated highly risky mobility tasks (e.g., crossing a street or tripping over construction obstacles; P1–P3, P5, P6). In contrast, sidewalks were considered less important due to their milder consequences (P1, P2, P4, P5).

Visual Difficulty. Four participants (P1, P3, P5, P6) considered visual difficulty a factor that affects object importance, with objects harder to perceive (e.g., low contrast, transparent) being more important to augment (P3, P5).

3.2.3 Design Guidelines for Object Augmentation. We summarize participants’ preferences for augmenting important objects and distinguishing importance levels via three guidelines (DG 1–3).

DG 1: Only use dynamic augmentations for crucial objects. Participants considered all four designs effective, including *static outline* (P1–P6), *solid overlay* (P1–P3, P6), *flashing outline* (P1, P3), and *text label* (P1, P2, P5). However, when augmenting multiple objects in a complex scene, all participants found the flashing augmentation visually distracting. They suggested that such dynamic augmentations should only be used for highly crucial and risky objects that require immediate attention, such as an approaching vehicle (P3) or construction cones (P5).

DG 2: Balance between visibility and visual overload. Four participants (P1, P2, P4, P6) considered the solid overlay visually obvious but also overwhelming when augmenting too many objects (P1–P3, P5). They preferred the static outline for less visual blockage. Similarly, four participants (P2–P5) considered the text label

²https://drive.google.com/file/d/1ASs8-ER3fscb_8lhTyX9ejRliixCwr9e/view?usp=sharing

overly complex if rendered on multiple objects. P6 thus suggested augmenting objects with icon labels to reduce visual clutter.

DG 3: Differentiate object importance on multiple visual dimensions. When augmenting important objects, five participants (P1–P5) considered it necessary to distinguish objects based on their importance levels. Participants praised all four distinction methods and three participants (P1, P2, P5) suggested combining multiple designs (e.g., color and visual effect) to better distinguish object importance. For example, P1 and P5 suggested augmenting the most important objects using flashing outlines in a bright color (e.g., yellow) and less important ones using static outlines in a darker color (e.g., blue) for stronger visual contrast.

4 SceneGlance

To demonstrate the concept of AR distinction for complex scenes, we designed and developed SceneGlance, a wearable AR system that detects and visually distinguishes objects of different importance levels through AR augmentations. We describe its design and implementation below.

4.1 Augmentation Design

Following the design guidelines from Section 3.2.3 and one pilot study, SceneGlance supports three base augmentation designs and three AR distinction methods to distinguish objects of different importance. We detail the designs below.

4.1.1 Base Augmentations. We refined the four initial base augmentation designs following the design guidelines (DG 1–3). Specifically, we replaced the text labels with icons to reduce visual complexity while maintaining necessary visual information (DG 2). While PLV liked the flashing effect for highly important objects (DG 1), we removed this design because it became indistinguishable from the static outline when object detection was unstable (the outline flickered due to occasional mis-detection). As a result, SceneGlance offers three base augmentations for important objects: (1) **Static Outline**, a bright contour around the object to enhance its visibility and minimize visual blockage; (2) **Solid Overlay**, a semi-transparent colored overlay over the object matching its exact shape; and (3) **Icon Label**, a floating icon of the object and a line connecting the icon to the object center. The icons are selected to be visually recognizable and easily understandable.

4.1.2 AR Distinction Methods. Based on the factors (i.e., risk severity, visual difficulty) influencing object importance (Section 3.2.2), SceneGlance categorized objects into three categories: (1) *primary-important* objects requiring immediate attention, (2) *secondary-important* objects needing subtle augmentations, and (3) *non-important* objects needing no augmentations. Following the pilot study, we removed the *By Visual Effect* method along with the flashing outline augmentation, ending with three distinction methods: (1) **By Form**, which assigns different base augmentations to different importance levels (e.g., *Solid Overlay* for primary-important objects vs. *Static Outline* for secondary ones; Figure 3A); (2) **By Color**, which assigns different colors when applying the same base augmentation to different importance levels (e.g., bright yellow outline for primary-important objects vs. darker blue for secondary ones; Figure 3B); (3) **By Additional Visual Information**, which adds additional

information to primary-important objects (e.g., primary-important objects are augmented with both *Static Outline* and *Icon Label* vs. secondary ones with *Static Outline* only; Figure 3C).

4.1.3 Customization & Combination. To accommodate users' different preferences and visual needs (DG 3), SceneGlance allows flexible customization and combination of the AR distinction methods. Color and opacity can be customized for all designs. Following prior research on PLV's perception of commercial AR devices [123], we used yellow as the default color for its high visibility and offered green, blue, red, white, and cyan as options. Users can also adjust the thickness of *Static Outline* (1 to 10 pixels) and the size of *Icon Label* (from 50×50 to 240×240 pixels in 10-pixel steps).

Users can also combine multiple AR distinction methods to further enhance the visual differences between objects. For example, they can combine the *By Form* and *By Color* methods to augment primary-important objects with a yellow solid overlay and secondary ones with a blue outline.

4.2 Prototype Implementation

We prototyped SceneGlance by fine-tuning object recognition models for important objects and implementing the system on Microsoft HoloLens. We describe each system component below.

4.2.1 Detection of Important Objects: Dataset Refinement and Model Finetuning. Based on the formative study (Section 3.2), we determined a list of important objects and their importance levels (primary vs. secondary-important) for the kitchen and street navigation scenarios to enable object detection. In the kitchen scenario, we included 11 object classes: three *safety-related* classes (*carafe*, *knife*, *scissors*), two *visually challenging* classes (*glasses*, *jars*), and six *frequently used* ones (*bowl*, *cup*, *fork*, *ladle*, *spatula*, *spoon*). Among them, we considered *knife*, *scissors*, *carafe*, *glasses*, *jars* as primary-important and the rest as secondary-important.

In the outdoor scenario, we covered 21 object classes. All objects were considered *safety-related* (*curb*, *curb cut*, *fence*, *crosswalk*, *pedestrian*, *bench*, *sewer drain*, *fire hydrant*, *utility box*, *mailbox*, *lamp pole*, *construction cone*, *pedestrian signal*, *trash can*, *bicycle*, *motorcycle*, *railway track*, *sidewalk*, *vehicle*, *traffic sign*, *vehicle signal*). Meanwhile, one also belonged to the *visually challenging* category (*pedestrian signal*) and two belonged to the *frequently used* category (*traffic sign*, *vehicle signal*). We excluded most visually challenging objects (e.g., small objects, low-contrast objects, low-lighting objects) as they were based on low-level visual features instead of high-level objects. Among these objects, we considered *curb*, *curb cut*, *crosswalk*, *sewer drain*, *construction cone*, *pedestrian signal*, *traffic sign*, *bicycle*, *motorcycle*, *railway track*, and *vehicle* as primary-important and the rest as secondary-important.

Based on the important object list, we constructed datasets to enable detection. For the kitchen scenario, we built the *Kitchen-Importance* dataset of 20,052 images by combining and relabeling MS-COCO [64] and the Kitchen Affordance dataset [58]. For outdoors, we constructed a dataset of 20,000 images, named *Street-Importance*, by relabeling the Mapillary Vistas dataset [81].



Figure 3: An illustration of the three AR distinction methods in SceneGlance. (A) By Form: e.g., solid overlay vs. static outline, (B) By Color: e.g., yellow vs. blue, (C) By Additional Visual Information: e.g., static outline & icon labels vs. static outline only.

We fine-tuned an object segmentation model on the two datasets to detect important objects in each scenario. We selected the *RTMDet* model [68] (the *RTMDet-Ins-l* variant) due to its balance between recognition accuracy (mAP=0.437 on MS-COCO) and inference speed (33 FPS on an NVIDIA 4070 GPU). We initialized the model with pre-trained weights on MS-COCO and fine-tuned the model by freezing the backbone and only training the recognition head. We refer to the fine-tuned models as *RTMDet-Ins-l-Kitchen* and *RTMDet-Ins-l-Street*, respectively, and evaluate them in Section 4.3.2.

4.2.2 System Framework. SceneGlance consists of a frontend on Microsoft HoloLens that senses the surrounding environment and renders object augmentations, and a backend on a remote server that receives the real-time video stream from HoloLens and recognizes important objects using the fine-tuned models. The video was streamed from HoloLens to the server using the *hl2ss* library [19], modified to support transmission via User Datagram Protocol (UDP). The recognition results and computed AR augmentation were serialized using *Protocol Buffer* [29], a cross-language protocol that supports compact data compression, and streamed to the frontend via UDP for augmentation rendering.

To accurately position the augmentations in 3D space, SceneGlance leveraged the RGB video input for 2D object localization and the environmental mesh captured by HoloLens for depth estimation. For each recognized object, we converted its 2D coordinates in image space into HoloLens’ camera space using camera intrinsics from HoloLens’ Research Mode API [107]. We then cast a ray from the camera position toward the object center and rendered the augmentation at the point where the ray collided with the environmental mesh. Augmentations were scaled by distance to match the actual object size. We rendered augmentations using the *RawImage* API in Unity 2022.3.25f1 [108].

4.3 Technical Evaluation of the System

We performed a technical evaluation of SceneGlance on two aspects: real-time performance and recognition accuracy.

4.3.1 Real-Time Performance. We ran SceneGlance continuously for three minutes and measured the frame rate and mean latency of each stage in the pipeline (Section 4.2.2). Specifically, video streaming from HoloLens to the backend ran at 29.78 FPS with a 120 ms latency; model inference and post-processing ran at 28.56 FPS with a 35 ms latency; result streaming from the backend to HoloLens ran

at 28.56 FPS (as it was capped by model inference) with a 10 ms latency; and object localization and rendering ran at 28.56 FPS with an 11 ms latency. Overall, SceneGlance ran at 28.56 FPS with a 176 ms per-frame latency, reaching near real-time performance. We offset the latency by reprojecting augmentations based on the user’s pose at the time of capture, so that the perceived visual delay of static object augmentations was 11 ms; however, this technique does not apply to moving objects (e.g., pedestrian in outdoor navigation).

4.3.2 Recognition Accuracy. We evaluated the recognition accuracy of the two fine-tuned models described in Section 4.2.1 on the test subsets of our kitchen and street object datasets. For each dataset, we used a random train/validation/test split of 80%/10%/10%. We used three common metrics [83]: average precision at a 50% Intersection over Union (IoU) threshold (AP@50), at a 75% IoU threshold (AP@75), and mean average precision (mAP), defined as the mean of average precision across IoU thresholds from 50% to 95% with step 5%. These IoU-based metrics holistically assess how accurately the predicted masks align with the ground truth [24].

We compared the two fine-tuned models against the baseline model, the *RTMDet-Ins-l* model pretrained on *MS-COCO* without fine-tuning. The results (Table 1) showed that our fine-tuned models outperformed the baseline model across all three metrics, reaching mAPs of 0.435 (Kitchen) and 0.324 (Street). Figure 5 shows examples of the fine-tuned models performing on test images. Our models demonstrated strong robustness in complex scenes, maintaining high accuracy in busy kitchen and street environments from an egocentric perspective. The models were also robust against partial occlusion (e.g., a spoon in a bowl).

Model	mAP	AP@50	AP@75
RTMDet-Ins-l (baseline)	0.190	0.313	0.200
RTMDet-Ins-l-Kitchen	0.435	0.683	0.477
RTMDet-Ins-l (baseline)	0.133	0.221	0.133
RTMDet-Ins-l-Street	0.324	0.558	0.337

Table 1: Recognition accuracy of the fine-tuned and baseline models. Both fine-tuned models outperformed the baseline.

4.3.3 False Negatives and False Discoveries. We analyzed the false negative rate (indicating the rate of unaugmented important objects) and the false discovery rate (indicating the rate of wrongly augmented non-important objects) [83] of the two fine-tuned models. Across object categories, the *RTMDet-Ins-l-Kitchen* model has a false negative rate of 29.3% and a false discovery rate of 29.7%; the *RTMDet-Ins-l-Street* model has a false negative rate of 42.3%

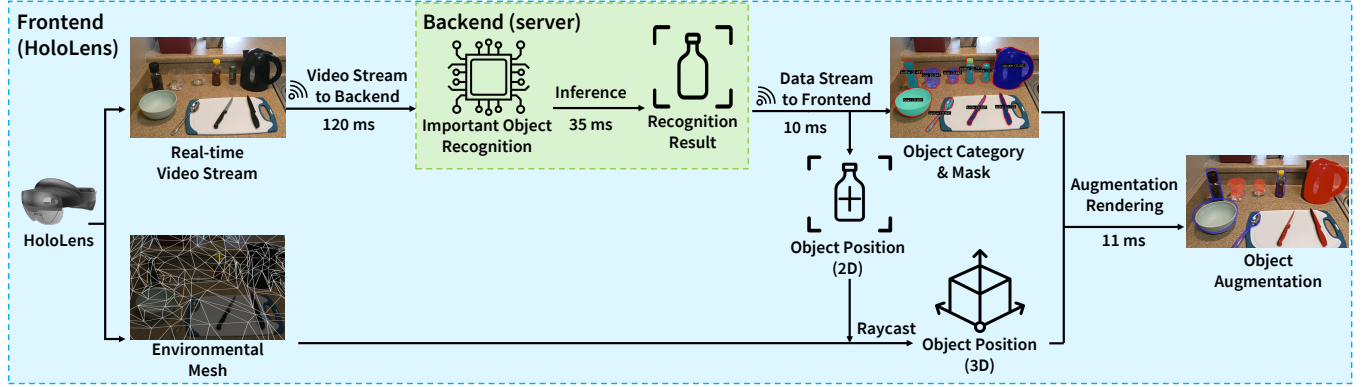


Figure 4: System pipeline of SceneGlance: the HoloLens (frontend) streams video to the backend; the backend runs the fine-tuned models to recognize and segment important objects and streams the result to the HoloLens; the HoloLens receives each object’s 2D position, raycasts it onto the 3D environmental mesh to locate the 3D position, and renders augmentations. We label each stage’s latency in the figure. The overall latency is 176 ms and the frame rate is 28.56 FPS.

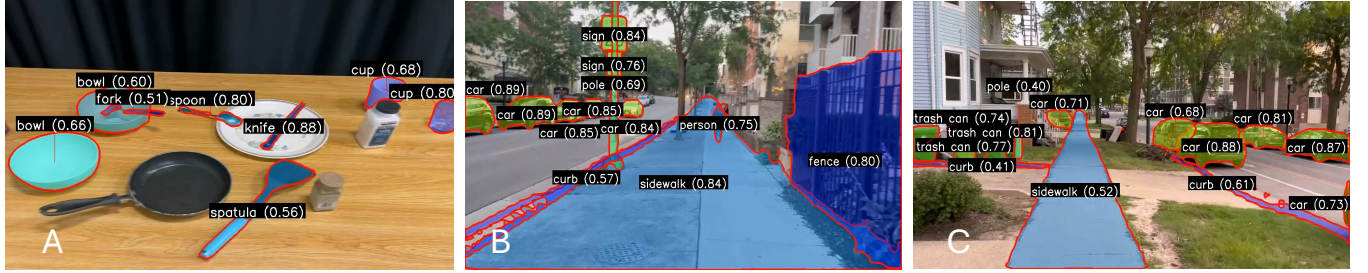


Figure 5: Example inference results of the two fine-tuned models on test images in the kitchen (A) and outdoor environment (B, C). These examples illustrated the models’ robustness in complex environments against partial occlusion.

and a false discovery rate of 32.2%. We report per-category rates in Tables 6 and 7 in Appendix C.

In both scenarios, the two error types were associated with different visual characteristics. In the kitchen scenario, false negatives were most common for transparent objects, such as glasses (39.4%) and jars (37.7%), echoing the visually difficult objects identified in the formative study (Section 3.2). False discoveries were most common for objects that are visually similar to other kitchen items, such as ladles (39.4%), bowls (36.5%), and scissors (34.9%). In the outdoor scenario, false negatives were most common for surfaces that lack distinct visual features, such as curb cuts (89.0%) and crosswalks (79.1%), and for objects that vary widely in visual appearances, such as fences (73.1%). False discoveries were most common for objects that are visually similar to other common obstacles in the scene, such as sewer drains (51.1%), utility boxes (42.6%), and benches (39.5%).

5 Study I: Using SceneGlance in a Complex Mock-up Kitchen

We first evaluated PLV’s complex scene perception with and without SceneGlance in a kitchen environment, understanding how AR distinction designs change people’s viewing strategy and mental map development. We also used SceneGlance as a technology probe

to uncover the unique perception challenges, strategies, and preferences of PLV viewing AR-enhanced complex scenes, shedding light on practical AR system designs for complex reality.

5.1 Methods

5.1.1 Participants. We recruited 12 participants with low vision (R1-R12) through contact lists from local low vision clinics and communities. A participant was eligible for the study if they were at least 18 years old and had best-corrected visual acuity between 20/100 and 20/800 in the better eye. Table 8 in Appendix D details participants’ demographic information. Participants were aged between 19 and 73 ($M = 48.75$, $SD = 18.68$), including six males and six females. All participants except R2, R4, and R10 were legally blind. R1 and R3 experienced peripheral vision loss, and six participants (R2, R4, R7, R10–R12) experienced central vision loss. Six participants (R1, R4, R9–R12) had prior experience with AR, but none used it regularly. Participants were compensated \$25 per hour. The study was approved by our university IRB.

5.1.2 Apparatus. We set up a mock-up kitchen countertop with a size of $150\text{ cm} \times 60\text{ cm}$ in a well-lit lab space. To simulate a complex kitchen environment, we prepared 12 categories of kitchen objects, including four primary-important (knives, scissors, glasses, jars), three secondary-important (bowls, spoons, forks), and five

non-important (snacks, fruits, cutting board, sponge, dish cloth) following the formative study.

With these kitchen objects, we generated six table layouts with similar amount and variability of objects to achieve similar scene complexity for the six trials (details in Section 5.1.3). Each layout contained 30–33 objects, with 10 or 11 objects from each object importance category (i.e., primary-important, secondary-important, and non-important). The objects were randomly placed on the table, and utensils (i.e., forks, spoons, knives) were either placed on the table or in containers (i.e., bowls, glasses). Figure 6A shows an example of the table layout and Figure 6B shows a participant’s view of the kitchen countertop with objects augmented by SceneGlance.

5.1.3 Procedure. The study consisted of one session lasting three hours, where participants conducted scene perception tasks in three conditions: (1) *Reality baseline*: not wearing an AR device but using their own vision with best correction (e.g., eyeglasses); (2) *AR baseline*: wearing the AR glasses with all important objects augmented by the same base design; and (3) *SceneGlance*: wearing SceneGlance that distinguished primary- and secondary-important objects with AR distinction designs. We describe the study details below.

System Tutorial and Customization. We started by collecting participants’ demographics, visual conditions, and prior AR experience. Then, we prepared a tutorial scene with 31 objects across all 12 categories randomly placed on the table, and demonstrated SceneGlance. We first walked through all three base augmentations. For each design, participants freely customized the augmentation parameters (e.g., color, opacity, outline thickness, icon label size) and discussed what they liked and disliked about the design. Participants then chose their preferred augmentation if only one type of augmentation could be used on both primary- and secondary-important objects (for the AR baseline). Next, we demonstrated the three AR distinction designs and asked about participants’ feedback and preferences. Participants then freely explored the tutorial scene using SceneGlance with their preferred designs until they felt fully familiar with the system and its augmentations.

Scene Perception Tasks. Participants then completed scene perception tasks on the kitchen countertop. We contextualized them in a scenario where they would cook on the kitchen countertop and needed to familiarize themselves with the table sufficiently. They could proceed at their own pace and walk around the table as needed. Once participants felt sufficiently familiar with the scene, they verbally described it without looking at the table, and drew a mental map of all objects they remembered. We also asked which objects caught their attention first and why.

Participants completed the tasks under three conditions: the Reality baseline, the AR baseline, and SceneGlance, with two trials each, totaling six. After the two trials in each condition, participants reflected on their experience and perceived effectiveness of scene perception under that condition. We counterbalanced the order of the conditions using Latin Square and randomly mapped the six table layouts to the six trials for each participant.

Exit Interview. We concluded the study with an interview where participants compared their scene perception experience across the three conditions. Participants also discussed whether two importance levels were sufficient and brainstormed augmentation designs to better distinguish object importance.

5.2 Data Analysis

We collected both quantitative and qualitative data from the study and analyzed them as follows.

5.2.1 Quantitative Analysis. We evaluated participants’ mental maps with two metrics: (1) *Object Recall*, defined as the proportion of correctly recalled objects among all objects present on the table, to reflect how well participants perceived the scene under each condition; (2) *Recall Ratio*, defined as the proportion of correctly recalled objects at each importance level among all correctly recalled objects, to reflect how participants allocated their attention across importance levels. We computed *Object Recall* at five levels: overall, all important objects (i.e., primary- and secondary-important combined), primary-important, secondary-important, and non-important objects. We computed *Recall Ratio* at four levels: all important, primary-important, secondary-important, and non-important objects. We had nine measures in total.

For all measures, we had one within-subject factor, *Condition*, with three levels: *Reality baseline*, *AR baseline*, and *SceneGlance*. To validate the counterbalancing, we included a between-subjects factor *Order*, and found no significant effect of *Order* on any measures. We applied the Shapiro-Wilk normality test [95] to all nine measures. Five measures (overall recall, important object recall, primary-important recall, secondary-important recall, and secondary-important ratio) were normally distributed, and four measures (non-important recall, important ratio, primary-important ratio, and non-important ratio) were not. If a measure was normally distributed, we fitted a Linear Mixed-Effects (LME) model and computed the ANOVA table to evaluate the effect of *Condition* [52, 73, 114], and used Tukey’s HSD for *post-hoc* comparison if significance was found [104]. For non-normal measures, we used Aligned Rank Transform (ART) ANOVA [119] to test significance and ART-C [22] for *post-hoc* comparison. Since we ran tests on nine measures, we applied Bonferroni correction to the significance threshold ($\alpha = 0.05/9 = 0.0056$) [6, 20]. We calculated the effect size using partial eta squared (η_p^2), with 0.01, 0.06, and 0.14 as the thresholds of small, medium, and large effects, respectively [15].

Moreover, we categorized the first object noticed in each trial by importance level and summarized the distribution across the three conditions in Table 2. We collected 24 valid responses for each of the Reality baseline and SceneGlance conditions. In the AR baseline, two participants (R1, R6) each reported in one trial that no object stood out at first glance as they felt confused by the augmentations, resulting in 22 valid responses. To test whether the distribution of first-noticed objects differed across conditions, we applied the Pearson chi-square test [86] followed by pairwise chi-square tests with Bonferroni correction [20].

5.2.2 Qualitative Analysis. We video-recorded all studies. All recordings were transcribed using an automatic transcription service and manually revised by the research team. We analyzed the transcripts using thematic analysis [8, 9]. Two researchers independently open-coded three participant transcripts (25% of the data) to develop an initial codebook upon agreement. Then, one researcher coded the remaining data based on the codebook. When a new code emerged, researchers discussed it and added it to the codebook upon agreement. Codes were grouped into themes and sub-themes guided

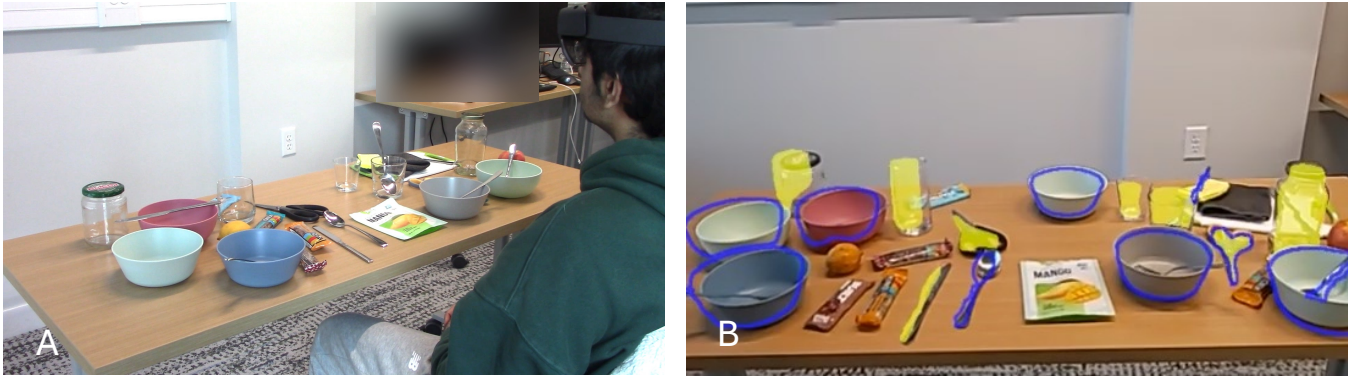


Figure 6: Kitchen Countertop Perception Task. (A) Participants sat at a mock-up kitchen table with 30–33 kitchen objects, contextualized in the scenario that they were going to cook on this table and needed to familiarize themselves with the table sufficiently for cooking. (B) Participants’ view in SceneGlance where primary-important objects (glasses, jars, knives, scissors) were augmented in semi-transparent yellow solid overlay, and secondary-important objects (bowls, spoons, forks) augmented in blue outline. Non-important objects (snacks, cutting board, sponge, dish cloth, fruits) were not augmented.

by our research focus, including perception challenges of complex scenes for PLV, effects of augmentation and distinction on attention and scene understanding, and design challenges of AR augmentations in complex scenes.

5.3 Findings

We report findings on how PLV perceived the complex scene under the three conditions, revealing their viewing strategy and shifts in attention allocation with and without SceneGlance. We found that SceneGlance shifted attention toward primary-important objects but reduced overall scene recall, revealing an attention-recall trade-off in augmenting complex scenes. Our study also revealed unique challenges of applying AR augmentations to complex, cluttered scenes, providing critical design insights on *what objects to augment and distinguish* and *how to effectively distinguish them* for future AR systems. We detail our findings below.

5.3.1 Visual Strategies for Perceiving Complex Scenes. The crowded kitchen countertop with over 30 objects posed significant perceptual challenges. Without any augmentations, participants adopted several strategies to perceive and organize the scenes.

Linear scanning & grouping. Most participants (8/12; e.g., R1, R7) scanned the scene linearly from one end to the other, sometimes repeating the scan multiple times. As R7 described: “With bare eyes, I was kind of just going left to right multiple times in a row to get a good scan of the whole table.” During linear scanning, four participants (R2, R8, R11, R12) grouped nearby objects of the same category and memorized object groups instead of each individual object. As R2 described: “Everything was very compartmentalized. Four glasses on the right side, three bowls on the back, one in the center left-ish.”

Anchor-based mental map construction. Four participants (R4, R6, R9, R11) actively looked for certain objects as “anchors” and developed mental maps for the rest of the scene relative to them. As R9 explained: “I’ve kind of been using [bowls, cups, and jars] as anchors for remembering the table [...] in my head I just made a map: okay, the bowls are here, the cups are here, the jars there.

And then I try to remember what’s in them.” Participants chose anchor objects with certain characteristics: R6 and R9 selected large objects (e.g., bowls, a large pack of snacks) as they were visually salient, R11 chose objects that were unique in the scene (e.g., the only cutting board) because “there aren’t multiples of it, so it’s easier to remember exactly where those items are,” and R4 focused on breakable objects (e.g., glasses, jars) as he was concerned about knocking them over. This anchor-based mental model provides insights into what should be augmented in complex scenes. We discuss this further in Section 7.1.

Relating to everyday activities. Three participants (R5, R10, R12) interpreted the scene by relating it to everyday activities, such as a table setting for a meal (R5, R12) or a kitchen around mealtime (R10). As R10 described: “I just see bowls, almost like someone’s ready to eat or they just finished eating. I see the cutting board, it looks like it had some action with the scissors and the knife.” R12 further said that this kind of association guided his attention and mental model. For example, he focused on a bowl with a utensil on the table since he imagined it as part of a meal: “It almost felt like I was at a restaurant or at home, and it could [look like] food [...] right in the middle of the table and a utensil in it. Some of those patterns or schemas are easy to recognize and immediately where some of my attention goes.”

Despite these strategies, it remained mentally demanding for some participants (R9–R12) to perceive the complex scenes because they lacked visual guidance on which objects to pay attention to first. As R9 explained: “I feel like [without augmentations] I was looking at more stuff. I didn’t have a clear priority of what to look at, you know? [With SceneGlance] it was like: okay, I remembered the utensils first, and I felt like that was what I needed to remember, whereas here I’m looking at everything all at once.”

5.3.2 AR Importance Distinction Reshaped Attention Allocation. We report how SceneGlance affected attention allocation in the complex scene. Compared to the Reality baseline, SceneGlance effectively shifted participants’ attention toward primary-important objects, reflected in both object recall ratio and first-noticed object

Condition (# trials)	Primary-important	Secondary-important	Important (Primary + Secondary)	Non-important
Reality baseline (24)	5 (20.8%)	8 (33.3%)	13 (54.2%)	11 (45.8%)
AR baseline (22)	10 (45.5%)	7 (31.8%)	17 (77.3%)	5 (22.7%)
SceneGlance (24)	19 (79.2%)	2 (8.3%)	21 (87.5%)	3 (12.5%)

Table 2: Distribution of first-noticed objects across the three conditions. Each participant reported their first-noticed objects in two trials per condition. In the AR baseline, two participants (R1, R6) each reported in one trial that nothing stood out at first glance, resulting in 22 valid responses.

distribution. We also summarize qualitative feedback on how the AR distinction supported PLV’s perception of complex scenes.

Attention allocation across conditions. We found a significant effect of *Condition* on the recall ratio of primary-important objects ($F_{2,58} = 6.11, p = 0.004, \eta_p^2 = 0.17$). Both the AR baseline ($diff = 13.75, p = 0.018$) and SceneGlance ($diff = 15.56, p = 0.006$) showed a significantly higher recall ratio than the Reality baseline. No significant difference was found between SceneGlance and the AR baseline ($p = 0.927$). No significant effect was found for the recall ratio of all important objects ($F_{2,58} = 3.60, p = 0.034 > 0.0056$ with correction) or secondary-important objects ($F_{2,58} = 0.31, p = 0.737$).

We also found a significant effect of *Condition* on the distribution of first-noticed objects ($\chi^2(4) = 17.45, p = 0.002$). Pairwise comparisons showed that SceneGlance significantly differed from the Reality baseline ($p < 0.001$), with participants first noticing primary-important objects in 79.2% of the trials with SceneGlance in contrast to 20.8% in the Reality baseline (Table 2). The AR baseline fell in-between with no significant difference from SceneGlance ($p = 0.149$) or the Reality baseline ($p = 0.426$). These results indicated that SceneGlance effectively shifted participants’ attention toward primary-important objects that they most desired to see.

Visual strategy changes. Participants reported that the importance distinction of SceneGlance helped them better perceive the complex scenes compared to the AR baseline. Augmenting all important objects equally did not fully address the high mental demand of perceiving complex scenes. Six participants (e.g., R1, R9) reported that, with over 20 objects augmented in the same way in the AR baseline, they still felt confused about where to look first. As R1 said: “The fact that everything was highlighted in the same color, so there’s highlights over here, there’s highlights there [...] where do I look first?” In contrast, the importance distinction from SceneGlance enabled participants to adopt new visual strategies to address such disorientation issues. We describe their new visual strategies below.

Mental snapshot from augmentation distribution. Before perceiving individual objects, participants used the distribution of different augmentations to establish a rough impression of the scene (R4, R6, R10). As R4 reflected: “[With the AR distinction] there’s a lot of icons over here but not a lot of icons over here, [it] helps understand more of a general danger zone [...] spatially remembering where more things that are fragile versus not [...] having that differentiation, it’s a nice snapshot when you assess the scene: okay, there’s a high concentration of glass over here and over here, but in this space there’s not as much.”

Hierarchical scanning with visual guidance. Using more pronounced augmentations for more important objects provided visual guidance for participants to start from primary-important objects before moving to secondary- and non-important ones (10/12; e.g., R5, R8).

Since the AR distinction divided the scene into smaller object groups by importance levels, participants could focus on one group at a time instead of processing all augmented objects at once (in the AR baseline), which was easier to manage mentally (6/12; e.g., R5, R7). As R5 said: “Everything wasn’t the same. There was a difference. You know, [with the AR distinction] I looked and tried to figure out [primary-important objects]: oh, these are jars, these are glasses, these are scissors, okay, did all that first. The next thing, okay, now what’s the other highlighting. It’s in blue. Let me check, oh, those are the bowls [...] So instead of overwhelming me with everything all [augmented equally] at once, it gave me a chance to sort it out.”

5.3.3 New Perception Challenges Posed by AR Augmentations. While SceneGlance shifted participants’ attention toward primary-important objects (Section 5.3.2), our study also revealed new barriers posed by AR augmentations to complex scene perception. We found a significant effect of *Condition* on overall recall ($F_{2,58} = 5.74, p = 0.005, \eta_p^2 = 0.17$), where participants correctly recalled significantly fewer objects in both the AR baseline ($diff = -0.088, p = 0.005$) and SceneGlance ($diff = -0.079, p = 0.013$) than the Reality baseline, with no significant difference between SceneGlance and the AR baseline ($p = 0.944$). This decline was mainly driven by non-important objects, recalled significantly less under both AR conditions than the Reality baseline ($F_{2,58} = 6.76, p = 0.002, \eta_p^2 = 0.19$). No significant effect was found for important objects overall ($F_{2,58} = 2.34, p = 0.105$), primary-important objects ($F_{2,58} = 0.51, p = 0.605$), or secondary-important objects ($F_{2,58} = 5.12, p = 0.009 > 0.0056$ after correction).

These results suggest an *attention-recall tradeoff* of the multi-object AR augmentations—shifted attention toward primary-important objects with reduced recall of non-important objects—indicating potential challenges posed by AR augmentations in cluttered scenes. We found that some augmentations amplified existing visual difficulties, such as distinguishing partially occluded or visually similar objects, while augmenting multiple objects further introduced new challenges, such as adjacent augmentations clustering into confusing shapes and spatial misalignment of icon labels. We detail these challenges below.

Amplifying confusion from partially occluded objects. One main challenge PLV faced was perceiving partially occluded objects, such as a utensil inside a bowl or scissors behind a bowl (8/12; e.g., R2, R9). In the cluttered kitchen, such occlusion was common, as nearby objects could hide distinguishing features (e.g., fork tines, scissor handles), making occluded objects hard to identify. For example, R2 misidentified the scissors as a knife as they were partially occluded by a bowl and only the blades were visible. On top of this challenge, the Static Outline and Solid Overlay only augmented the visible part, which could even amplify the confusion (8/12; e.g., R4,

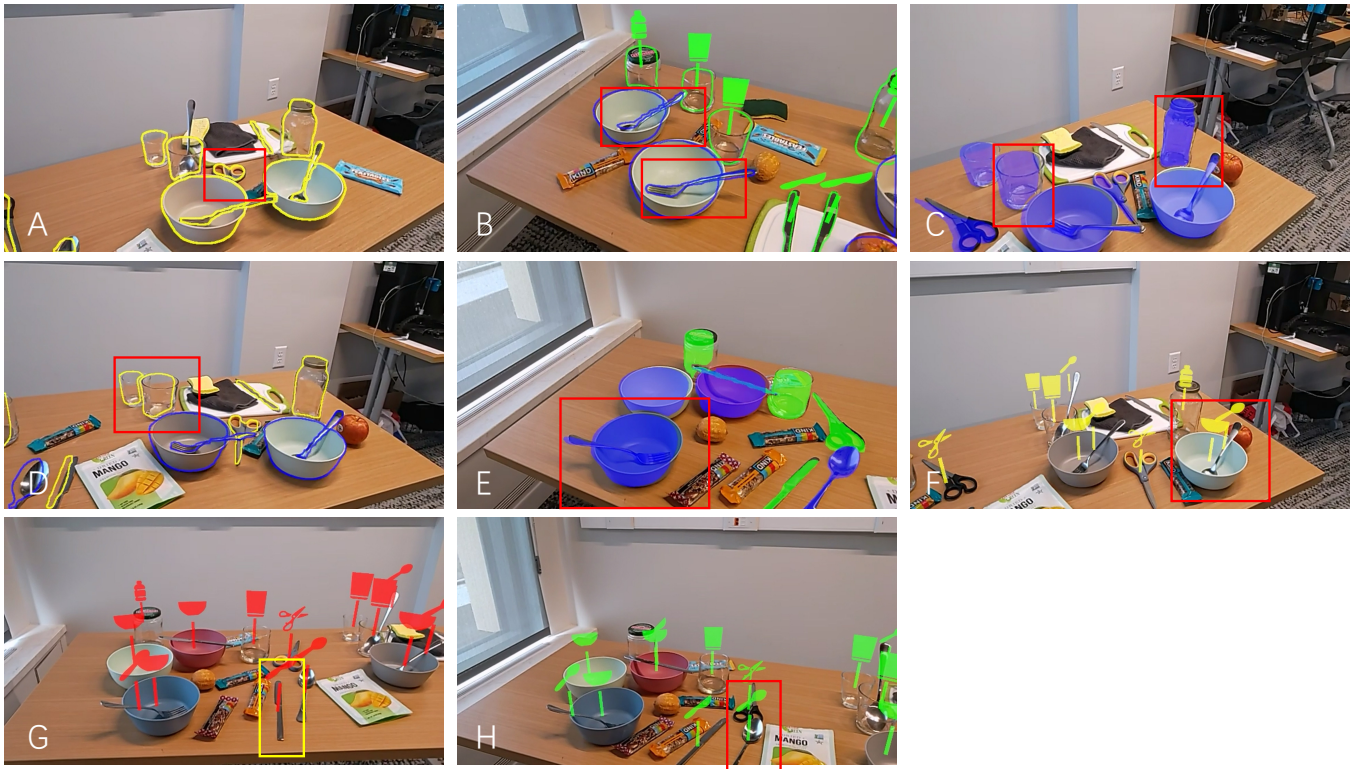


Figure 7: Examples of perception challenges in Study I. (A) *Partial occlusion*: the outline of the scissors handle was misidentified as a “heart,” as the outline did not reveal the full object. (B) *Visual similarity*: outlines of a fork and spoon looked similar because outlines did not preserve fine details such as fork tines. (C) *Solid overlay hindering distinguishing similar objects*: the solid overlay covered the lid of a jar, making it indistinguishable from a glass. (D) *Adjacent augmentations merging*: outlines of two adjacent glasses merged and appeared like a “pitcher.” (E) *Solid overlays merging*: solid overlays of a bowl and a fork inside it merged into a single shape looking like “a bowl with a handle.” (F) *Icon spatial misalignment*: icon labels of a bowl and a spoon inside it could not convey their spatial relationship (containment). (G) *Icon label occluding the augmented object itself*: the connector line of the icon label of a knife occluded the upper part of the knife. (H) *Icon label occluding objects behind*: the icon label of the spoon occluded part of the scissors behind it.

R10). For example, when perceiving the occluded scissors with the static outline, R8 misidentified the handle as “a heart” (Figure 7A).

The icon label helped participants identify partially occluded objects by visualizing the complete shape and indicating its category (5/12; e.g., R3, R8). As R8 said: “[The icon] really makes the object clear as to what it is. It really defines that that is scissors, that is a knife. [Without augmentation] I would have had to move myself closer to the object to figure it out, versus the icon was like popped up so you knew [an object] was behind or around something.”

Obscuring differences between visually similar objects.

Participants also struggled to distinguish similar objects with subtle differences (e.g., spoons and forks, or jars and glasses; 7/12; e.g., R2, R12). As with partial occlusion, icon labels helped because they conveyed object type through clearly distinguishable icons (R3, R4, R5, R11). However, the AR outline did not help as it did not preserve fine shape details (e.g., fork tines), providing similar outlines for different objects (R5, R9, R11, R12). For example, the spoon and fork in Figure 7B looked more similar due to similar AR outlines. The solid overlay further exacerbated this problem by obscuring object

details, removing the subtle visual differences (e.g., shape, color) that participants relied on to distinguish similar objects (9/12; e.g., R4, R8). For example, R8 reported that the solid overlay made it harder to distinguish jars from glasses as it obscured whether the object had a lid: “[with the overlay] I can’t tell if there’s a cover on [the glass]” (Figure 7C).

Adjacent augmentations blending in. We found that adjacent objects with similar importance levels posed unique challenges as the same augmentations clustered together and introduced visual confusion. The augmentations (e.g., outline, solid overlay) could overlap or merge, creating misleading shapes. For example, five participants (e.g., R1, R9) reported that multiple adjacent outlines would intersect to “look like one big highlight” (R9) and be mistaken for one object. R1 also misidentified the outlines of two adjacent glasses of different heights as a pitcher, with the shorter glass as the “handle” and taller one as the “body” (Figure 7D). Similarly, solid overlays on multiple adjacent objects would visually merge to create one shape that was hard to identify (6/12; e.g., R5, R12). For example, R12 misidentified a bowl with an extruding fork as “a

bowl with a handle” from the two combined overlays (Figure 7E). In contrast, icon labels did not suffer from this “merging” problem as they were spatially separated above the objects (R5, R10).

Spatial misalignment of icon labels. Despite the advantages of icon labels mentioned above, they do not directly augment the original objects, generating spatial misalignment. This offset resulted in multiple problems in a complex scene: First, as icons did not precisely indicate object boundaries, they failed to convey the exact relative position and relationships between overlapping objects (5/12; e.g., R2, R6). For example, when augmenting a spoon inside a bowl, R6 reported that both icons floated side-by-side at similar heights, giving no indication of whether the spoon was inside or beside the bowl (Figure 7F). Second, as icons floated above the augmented objects, they created new occlusion, blocking part of the augmented object itself (R2, Figure 7G), objects behind them (R5, R7, R9, Figure 7H), or the icons of other objects (R7).

Moreover, the spatial misalignment required participants to consistently switch attention between icons and their corresponding objects, causing increased mental load (5/12; e.g., R1, R6). As R6 elaborated: “I’d rather have the outline right on the bowl than a little [...] arrow pointing to it. It’s kind of like: okay, I see a scissors [icon]. Now, where’s the scissors? Oh, the scissors is down there.” Some participants (R1, R5) also raised that in tasks requiring object interaction (e.g., cooking), the icons would mislead them as they “tend to reach for the icon” (R1).

5.3.4 Preferences on AR Distinction Designs. Compared to augmenting all important objects equally, eleven participants (all except R6) preferred the AR distinction in SceneGlance because it helped guide their attention (see Section 5.3.2). Ten participants (e.g., R2, R7) appreciated that the AR distinction emphasized dangerous objects, allowing them to pinpoint these objects more quickly. Eight participants (e.g., R5, R10) also found that the AR distinction helped with object identification because each augmentation represented a smaller set of objects and narrowed down their identification scope. For example, R1 used yellow outline for primary-important objects and blue outline for secondary ones: “When I saw something in yellow, I knew that it would have to be one of those [four] items, like the glass or the scissors, and when I saw something in blue I knew it was a bowl, a fork, or a spoon.” In contrast, R6 preferred augmenting all objects equally, as she found that the AR distinction added cognitive load: “I think I’m doing too much thinking [with distinction] like, oh, some [augmentations] are blue, why are they blue? So I’m kind of going down that way.”

When choosing distinction designs, all 12 participants chose distinguishing *By Color*, and five combined it with *By Form* (R6, R12) or *By Additional Visual Information* (R4, R5, R8). Table 3 lists participants’ distinction design choices. Participants chose distinction design based on two criteria: intuitive mapping between augmentation and importance, and avoiding visual clutter.

Intuitive mapping between augmentation and importance. Participants evaluated the AR distinction designs on how intuitively the visual difference mapped to importance levels. Six participants (e.g., R5, R9) considered *By Color* intuitive and easy to interpret. By assigning a brighter color to primary-important objects (e.g., yellow or green) and a less bright color to secondary ones (e.g., blue), nine participants (e.g., R4, R8) reported that the brighter augmentations

stood out and naturally drew their attention at first glance without additional thinking. As R9 said: “When it was different colors, that was very clear that this is a [primary-important] object, whereas [with the same color] I have to kind of stop and read the icon a little bit, so it’s not as immediate.”

When evaluating *By Form* distinction, three participants (R6, R7, R12) found the pairing of solid overlay for primary-important objects and outline for secondary ones intuitive, because both augmentations were rendered on the object and differed only in visual coverage. As R12 explained, the solid overlay “adds a very clear distinction” without “the complexity of adding the vertical lines and the icons.” In contrast, no participant chose to use outlines for one importance level and icon labels for the other. While these two designs were visually different, they did not intuitively convey importance levels (R5, R6, R9). R5 and R9 also found *By Additional Visual Information* unintuitive to indicate importance.

Avoiding visual clutter. Three participants (R4, R5, R8) combined *By Color* with *By Additional Visual Information* to highlight primary-important objects. They felt that outline alone did not help with object identification, but adding icon labels to objects would create high visual clutter. Selectively adding icons to only primary-important objects balanced visual clutter and clarity (R2, R4, R5, R8). As R4 described, this method “is a happy medium of helping me identify, but also not like an insane level of distraction.” R3 and R9 cautioned that combining multiple augmentations on the same object could create visual clutter, as the augmentations intertwined and became difficult to parse. For example, R3 found that the outlines “get messed up with the icons,” making it messy and “hard to concentrate.”

6 Study II: Using SceneGlance in Busy Street Navigation

While Study I revealed the impact of SceneGlance as well as the challenges of augmenting multiple objects in a complex kitchen scene, it contained only static objects, whereas complex scenes can also involve dynamic objects, a key dimension of complexity absent in Study I. Therefore, we supplemented it with an outdoor scenario that involves moving objects. Specifically, we conducted a think-aloud study with 13 PLV who navigated a preplanned outdoor route using SceneGlance to explore the unique opportunities and challenges in complex, dynamic outdoor scenes. Although prior work proposed head-mounted AR systems for the outdoor mobility of visually impaired people [66, 75], to our knowledge, none were evaluated in real-world outdoor scenarios.

6.1 Methods

6.1.1 Participants. We recruited 13 participants (T1–T13) from the same source as Study I (Section 5.1.1). Table 9 in Appendix E details participants’ demographic information and visual conditions. All except T1, T8, and T11 were legally blind. Five participants (T3, T5, T8, T10, T12) had low visual acuity (i.e., no better than 20/100), eight (T2–T4, T6, T7, T9, T12, T13) had peripheral vision loss, and four (T5, T8, T11, T13) had central vision loss. Three participants (T3, T6, T9) had prior experience with AR, but none used it regularly. Three participants attended both studies (R1/T3;

AR Distinction Design	R1	R2	R3	R4	R5	R6	R7	R8	R9	R10	R11	R12
By Color	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
By Form						✓						✓
By Additional Visual Information				✓	✓			✓				

Table 3: Participants’ selection of AR distinction designs in Study I. Participants could combine multiple distinction methods to distinguish primary- and secondary-important objects. All 12 participants used color distinction. Five participants additionally combined it with distinguishing by form (R6, R12) or by additional visual information (R4, R5, R8).

R6/T5; R8/T10). Participants were compensated \$25 per hour. The study was approved by our university IRB.

6.1.2 Apparatus. We pre-planned an outdoor route of approximately 1,100 feet (335 meters) long with two intersections. The first intersection was labeled by a stop sign and the second by traffic lights. A railway crossed the route near the second intersection. Participants walked on the sidewalk except when crossing the two intersections. Overall, the route represented a complex walking path for PLV. See Figure 8 for an overview of the route.

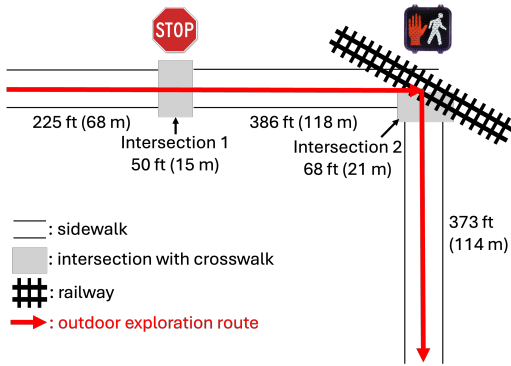


Figure 8: The route for the outdoor navigation study, which crossed two intersections and passed a railway crossing.

6.1.3 Procedure. Due to the difficulty and potential safety risks of conducting a quantitative outdoor navigation experiment, we conducted a free-form think-aloud study to gain a preliminary understanding of PLV’s experiences and challenges with SceneGlace. The study lasted approximately one hour. Participants navigated along the preplanned route using SceneGlace (Figure 8). At the beginning of the route, participants tried out all base augmentations and AR distinction designs, customized them, and selected their preferred designs. They were allowed to not use AR distinction but only one base augmentation if they preferred due to safety consideration. Then, participants walked along the route with SceneGlace using their preferred augmentations. During the walk, participants “thought aloud,” discussing their experience with SceneGlace, including benefits, challenges, and suggestions for improvement. One researcher on the team walked with them to ensure safety and a second researcher video-recorded the session for later analysis.

We followed the same procedure as in Study I to analyze the qualitative data (Section 5.2.2): two researchers open-coded three participants’ transcripts (23% of the data) to develop an initial codebook upon agreement. One researcher then coded the remaining transcripts and derived themes on participants’ experiences with

SceneGlace, including its benefits in outdoor navigation, the challenges of perceiving AR-enhanced dynamic scenes, and participants’ suggestions and design preferences.

6.2 Findings

We identified two unique scenarios in complex outdoor environments, where SceneGlace was helpful but also encountered augmentation challenges: perceiving ground surfaces and dynamic objects. We also found that different recognition errors (false negatives vs. false discoveries) affected participants’ safety asymmetrically. Moreover, participants’ preferences on AR distinction designs changed due to outdoor lighting conditions. We elaborate on these findings below.

6.2.1 Perceiving ground surfaces. Unlike the kitchen environment where augmented objects were small, discrete, and bounded (e.g., bowls, knives, glasses), the outdoor scene contained large, continuous ground surfaces (e.g., sidewalks, crosswalks, railway tracks). Participants reported that augmenting these surfaces supported safer navigation. Ten participants (e.g., T1, T8) appreciated that the AR outline clarified sidewalk and crosswalk boundaries and helped them stay within walkable areas. Augmenting walkable surfaces also helped participants better notice and avoid tripping hazards, such as curbs (11/13; e.g., T2, T13) and railway crossings (7/13; e.g., T3, T7). As T4 said: “[The outline] helps you with taking your path [...] [knowing] the sidewalk and where you can’t go. That makes you aware that: okay, that’s the [sidewalk] edge there, or there’s an up or down [curb] that could trip you.”

Since tripping hazards were categorized as primary-important objects, the AR distinction further helped participants attend to them first among all augmented objects (6/13; e.g., T8, T11). As T12 explained: “I think having the differentiation kind of outlined what was more important [...] you can only take in so much information and you want to take in the most relevant stuff first. Like the fact that there’s a curb coming, that’s important so you don’t fall.”

Despite its promise, participants reported several augmentation challenges associated with ground surfaces:

Augmentations interfered with objects on the surface. Because ground surfaces stretched continuously across the scene, the AR outlines frequently intersected with other objects on the surface and their augmentations (e.g., people, poles), creating confusing shapes (5/13; e.g., T3, T13). For example, T13 felt confused when the sidewalk outline crossed the outline of a pedestrian (Figure 9A). The solid overlay exacerbated this problem, as overlays of surfaces and overlays of objects on them merged into one colored shape (5/13; e.g., T8, T9). While the AR distinction helped reduce this visual interference (T3), it did not work when the surface and objects on it had the same importance level. T11 thus suggested using a designated color for walkable surfaces to visually separate them from other augmented objects.

Icon labels did not convey surface boundaries. While icon labels avoided the intersection problem, they did not convey surface boundaries, which were critical information for navigation (T7–T9, T11). Moreover, without explicitly highlighting surface boundaries, three participants (T5, T8, T11) found it hard to understand what the icons were referring to (Figure 9B). As T8 noted: “I think [the icon] is cool, but I think I like the outlines better. Just being outdoors in the wide open space, it’s nicer to have the outlines, whereas [with the icon] it’s one object kind of defining a lot more of the area.” T7 suggested combining icon labels with outlines to both identify surfaces and convey their boundaries.

Amplified impact of object segmentation errors on large surfaces. Since ground surfaces were large and extended across the scene, the inaccuracies in the mask from the object segmentation model were amplified and appeared obvious. As a result, six participants (e.g., T6, T9) reported that the AR outlines of sidewalks and crosswalks were confusing because they squiggled rather than precisely tracing the surface edge (Figure 9C). As T9 said: “[SceneGlance] doesn’t tell you exactly what the sidewalk is. It just gives me some lines, some squiggly lines, some straight lines [...] so it’s not really accurate.” T12 also said that she “wouldn’t rely on [surface augmentations] for safety” due to the inaccurate outlines. Since the icon labels were placed at the center of the segmentation masks, T5 also found the icons for the sidewalk constantly “running around” as she walked, bringing more confusion.

6.2.2 Perceiving and distinguishing dynamic objects. Unlike the static kitchen environment, the outdoor scene contained dynamic objects (e.g., pedestrians and cyclists). Participants reported that augmentation helped them perceive these moving objects, and suggested that SceneGlance should distinguish dynamic objects based on the likelihood of collision.

Augmentations revealed object motion. Ten participants (e.g., T1, T9) reported that augmentation helped them perceive dynamic objects such as pedestrians and cyclists. Because the augmentations moved together with the dynamic objects, participants could notice when an object was moving and judge its direction. As T9 described: “[From the augmentation] you can tell that [pedestrians] are walking. They’re walking towards me. You can tell by their legs when they’re moving.” Seven participants (e.g., T4, T10) further noted that this awareness helped them anticipate potential collisions and plan ahead to avoid risks.

Importance depended on path-crossing likelihood. While participants appreciated augmenting moving objects, seven participants (e.g., T1, T3) also suggested that the importance level of dynamic objects should depend on their likelihood of crossing the user’s path. Objects more likely to cross their paths, such as cars at crosswalks (T3, T8, T10) or incoming pedestrians on the same sidewalk (T3, T7, T8), should have higher importance so that participants could notice them quickly. Objects less likely to do so, such as moving cars on the road while walking on the sidewalk (T1, T4, T7), should be of lower importance or not augmented. As T4 described: “with walking on the sidewalk [...] I don’t think I need to see the cars as much [...] But getting to the crosswalk, augmenting the cars might be helpful.”

6.2.3 Asymmetric Impact of Recognition Errors on Safety. Beyond the augmentation challenges, participants’ safety also depended on

the recognition accuracy. We found that recognition errors affected safety asymmetrically. False negatives were safety-critical, since missing an important object would leave participants unaware of a hazard (T7, T9). For example, when T9 approached the other side of a crossing, neither the curb nor the curb cut was augmented, so she could not tell where to step onto the sidewalk without tripping over the curb. In contrast, false discoveries were less safety-critical and mainly caused confusion (T9, T11), such as seeing an augmentation for a person who was not there (T9). T2 even found false discoveries helpful: when a bike rack (not in the original important object list) was augmented as a fence, the augmentation made her aware of the bike rack so she could avoid it.

6.2.4 Preferences on AR distinction designs. Unlike the indoor scenario in Study I, where all participants chose the *By Color* distinction, participants’ preferences outdoors appeared more diverse: only eight of the 13 participants chose *By Color*, with T1 proposing a new design (*By Outline Thickness*) and T10 preferring no distinction at all. Table 4 lists participants’ selection of distinction designs in the outdoor environment.

While applying similar augmentation selection criteria to Study I (Section 5.3.4)—the intuitiveness of mapping between augmentation and importance (5/13; e.g., T1, T7) and the severity of visual clutter (6/13; e.g., T4, T8)—participants further reported that their preferences were affected by the outdoor lighting conditions. Unlike the uniform lighting of the indoor kitchen, which made color differences easy to perceive, the variable outdoor lighting distorted the augmentation colors. Eight participants (e.g., T4, T11) found that color differences were more subtle and sometimes hard to perceive. For example, T11 used red icon labels for primary-important objects and white icon labels for secondary ones but found that the white icons appeared “pinkish purple” and looked similar to the red icons when viewed under the sun. As a result, four participants (T7, T8, T12, T13) preferred the *By Form* or *By Additional Visual Information* AR distinction as these designs distinguished objects by shape and remained visible with diminished color contrast. As T8 described, when augmenting primary-important objects with solid overlay and secondary ones with outline, the solid overlay appeared “very distinguished compared to just the outline of the sidewalk or the outline of the person,” whereas distinction by color “kind of blended in together.” Similarly, T1’s *By Outline Thickness* (a thicker outline for primary-important objects and a thinner one for secondary ones) remained visually distinct in the outdoor environment.

7 Discussion

In this paper, we explored the opportunities and challenges of AR augmentation and distinction in supporting PLV’s perception of complex scenes. We designed SceneGlance, a wearable AR system that recognizes, locates, and augments multiple important objects in 3D space with AR distinction. Through a well-controlled study at a mock-up kitchen counter with 12 PLV, we found that SceneGlance shifted participants’ attention toward primary-important objects. However, we also identified perceptual challenges from multi-object augmentations in complex scenes, including perceiving partially occluded and visually similar objects, adjacent augmentations creating visual confusion, and spatial misalignment of icon labels. A second free-form think-aloud study with 13 PLV in the outdoor



Figure 9: Examples of perception challenges identified in Study II. (A) Outlines of the sidewalk and a railway track crossed the outline of a pedestrian, creating visually confusing shapes. (B) An icon label on a walkable surface did not indicate which area it was augmenting. (C) Outlines of the sidewalk squiggled rather than precisely tracing the sidewalk edge.

AR Distinction Design	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11	T12	T13
By Color		✓	✓	✓		✓	✓	✓	✓		✓		
By Form					✓			✓				✓	
By Additional Visual Information		✓				✓	✓						✓
By Outline Thickness	✓												
No Distinction										✓			

Table 4: Participants’ selection of AR distinction designs in the outdoor study. Unlike the kitchen study where all 12 participants chose distinction by color, outdoor participants’ choices were more varied. Eight of 13 used color distinction, three used form, and four used additional visual information. T10 preferred no distinction and augmented all important objects using icon labels. T1 proposed distinguishing by outline thickness during the study (i.e., a thicker outline for primary-important objects and a thinner outline for secondary ones).

street environment further revealed design challenges and needs specific to dynamic, complex scenes, including augmenting continuous ground surfaces and distinguishing dynamic objects by collision likelihood. In this section, we discuss the design implications for multiple object augmentations in complex scenes, as well as limitations and future directions.

7.1 Design Implications for Multiple Object Augmentations in Complex Scenes

Summarizing participants’ experiences using SceneGlance across indoor and outdoor scenes, we discuss design implications for future AR systems that augment multiple objects in complex scenes.

Distinguish objects by spatial relations. Scene complexity is influenced not only by the number of objects but also by their spatial arrangement [54, 82]. During the kitchen perception task, augmentations on adjacent objects of the same importance level intersected and merged into confusing shapes (Section 5.3.3). This pattern is consistent with the Gestalt principle of uniform connectedness [84, 89] that adjacent regions with the same visual properties would form a connected uniform region and tend to be perceived as a single shape rather than as distinct objects. SceneGlance’s distinction did not prevent this problem because it evaluated object importance based only on per-object properties (e.g., risk severity, visual difficulty), not on inter-object relations. Future research should explore modeling spatial relations between objects (e.g., adjacency, overlap, and containment) when distinguishing objects. For example, systems could integrate scene graph generation [12, 30] to model objects’ spatial relations, or CV techniques such as amodal instance segmentation [61] to predict the occluded parts of objects.

Given such spatial relations, systems could break the uniformity of adjacent same-importance augmentations by varying a visual dimension, such as color, brightness, texture, or motion [84]. Beyond distinguishing overlapping objects, future systems could register static objects and continue augmenting them when partially or fully occluded (e.g., using dotted outlines) to avoid misleading blended shapes (Section 5.3.3) and keep hidden objects visible [23, 69].

Support anchor-based scene perception. While the formative study characterized factors of object importance (risk severity, visual difficulty), these factors did not fully capture the perceptual roles some objects played during PLV’s perception of complex scenes. During the kitchen perception task, some participants actively looked for specific objects as anchors and organized other objects relative to them when building mental maps (Section 5.3.1). They selected anchors based on visual salience, uniqueness in the scene, or safety concern, which only partially aligned with the importance factors used by SceneGlance. Some anchor objects, such as bowls and cutting boards, were categorized as secondary- or non-important because they were visually obvious, and therefore received less prominent augmentation or were not augmented at all. However, these objects still played a central role in PLV’s perception and may warrant visual support. Prior work has investigated perceptual landmarks at the scale of buildings, such as when understanding building layouts [45] and conducting indoor navigation [13], but it remains underexplored how PLV would select and use anchor objects in dense, complex scenes. Future research could investigate what anchors PLV select in different complex scenarios, and visually distinguish them to support anchor-based scene perception.

Predict trajectory of dynamic objects. Our outdoor findings suggest that augmenting dynamic objects not only made their motion visible but also gave PLV advance notice of potential conflict. Participants appreciated augmentation on dynamic objects (e.g., pedestrians, cyclists) because it helped them anticipate conflicts and react earlier (Section 6.2.2). Seven participants further suggested that the importance of dynamic objects should depend on their likelihood of crossing the user’s path: a car was important at a crosswalk but not while walking on the sidewalk, and an incoming pedestrian on the same sidewalk was more important than one across the street. Future research should predict which objects will soon be relevant from their trajectories and the user’s path, such as by incorporating pedestrian trajectory prediction techniques [2, 32]. Given such predictions, future systems could adapt augmentation to give advance notice, such as augmenting with increasing salience (e.g., brighter color, flashing outline) as an object approaches, or showing the object direction with an arrow [47]. In safety-critical cases, such as fast-approaching or suddenly appearing objects, future systems could add audio or tactile alerts alongside visual cues to convey urgency [28, 38].

Treat importance as context-dependent rather than object-intrinsic. Our findings suggest that importance is context-dependent rather than a fixed, per-object property. While the formative study derived importance from object-level features (risk severity, visual difficulty), Study II participants evaluated importance by the likelihood of collision (Section 6.2.2). Importance also depended on scene familiarity: in the formative study, a participant wanted only out-of-place objects augmented in familiar settings since he knew where most objects were. These findings suggest that importance is shaped not only by an object’s properties but also by the user’s situation and prior experience with a scene, echoing prior work showing that attention in real-world scenes is guided by context and prior knowledge [111]. Future systems should model importance not only from object properties but also from the user’s task, path, and familiarity with the scene [65], such as treating a car as primary-important at a crosswalk but non-important while the user walks on a sidewalk, or tracking a user’s visit history to prioritize unfamiliar and out-of-place objects.

Adapt augmentation granularity. Participants relied on grouping strategies to perceive complex scenes. During the kitchen perception task without augmentation, participants grouped nearby objects of the same category and memorized these groups rather than individual objects (Section 5.3.1). With SceneGlance, participants also used the spatial distribution of augmentations to build a rough impression of where important objects were concentrated before perceiving individual objects (Section 5.3.2). Consistent with prior work showing that grouping by proximity and similarity can support visual working memory [84, 89], this pattern suggests that PLV may group similar and nearby visual elements into larger units to manage the visual complexity of dense scenes. Adaptive information density techniques in AR have long been studied [65, 85, 101], and LLM-driven approaches have recently enabled contextually adaptive mixed-reality layouts [62], but none have extended them to PLV. Future systems for PLV could adapt these techniques to adjust augmentation granularity based on the user’s current focus. For example, for scene overview, the system could merge nearby objects of the same category into a single aggregate augmentation,

such as a group outline with a count label (e.g., “four glasses”). When users focus on a region, the aggregate could expand into per-object augmentations to augment individual objects. Users’ focused region could be inferred from head orientation or gaze direction [55, 113], or from explicit input such as voice or gesture [127].

7.2 Limitations and Future Directions

Our research has three main limitations. First, although we involved 22 participants in the two studies exploring design opportunities and challenges, we had only 12 participants for the well-controlled lab study (Study I), limiting the statistical power of our quantitative results. Future studies should involve more participants across diverse visual conditions to fully understand the effects of multi-object augmentation on scene perception for PLV. Second, three participants enrolled in both studies, potentially causing carryover effects, affecting their AR augmentation choices in the second study. The order of design presentation in the tutorial could also affect participant choices. While we encouraged all participants to freely explore and adjust every design until satisfied before the formal trials to minimize such effects, we could not fully rule them out. Future research could counterbalance the order of designs in the tutorial, and recruit independent samples across studies to further reduce these effects. Third, we conducted a free-form think-aloud study (Study II) instead of a controlled experiment for the outdoor environment due to safety risks and the difficulty of controlling outdoor environments. Future work should consider conducting a well-controlled study in complex outdoor scenes with quantitative measures to rigorously assess the benefits and challenges of multi-object augmentation in dynamic environments.

8 Conclusion

In this paper, we explored the opportunities and challenges of augmenting and distinguishing multiple objects in complex scenes for people with low vision. We designed SceneGlance, a wearable AR system that augments and visually distinguishes objects in 3D space based on their perceived importance, and conducted two studies to surface PLV’s perception strategies, preferences, and challenges. A well-controlled lab study with 12 PLV at a mock-up kitchen counter showed that SceneGlance effectively shifted participants’ attention toward more important objects, and also surfaced perception challenges specific to multi-object augmentation, including perceiving partially occluded and visually similar objects, merging of adjacent augmentations, and spatial misalignment of icon labels. A second free-form think-aloud study with 13 PLV in outdoor scenes further revealed challenges unique to complex outdoor scenes, such as augmenting continuous ground surfaces, and highlighted the need to distinguish dynamic objects by their likelihood of collision with the user. Based on these findings, we derive design implications for future AR vision enhancement systems in the complex real world.

Acknowledgments

This work was supported in part by the National Eye Institute of the National Institutes of Health under Grant No. R01EY037100, an Apple Seed Grant, and the McPherson Eye Research Institute Grant Accelerator Program at the University of Wisconsin-Madison. We thank all participants for their time and insights.

References

- [1] Walid I Al-Atabany, Muhammad A Memon, Susan M Downes, and Patrick A Degenaar. 2010. Designing and testing scene enhancement algorithms for patients with retina degenerative disorders. *Biomedical engineering online* 9, 1 (2010), 27.
- [2] Alexandre Alahi, Kratharth Goel, Vignesh Ramanathan, Alexandre Robicquet, Li Fei-Fei, and Silvio Savarese. 2016. Social lstm: Human trajectory prediction in crowded spaces. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 961–971.
- [3] George A Alvarez and Patrick Cavanagh. 2004. The capacity of visual short-term memory is set both by visual information load and by number of objects. *Psychological science* 15, 2 (2004), 106–111.
- [4] American Optometric Association. 2023. Legal blindness in America — aoa.org. <https://www.aoa.org/news/clinical-eye-care/diseases-and-conditions/legal-blindness-in-america?sso=y>. [Accessed 12-03-2025].
- [5] Anastasios Nikolas Angelopoulos, Hossein Ameri, Debbie Mitra, and Mark Humayun. 2019. Enhanced depth navigation through augmented reality depth mapping in patients with low vision. *Scientific reports* 9, 1 (2019), 11230.
- [6] Richard A Armstrong. 2014. When to use the Bonferroni correction. *Ophthalmic and physiological optics* 34, 5 (2014), 502–508.
- [7] Marie Claire Bilyk, Jessica M Sontrop, Gwen E Chapman, Susan I Barr, and Linda Mamer. 2009. Food experiences and eating patterns of visually impaired and blind people. *Canadian Journal of Dietetic practice and research* 70, 1 (2009), 13–18.
- [8] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative research in psychology* 3, 2 (2006), 77–101.
- [9] Virginia Braun and Victoria Clarke. 2024. Thematic analysis. In *Encyclopedia of quality of life and well-being research*. Springer, 7187–7193.
- [10] Xiaoqian J Chai, Noa Ofen, Lucia F Jacobs, and John DE Gabrieli. 2010. Scene complexity: influence on perception, memory, and development in the medial temporal lobe. *Frontiers in human neuroscience* 4 (2010), 1021.
- [11] Ruei-Che Chang, Yuxuan Liu, and Anhong Guo. 2024. Worldscribe: Towards context-aware live visual descriptions. In *Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology*. 1–18.
- [12] Xiaojun Chang, Pengzhen Ren, Pengfei Xu, Zhihui Li, Xiaojiang Chen, and Alex Hauptmann. 2021. A comprehensive survey of scene graphs: Generation and application. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 45, 1 (2021), 1–26.
- [13] Ruijia Chen, Junru Jiang, Pragati Maheshwary, Brianna R Cochran, and Yuhang Zhao. 2025. Visimark: Characterizing and augmenting landmarks for people with low vision in augmented reality to support indoor navigation. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. 1–20.
- [14] Ruijia Chen, Yuheng Wu, Charlie Houseago, Filipe Gaspar, Filippo Aleotti, Dorian Gálvez-López, Oliver Johnston, Diego Mazala, Guillermo Garcia-Hernando, Maryam Bandukda, Gabriel Brostow, and Jessica Van Brummelen. 2026. NaviNote: Enabling In-situ Spatial Annotation Authoring to Support Exploration and Navigation for Blind and Low Vision People. In *Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems (CHI '26)*. Association for Computing Machinery, New York, NY, USA, Article 948, 22 pages. doi:10.1145/3772318.3790589
- [15] Jacob Cohen. 2013. *Statistical power analysis for the behavioral sciences*. Routledge.
- [16] Dima Damen, Hazel Doughty, Giovanni Maria Farinella, Sanja Fidler, Antonino Furnari, Evangelos Kazakos, Davide Moltisanti, Jonathan Munro, Toby Perrett, Will Price, and Michael Wray. 2018. Scaling Egocentric Vision: The EPIC-KITCHENS Dataset. In *European Conference on Computer Vision (ECCV)*.
- [17] Raghavendra Singh Dasila, Meet Trivedi, Shubham Soni, M Senthil, and M Narendran. 2017. Real time environment perception for visually impaired. In *2017 IEEE Technological Innovations in ICT for Agriculture and Rural Development (TIAR)*. IEEE, 168–172.
- [18] Johanne Desrosiers, Marie-Chantal Wanet-Defalque, Khatoune Témisjjan, Jacques Gresset, Marie-France Dubois, Judith Renaud, Claude Vincent, Jacqueline Rousseau, Mathieu Carignan, and Olga Overbury. 2009. Participation in daily activities and social roles of older adults with visual impairment. *Disability and rehabilitation* 31, 15 (2009), 1227–1234.
- [19] Juan C Dibene and Enrique Dunn. 2022. HoloLens 2 Sensor Streaming. *arXiv preprint arXiv:2211.02648* (2022).
- [20] Olive Jean Dunn. 1961. Multiple comparisons among means. *Journal of the American statistical association* 56, 293 (1961), 52–64.
- [21] Martin Eckert, Matthias Blex, Christoph M Friedrich, et al. 2018. Object detection featuring 3D audio localization for Microsoft HoloLens. In *Proc. 11th Int. Joint Conf. on Biomedical Engineering Systems and Technologies*, Vol. 5. 555–561.
- [22] Lisa A. Elkin, Matthew Kay, James J. Higgins, and Jacob O. Wobbrock. 2021. An Aligned Rank Transform Procedure for Multifactor Contrast Tests. In *The 34th Annual ACM Symposium on User Interface Software and Technology (Virtual Event, USA) (UIST '21)*. Association for Computing Machinery, New York, NY, USA, 754–768. doi:10.1145/3472749.3474784
- [23] Niklas Elmquist and Philippas Tsigas. 2008. A taxonomy of 3d occlusion management for visualization. *IEEE transactions on visualization and computer graphics* 14, 5 (2008), 1095–1109.
- [24] Mark Everingham, Luc Van Gool, Christopher KI Williams, John Winn, and Andrew Zisserman. 2010. The pascal visual object classes (voc) challenge. *International journal of computer vision* 88 (2010), 303–338.
- [25] M. R. Everingham, B. T. Thomas, and T Troscianko. 1998. Head-Mounted Mobility Aid for Low Vision Using Scene Classification Techniques. *International Journal of Virtual Reality* 3, 4 (Jan. 1998), 1–10. doi:10.20870/IJVR.1998.3.4.2629
- [26] Dylan R. Fox, Ahmad Ahmadzadeh, Clara T. Friedman, Shiri Azenkott, Marlena A. Chu, Roberto Manduchi, and Emily A. Cooper. 2023. Using augmented reality to cue obstacles for people with low vision. *Opt. Express* 31, 4 (Feb 2023), 6827–6848. doi:10.1364/OE.479258
- [27] Bhanuka Gamage, Nicola McDowell, Dijana Kovacic, Leona Holloway, Thanh-Toan Do, Arthur James Lowery, Nicholas Price, and Kim Marriot. 2025. Smart Glasses for CVI: Co-Designing Extended Reality Solutions to Support Environmental Perception by People with Cerebral Visual Impairment. In *Proceedings of the 27th International ACM SIGACCESS Conference on Computers and Accessibility*. 1–24.
- [28] Yun Gao, Dan Wu, Jie Song, Xueyi Zhang, Bangbang Hou, Hengfa Liu, Junqi Liao, and Liang Zhou. 2025. A wearable obstacle avoidance device for visually impaired individuals with cross-modal learning. *Nature Communications* 16, 1 (2025), 2857.
- [29] Google. 2025. Protocol Buffers. <https://developers.google.com/protocol-buffers>. Accessed: 2025-03-30.
- [30] Qiao Gu, Alihusein Kuwajerwala, Sacha Morin, Krishna Murthy Jatavallabhula, Bipasha Sen, Aditya Agarwal, Corban Rivera, William Paul, Kirsty Ellis, Rama Chellappa, Chuang Gan, Celso Miguel de Melo, Joshua B. Tenenbaum, Antonio Torralba, Florian Shkurti, and Liam Paull. 2023. Concept-Graphs: Open-Vocabulary 3D Scene Graphs for Perception and Planning. arXiv:2309.16650 [cs.RO] <https://arxiv.org/abs/2309.16650>
- [31] Fangli Guan, Zhixiang Fang, Lubin Wang, Xucai Zhang, Haoyu Zhong, and Haosheng Huang. 2022. Modelling people's perceived scene complexity of real-world environments using street-view panoramas and open geodata. *ISPRS Journal of Photogrammetry and Remote Sensing* 186 (2022), 315–331.
- [32] Agrim Gupta, Justin Johnson, Li Fei-Fei, Silvio Savarese, and Alexandre Alahi. 2018. Social gan: Socially acceptable trajectories with generative adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2255–2264.
- [33] Yu Hao, Alexey Magay, Hao Huang, Shuaihang Yuan, Congcong Wen, and Yi Fang. 2024. ChatMap: A Wearable Platform Based on the Multi-modal Foundation Model to Augment Spatial Cognition for People with Blindness and Low Vision. In *2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 129–134.
- [34] Yu Hao, Fan Yang, Hao Huang, Shuaihang Yuan, Sundeep Rangan, John-Ross Rizzo, Yao Wang, and Yi Fang. 2024. A multi-modal foundation model to assist people with blindness and low vision in environmental interaction. *Journal of Imaging* 10, 5 (2024), 103.
- [35] Sharon A Haymes, Alan W Johnston, and Anthony D Heyes. 2002. Relationship between vision impairment and ability to perform activities of daily living. *Ophthalmic and Physiological Optics* 22, 2 (2002), 79–91.
- [36] John M Henderson, Myriam Chaneaux, and Tim J Smith. 2009. The influence of clutter on real-world scene search: Evidence from search efficiency and eye movements. *Journal of vision* 9, 1 (2009), 32–32.
- [37] John M Henderson, Taylor R Hayes, Candace E Peacock, and Gwendolyn Rehrig. 2019. Meaning and attentional guidance in scenes: A review of the meaning map approach. *Vision* 3, 2 (2019), 19.
- [38] Marion Hersh. 2022. Wearable travel aids for blind and partially sighted people: A review with a focus on design issues. *Sensors* 22, 14 (2022), 5454.
- [39] Stephen L Hicks, Iain Wilson, Louwai Muhammed, John Worsfold, Susan M Downes, and Christopher Kennard. 2013. A depth-based head-mounted visual display to aid navigation in partially sighted individuals. *PLOS ONE* 8, 7 (2013), e67695.
- [40] Jonathan Huang, Max Kinader, Matt J Dunn, Wojciech Jarosz, Xing-Dong Yang, and Emily A Cooper. 2019. An augmented reality sign-reading assistant for users with reduced vision. *PLoS one* 14, 1 (2019), e0210630.
- [41] Alex D Hwang and Eli Peli. 2014. An augmented-reality edge enhancement application for Google Glass. *Optometry and vision science* 91, 8 (2014), 1021–1030.
- [42] Md Touhidul Islam, Imran Kabir, Elena Ariel Pearce, Md Alimoor Reza, and Syed Masum Billah. 2024. Identifying Crucial Objects in Blind and Low-Vision Individuals' Navigation. In *Proceedings of the 26th International ACM SIGACCESS Conference on Computers and Accessibility*. 1–8.
- [43] Wathanasak Jeamwathanachai, Mike Wald, and Gary Wills. 2019. Indoor navigation by blind people: Behaviors and challenges in unfamiliar spaces and buildings. *British Journal of Visual Impairment* 37, 2 (2019), 140–153. doi:10.1177/0264619619833723

- [44] Nabila Jones, Hannah Elizabeth Bartlett, and Richard Cooke. 2019. An analysis of the impact of visual impairment on activities of daily living and vision-related quality of life in a visually impaired adult population. *British Journal of Visual Impairment* 37, 1 (2019), 50–63.
- [45] Amy A Kalia, Gordon E Legge, and Nicholas A Giudice. 2008. Learning building layouts with non-geometric visual information: The effects of visual impairment and age. *Perception* 37, 11 (2008), 1677–1699.
- [46] Avyay Ravi Kashyap. 2020. Behaviors, Problems and Strategies of Visually Impaired Persons During Meal Preparation in the Indian Context: Challenges and Opportunities for Design. In *Proceedings of the 22nd International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS '20)*. Association for Computing Machinery, New York, NY, USA, Article 101, 3 pages. doi:10.1145/3373625.3417083
- [47] Liam Kettle and Yi-Ching Lee. 2022. Augmented reality for vehicle-driver communication: a systematic review. *Safety* 8, 4 (2022), 84.
- [48] Doaa Khattab, Julie Buelow, and Donna Saccuteli. 2015. Understanding the barriers: Grocery stores and visually impaired shoppers. *Journal of accessibility and design for all: JACCESS* 5, 2 (2015), 157–173.
- [49] Daniel Killough, Justin Feng, Zheng Xue Ching, Daniel Wang, Rithvik Dyava, Yapeng Tian, and Yuhang Zhao. 2025. VRSight: An AI-Driven Scene Description System to Improve Virtual Reality Accessibility for Blind People. In *Proceedings of the 38th Annual ACM Symposium on User Interface Software and Technology (UIST '25)*. Association for Computing Machinery, New York, NY, USA, 1–17. doi:10.1145/3746059.3747641
- [50] Benjamin Kommey, Kumbong Herrman, and Ernest Ofori Addo. 2019. A smart vision based navigation aid for the visually impaired. *Asian Journal of Research in Computer Science* 4, 3 (2019), 1–8.
- [51] Masaki Kuribayashi, Kohei Uehara, Allan Wang, Shigeo Morishima, and Chieko Asakawa. 2025. Wanderguide: Indoor map-less robotic guide for exploration by blind people. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. 1–21.
- [52] Alexandra Kuznetsova, Per B Brockhoff, and Rune HB Christensen. 2017. lmerTest package: tests in linear mixed effects models. *Journal of statistical software* 82 (2017), 1–26.
- [53] MiYoung Kwon, Chaithanya Ramachandra, PremNandhini Satgunam, Bartlett W Mel, Eli Peli, and Bosco S Tjan. 2012. Contour enhancement benefits older adults with simulated central field loss. *Optometry and vision science* 89, 9 (2012), 1374–1384.
- [54] Cameron Kyle-Davidson, Elizabeth Yue Zhou, Dirk B Walther, Adrian G Bors, and Karla K Evans. 2023. Characterising and dissecting human perception of scene complexity. *Cognition* 231 (2023), 105319.
- [55] Mikko Kytö, Barrett Ens, Thammathip Piumsomboon, Gun A Lee, and Mark Billingham. 2018. Pinpointing: Precise head-and eye-based target selection for augmented reality. In *Proceedings of the 2018 CHI conference on human factors in computing systems*. 1–14.
- [56] Florian Lang and Tonja Machulla. 2021. Pressing a button you cannot see: evaluating visual designs to assist persons with low vision through augmented reality. In *Proceedings of the 27th ACM Symposium on Virtual Reality Software and Technology*. 1–10.
- [57] Jaewook Lee, Yang Li, Dylan Bunarto, Eujan Lee, Olivia H Wang, Adrian Rodriguez, Yuhang Zhao, Yapeng Tian, and Jon E Froehlich. 2024. Towards ai-powered ar for enhancing sports playability for people with low vision: An exploration of arsports. In *2024 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*. IEEE, 228–233.
- [58] Jaewook Lee, Andrew D Tjahjedi, Jiho Kim, Junpu Yu, Minji Park, Jiawen Zhang, Jon E Froehlich, Yapeng Tian, and Yuhang Zhao. 2024. CookAR: Affordance Augmentations in Wearable AR to Support Kitchen Tool Interactions for People with Low Vision. In *Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology*. 1–16.
- [59] Meesung Lee, Heerim Lee, Sungjoo Hwang, and Minji Choi. 2021. Understanding the impact of the walking environment on pedestrian perception and comprehension of the situation. *Journal of Transport & Health* 23 (2021), 101267.
- [60] Franklin Mingzhe Li, Jamie Dorst, Peter Cederberg, and Patrick Carrington. 2021. Non-Visual Cooking: Exploring Practices and Challenges of Meal Preparation by People with Visual Impairments. In *Proceedings of the 23rd International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS '21)*. Association for Computing Machinery, New York, NY, USA, Article 30, 11 pages. doi:10.1145/3441852.3471215
- [61] Ke Li and Jitendra Malik. 2016. Amodal instance segmentation. In *European Conference on Computer Vision*. Springer, 677–693.
- [62] Zhipeng Li, Christoph Gebhardt, Yves Inglin, Nicolas Steck, Paul Strelci, and Christian Holz. 2024. Situationadapt: Contextual ui optimization in mixed reality with situation awareness via llm reasoning. In *Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology*. 1–13.
- [63] Haozhe Lin, Jiangtao Gong, Yu Wang, Jinsong Zhang, Bing Bai, Yan Zhang, Luyao Wang, Chenyu Wei, Yancheng Cao, Kun Li, et al. 2025. AI system facilitates people with blindness and low vision in interpreting and experiencing unfamiliar environments. *npj Artificial Intelligence* 1, 1 (2025), 7.
- [64] Tsung-Yi Lin, Michael Maire, Serge Belongie, Lubomir Bourdev, Ross Girshick, James Hays, Pietro Perona, Deva Ramanan, C. Lawrence Zitnick, and Piotr Dollár. 2015. Microsoft COCO: Common Objects in Context. arXiv:1405.0312 [cs.CV] <https://arxiv.org/abs/1405.0312>
- [65] David Lindlbauer, Anna Maria Feit, and Otmar Hilliges. 2019. Context-aware online adaptation of mixed reality interfaces. In *Proceedings of the 32nd annual ACM symposium on user interface software and technology*. 147–160.
- [66] Alice Lo Valvo, Daniele Croce, Domenico Garlisi, Fabrizio Giuliano, Laura Giarré, and Ilenia Tinnirello. 2021. A navigation and augmented reality system for visually impaired people. *Sensors* 21, 9 (2021), 3061.
- [67] Gang Luo and Eli Peli. 2006. Use of an augmented-vision device for visual search by patients with tunnel vision. *Investigative ophthalmology & visual science* 47, 9 (2006), 4152–4159.
- [68] Chengqi Lyu, Wenwei Zhang, Haian Huang, Yue Zhou, Yudong Wang, Yanyi Liu, Shilong Zhang, and Kai Chen. 2022. Rtmtdet: An empirical study of designing real-time object detectors. *arXiv preprint arXiv:2212.07784* (2022).
- [69] Márcio CF Macedo and Antonio L Apolinario. 2021. Occlusion handling in augmented reality: Past, present and future. *IEEE Transactions on Visualization and Computer Graphics* 29, 2 (2021), 1590–1609.
- [70] Sean P MacEvoy and Russell A Epstein. 2011. Constructing scenes from objects in human occipitotemporal cortex. *Nature neuroscience* 14, 10 (2011), 1323–1329.
- [71] Alexey Magay, Dhurba Tripathi, Yu Hao, and Yi Fang. 2024. A Light and Smart Wearable Platform with Multimodal Foundation Model for Enhanced Spatial Reasoning in People with Blindness and Low Vision. In *European Conference on Computer Vision*. Springer, 323–339.
- [72] Florian Mathis and Johannes Schöning. 2025. LifeInsight: Design and Evaluation of an AI-Powered Assistive Wearable for Blind and Low Vision People Across Multiple Everyday Life Scenarios. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. 1–25.
- [73] Scott E Maxwell, Harold D Delaney, and Ken Kelley. 2017. *Designing experiments and analyzing data: A model comparison perspective*. Routledge.
- [74] Microsoft. 2017. Seeing AI - Talking Camera for the Blind — seeingai.com. <https://www.seeingai.com/>. [Accessed 09-09-2025].
- [75] Hein Min Htike, Tom H. Margrain, Yu-Kun Lai, and Parisa Eslambolchilar. 2021. Augmented reality glasses as an orientation and mobility aid for people with low vision: a feasibility study of experiences and requirements. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–15.
- [76] Aliaksei Miniukovich and Antonella De Angeli. 2014. Quantification of interface visual complexity. In *Proceedings of the 2014 international working conference on advanced visual interfaces*. 153–160.
- [77] Pilar Montero. 2005. Nutritional assessment and diet quality of visually impaired Spanish children. *Annals of Human Biology* 32, 4 (2005), 498–512.
- [78] Karin Müller, Christin Engel, Claudia Loitsch, Rainer Stiefelhagen, and Gerhard Weber. 2022. Traveling more independently: a study on the diverse needs and challenges of people with visual or mobility impairments in unfamiliar indoor environments. *ACM Transactions on Accessible Computing (TACCESS)* 15, 2 (2022), 1–44.
- [79] National Eye Institute. 2025. Low Vision | National Eye Institute — nei.nih.gov. <https://www.nei.nih.gov/learn-about-eye-health/eye-conditions-and-diseases/low-vision>. [Accessed 19-06-2024].
- [80] Mark B Neider and Gregory J Zelinsky. 2011. Cutting through the clutter: Searching for targets in evolving complex scenes. *Journal of Vision* 11, 14 (2011), 7–7.
- [81] Gerhard Neuhold, Tobias Ollmann, Samuel Rota Buló, and Peter Kontschieder. 2017. The mapillary vistas dataset for understanding of street scenes. In *Proceedings of the IEEE international conference on computer vision*. 4990–4999.
- [82] Aude Oliva, Michael L Mack, Mochan Shrestha, and Angela Peeper. 2004. Identifying the perceptual dimensions of visual complexity of scenes. In *Proceedings of the annual meeting of the cognitive science society*, Vol. 26.
- [83] Rafael Padilla, Wesley L Passos, Thadeu LB Dias, Sergio L Netto, and Eduardo AB Da Silva. 2021. A comparative analysis of object detection metrics with a companion open-source toolkit. *Electronics* 10, 3 (2021), 279.
- [84] Stephen Palmer and Irvin Rock. 1994. Rethinking perceptual organization: The role of uniform connectedness. *Psychonomic bulletin & review* 1, 1 (1994), 29–55.
- [85] Ji Hwan Park, Braden Roper, Amirhossein Arezoumand, and Tien Tran. 2025. Exploring AR Label Placements in Visually Cluttered Scenarios. In *2025 IEEE Visualization and Visual Analytics (VIS)*. IEEE, 336–340.
- [86] Karl Pearson. 1900. X. On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* 50, 302 (1900), 157–175.
- [87] Denis G Pelli and Katharine A Tillman. 2008. The uncrowded window of object recognition. *Nature neuroscience* 11, 10 (2008), 1129–1135.
- [88] Ricardo E. Gonzalez Penuela, Jazmin Collins, Cynthia L. Bennett, and Shiri Azenkot. 2024. Investigating Use Cases of AI-Powered Scene Description Applications for Blind and Low Vision People. *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (2024). doi:10.1145/3613904.3642211

- [89] Dwight J Peterson and Marian E Berryhill. 2013. The Gestalt principle of similarity benefits visual working memory. *Psychonomic bulletin & review* 20, 6 (2013), 1282–1289.
- [90] Sanjeev U Rao, Swaroop Ranganath, TS Ashwin, Guddeti Ram Mohana Reddy, et al. 2021. A Google glass based real-time scene analysis for the visually impaired. *IEEE Access* 9 (2021), 166351–166369.
- [91] Elena T Remillard, Lyndsie M Koon, Tracy L Mitzner, and Wendy A Rogers. 2024. Everyday Challenges for Individuals Aging with Vision Impairment: Technology Implications. *The Gerontologist* 64, 6 (2024), gnad169.
- [92] Ruth Rosenholtz, Yuanzhen Li, and Lisa Nakano. 2007. Measuring visual clutter. *Journal of vision* 7, 2 (2007), 17–17.
- [93] Arezoo Sadeghzadeh, Md Baharul Islam, Md Nur Uddin, and Tarkan Aydin. 2024. ARVA: An Augmented Reality-Based Visual Aid for Mobility Enhancement Through Real-Time Video Stream Transformation. *IEEE Access* 12 (2024), 137268–137283. doi:10.1109/ACCESS.2024.3462628
- [94] Ahmed M Sayed, Mohamed Abou Shousha, MD Baharul Islam, Taher K Eleiwa, Rashed Kashem, Mostafa Abdel-Mottaleb, Eyup Ozcan, Mohamed Tolba, Jane C Cook, and Richard K Parrish. 2020. Mobility improvement of patients with peripheral visual field losses using novel see-through digital spectacles. *PLOS ONE* 15, 10 (2020), e0240509.
- [95] S Shapiro, M.B. Wilk, et al. 1965. An analysis of variance test for normality. *Biometrika* 52, 3 (1965), 591–611.
- [96] Sandra D Starke, Eugenie Golubova, Michael D Crossland, and James S Wolffsohn. 2020. Everyday visual demands of people with low vision: A mixed methods real-life recording study. *Journal of Vision* 20, 9 (2020), 3–3.
- [97] Lee Stearns, Leah Findlater, and Jon E. Froehlich. 2018. Design of an Augmented Reality Magnification Aid for Low Vision Users. In *Proceedings of the 20th International ACM SIGACCESS Conference on Computers and Accessibility* (Galway, Ireland) (ASSETS '18). Association for Computing Machinery, New York, NY, USA, 28–39. doi:10.1145/3234695.3236361
- [98] Sarit Szpiro, Yuhang Zhao, and Shiri Azenkot. 2016. Finding a store, searching for a product: a study of daily challenges of low vision people. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (Heidelberg, Germany) (UbiComp '16). Association for Computing Machinery, New York, NY, USA, 61–72. doi:10.1145/2971648.2971723
- [99] Sarit Felicia Anais Szpiro, Shafeka Hashash, Yuhang Zhao, and Shiri Azenkot. 2016. How people with low vision access computing devices: Understanding challenges and opportunities. In *Proceedings of the 18th International ACM SIGACCESS Conference on Computers and Accessibility*. 171–180.
- [100] Dujie Tadin, Jeffrey B Nyquist, Kelly E Lusk, Anne L Corn, and Joseph S Lappin. 2012. Peripheral vision of youths with low vision: motion perception, crowding, and visual search. *Investigative ophthalmology & visual science* 53, 9 (2012), 5860–5868.
- [101] Markus Tatzgern, Valeria Orso, Denis Kalkofen, Giulio Jacucci, Luciano Gamberini, and Dieter Schmalstieg. 2016. Adaptive information density for augmented reality displays. In *2016 IEEE Virtual Reality (VR)*. IEEE, 83–92.
- [102] Jodi Teitelman and Al Copolillo. 2005. Psychosocial issues in older adults' adjustment to vision loss: findings from qualitative interviews and focus groups. *The American journal of occupational therapy* 59, 4 (2005), 409–417.
- [103] Miguel Thibaut, Muriel Boucart, and Thi Ha Chau Tran. 2020. Object search in neovascular age-related macular degeneration: the crowding effect. *Clinical and Experimental Optometry* 103, 5 (2020), 648–655.
- [104] John W Tukey. 1949. Comparing individual means in the analysis of variance. *Biometrics* (1949), 99–114.
- [105] Sandra Tullio-Pow, Hong Yu, and Megan Strickfaden. 2021. Do You See What I See? The shopping experiences of people with visual impairment. *Interdisciplinary Journal of Signage and Wayfinding* 5, 1 (2021), 42–61.
- [106] Kathleen A Turano, Duane R Geruschat, and Julie W Stahl. 1998. Mental effort required for walking: effects of retinitis pigmentosa. *Optometry and Vision Science* 75, 12 (1998), 879–886.
- [107] Dorin Ungureanu, Federica Bogo, Silvano Galliani, Pooja Sama, Xin Duan, Casey Meekhof, Jan Stühmer, Thomas J. Cashman, Bugra Tekin, Johannes L. Schönberger, Pawel Olszta, and Marc Pollefeys. 2020. HoloLens 2 Research Mode as a Tool for Computer Vision Research. arXiv:2008.11239 [cs.CV] <https://arxiv.org/abs/2008.11239>
- [108] Unity Technologies. 2022. Unity - Manual: Raw Image — docs.unity3d.com. <https://docs.unity3d.com/2022.3/Documentation/Manual/script-RawImage.html>. [Accessed 07-04-2025].
- [109] Joram J van Rheede, Iain R Wilson, Rose I Qian, Susan M Downes, Christopher Kennard, and Stephen L Hicks. 2015. Improving mobility performance in low vision with a distance-based representation of the visual scene. *Investigative ophthalmology & visual science* 56, 8 (2015), 4802–4809.
- [110] Melissa Le-Hoa Võ. 2021. The meaning and structure of scenes. *Vision Research* 181 (2021), 10–20.
- [111] Melissa Le-Hoa Võ, Sage EP Boettcher, and Dejan Draschkow. 2019. Reading scenes: How scene grammar guides attention and aids perception in real-world environments. *Current opinion in psychology* 29 (2019), 205–210.
- [112] Julian M Wallace, Susana TL Chung, and Bosco S Tjan. 2017. Object crowding in age-related macular degeneration. *Journal of Vision* 17, 1 (2017), 33–33.
- [113] Ru Wang, Ruijia Chen, Anqiao Erica Cai, Zhiyuan Li, Sanbrita Mondal, and Yuhang Zhao. 2025. Characterizing Visual Intents for People with Low Vision through Eye Tracking. In *Proceedings of the 27th International ACM SIGACCESS Conference on Computers and Accessibility*. 1–18.
- [114] Ru Wang, Zach Potter, Yun Ho, Daniel Killough, Linxiu Zeng, Sanbrita Mondal, and Yuhang Zhao. 2024. GazePrompt: Enhancing Low Vision People's Reading Experience with Gaze-Aware Augmentations. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*. 1–17.
- [115] Ru Wang, Nihan Zhou, Tam Nguyen, Sanbrita Mondal, Bilge Mutlu, and Yuhang Zhao. 2023. Characterizing barriers and technology needs in the kitchen for blind and low vision people. *arXiv preprint arXiv:2310.05396* (2023).
- [116] David Whitney and Dennis M Levi. 2011. Visual crowding: A fundamental limit on conscious perception and object recognition. *Trends in cognitive sciences* 15, 4 (2011), 160–168.
- [117] Sandro L Wiesmann and Melissa Le-Hoa Võ. 2023. Disentangling diagnostic object properties for human scene categorization. *Scientific reports* 13, 1 (2023), 5912.
- [118] Michele A Williams, Amy Hurst, and Shaun K Kane. 2013. "Pray before you step out": Describing Personal and Situational Blind Navigation Behaviors. In *Proceedings of the 15th international ACM SIGACCESS conference on computers and accessibility*. 1–8.
- [119] Jacob O. Wobbrock, Leah Findlater, Darren Gergle, and James J. Higgins. 2011. The aligned rank transform for nonparametric factorial analyses using only anova procedures. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Vancouver, BC, Canada) (CHI '11). Association for Computing Machinery, New York, NY, USA, 143–146. doi:10.1145/1978942.1978963
- [120] Jeremy M Wolfe. 2021. Guided Search 6.0: An updated model of visual search. *Psychonomic bulletin & review* 28, 4 (2021), 1060–1092.
- [121] Limin Zeng. 2015. A survey: outdoor mobility experiences by the visually impaired. In *Mensch und Computer 2015—Workshopband*. De Gruyter Oldenbourg, 391–397.
- [122] Xi Zhao, Kentaro Go, Kenji Kashiwagi, Masahiro Toyoura, Xiaoyang Mao, and Issei Fujishiro. 2019. Computational alleviation of homonymous visual field defect with ost-hmd: The effect of size and position of overlaid overview window. In *2019 International Conference on Cyberworlds (CW)*. IEEE, 175–182.
- [123] Yuhang Zhao, Michele Hu, Shafeka Hashash, and Shiri Azenkot. 2017. Understanding Low Vision People's Visual Perception on Commercial Augmented Reality Glasses. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (Denver, Colorado, USA) (CHI '17). Association for Computing Machinery, New York, NY, USA, 4170–4181. doi:10.1145/3025453.3025949
- [124] Yuhang Zhao, Elizabeth Kupferstein, Brenda Veronica Castro, Steven Feiner, and Shiri Azenkot. 2019. Designing AR Visualizations to Facilitate Stair Navigation for People with Low Vision. In *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology* (New Orleans, LA, USA) (UIST '19). Association for Computing Machinery, New York, NY, USA, 387–402. doi:10.1145/3332165.3347906
- [125] Yuhang Zhao, Sarit Szpiro, and Shiri Azenkot. 2015. Foresee: A customizable head-mounted vision enhancement system for people with low vision. In *Proceedings of the 17th international ACM SIGACCESS conference on computers & accessibility*. 239–249.
- [126] Yuhang Zhao, Sarit Szpiro, Jonathan Knights, and Shiri Azenkot. 2016. CueSee: exploring visual cues for people with low vision to facilitate a visual search task. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. 73–84.
- [127] Yuhang Zhao, Sarit Szpiro, Lei Shi, and Shiri Azenkot. 2019. Designing and evaluating a customizable head-mounted vision enhancement system for people with low vision. *ACM Transactions on Accessible Computing (TACCESS)* 12, 4 (2019), 1–46.

A Design Probes in the Formative Study

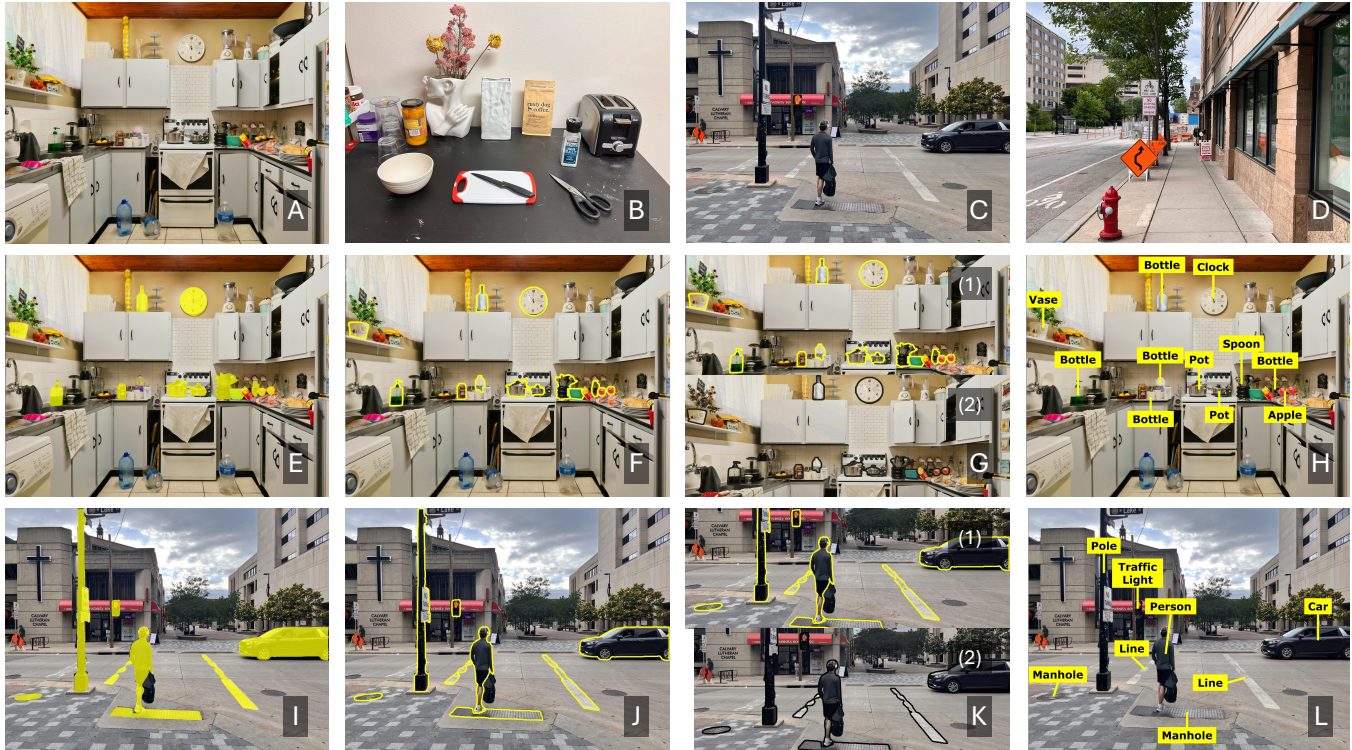


Figure 10: Example scenarios and augmentation designs in the formative study design probe. (A)-(B) Two example kitchen scenarios; (C)-(D) Two example outdoor scenarios; (E) The solid overlay in kitchen with randomly selected objects (vase, bottle, clock, pot, spoon, apple); (F) The static outline in kitchen; (G1-2) The flashing outline in kitchen; (H) The text label in kitchen; (I) The solid overlay in outdoor environment with randomly selected objects (pole, traffic light, person, crosswalk line, manhole, car); (J) The static outline in outdoor environment; (K1-2) The flashing outline in outdoor environment; (L) The text label in outdoor environment.

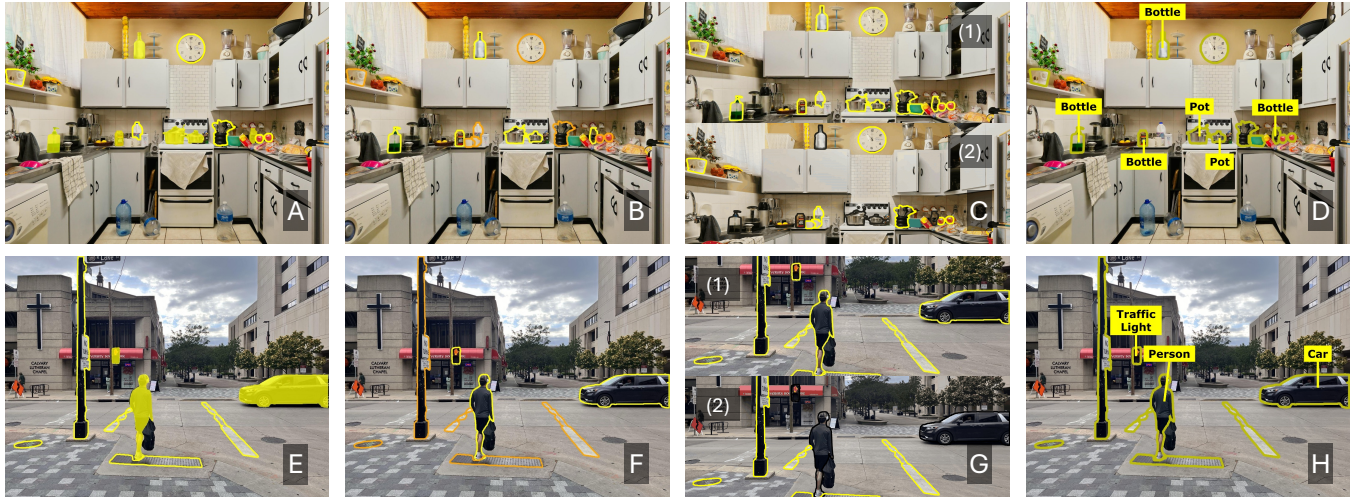


Figure 11: Example distinction methods in the formative study design probe. (A) The *By Form* in kitchen with randomly selected high-importance objects (bottle, pot) in solid overlay and low-importance objects (vase, spoon, clock, apple) in static outline; (B) The *By Color* in kitchen with high-importance objects in yellow static outline and low-importance ones in orange static outline; (C1-2) The *By Visual Effect* in kitchen with high-importance objects in flashing outline and low-importance ones in static outline; (D) The *By Additional Visual Information* in kitchen with high-importance objects in static outline and text label combined, and low-importance ones in static outline only; (E) The *By Form* in outdoor environment with randomly selected high-importance objects (person, car, traffic light) in solid overlay and low-importance ones (manhole, crosswalk line, pole) in static outline; (F) The *By Color* in outdoor environment; (G1-2) The *By Visual Effect* in outdoor environment; (H) The *By Additional Visual Information* in outdoor environment.

B Demographic Information of Formative Study Participants

ID	Age/ Gender	Diagnosis	Legally Blind	Visual Acuity	Field of View	Other Visual Difficulties	Prior AR Experience
P1	55/M	Stargardt disease	N	L: 20/100; R: 20/75	Blurry and hazy spots on central vision	Sensitive to light Cannot distinguish between dark colors	N
P2	58/M	Optic neuropathy	Y	L: 20/70; R: 20/300	Both eyes <19°	Sensitive to light	Y
P3	51/F	Retinopathy of prematurity, Chronic uveitis	Y	L: 20/300; R: 20/125	Peripheral vision loss Full central vision	Sensitive to light	N
P4	85/F	Macular Edema, Glaucoma	N	Unknown, but with blurry central vision	Peripheral vision loss Full central vision	Sensitive to light	N
P5	62/F	Spinal meningitis	Y	L: 20/2200; R: 20/400	Cannot see lower half of vision	Sensitive to light	Y
P6	85/F	Macular degeneration, Charles Bonnet syndrome	Unknown	Unknown, but with blurry central vision	Blurry central vision	Occasional optical illusions	N

Table 5: Participant demographics and visual conditions in the formative study.

C Per-Category Recognition Accuracy of the Fine-Tuned Models

Class	Instances	<i>RTMDet-Ins-l-Kitchen</i> (fine-tuned)					<i>RTMDet-Ins-l</i> (baseline)		
		False Negative Rate	False Discovery Rate	AP@50	AP@75	AP	AP@50	AP@75	AP
Bowl	626	0.364	0.365	0.628	0.492	0.454	0.357	0.294	0.267
Carafe	283	0.078	0.199	0.887	0.843	0.755	0.000	0.000	0.000
Cup	1926	0.268	0.270	0.746	0.608	0.538	0.489	0.388	0.344
Fork	616	0.263	0.210	0.703	0.224	0.316	0.441	0.128	0.193
Glasses	343	0.394	0.197	0.640	0.423	0.385	0.612	0.383	0.375
Jars	1025	0.377	0.330	0.644	0.441	0.404	0.327	0.242	0.218
Knife	982	0.369	0.322	0.582	0.312	0.312	0.392	0.226	0.214
Ladle	61	0.344	0.394	0.627	0.429	0.373	0.000	0.000	0.000
Scissors	109	0.248	0.349	0.740	0.599	0.465	0.519	0.405	0.330
Spatula	278	0.165	0.341	0.750	0.592	0.497	0.000	0.000	0.000
Spoon	929	0.351	0.294	0.566	0.282	0.288	0.304	0.135	0.149
All classes	7178	0.293	0.297	0.683	0.477	0.435	0.313	0.200	0.190

Table 6: Per-category recognition performance of the kitchen model (*RTMDet-Ins-l-Kitchen*) against the baseline (*RTMDet-Ins-l*) on the kitchen test set. *Instances* is the number of ground-truth instances of each class. The *false negative rate* (false negatives / ground truth) indicates the rate of unaugmented important objects, and the *false discovery rate* (false positives / detections) indicates the rate of wrongly augmented non-important objects. For each AP metric, the better model is shown in bold; the fine-tuned model outperforms the baseline on all classes. A baseline score of 0.000 indicates a class absent from MS-COCO (e.g., carafe, ladle, spatula). Detection-level metrics use an IoU threshold of 0.3 and a confidence threshold of 0.45, matching the deployment parameters.

Class	Instances	RTMDet-Ins-l-Street (fine-tuned)					RTMDet-Ins-l (baseline)		
		False Negative Rate	False Discovery Rate	AP@50	AP@75	AP	AP@50	AP@75	AP
Bench	47	0.511	0.395	0.403	0.084	0.182	0.336	0.119	0.175
Bicycle	195	0.364	0.205	0.659	0.126	0.268	0.541	0.127	0.222
Construction cone	41	0.171	0.393	0.781	0.741	0.517	0.000	0.000	0.000
Crosswalk	277	0.791	0.293	0.309	0.174	0.173	0.000	0.000	0.000
Curb	2766	0.387	0.260	0.517	0.187	0.231	0.000	0.000	0.000
Curb cut	337	0.890	0.327	0.150	0.022	0.052	0.000	0.000	0.000
Fence	1857	0.731	0.223	0.363	0.173	0.186	0.000	0.000	0.000
Fire hydrant	16	0.125	0.263	0.864	0.864	0.618	0.792	0.754	0.603
Lamp pole	4579	0.340	0.247	0.415	0.038	0.122	0.000	0.000	0.000
Mailbox	12	0.750	0.625	0.180	0.077	0.101	0.000	0.000	0.000
Motorcycle	183	0.262	0.224	0.752	0.270	0.357	0.623	0.167	0.264
Pedestrian	1315	0.157	0.231	0.835	0.378	0.435	0.738	0.397	0.409
Pedestrian signal*	124	0.387	0.420	0.617	0.385	0.357	0.030	0.017	0.016
Railway track	89	0.596	0.294	0.356	0.165	0.190	0.000	0.000	0.000
Sewer drain	99	0.566	0.511	0.414	0.195	0.230	0.000	0.000	0.000
Sidewalk	2615	0.391	0.202	0.580	0.254	0.286	0.000	0.000	0.000
Traffic sign	908	0.227	0.371	0.762	0.724	0.588	0.106	0.099	0.086
Trash can	202	0.441	0.358	0.590	0.449	0.396	0.000	0.000	0.000
Utility box	177	0.582	0.426	0.417	0.311	0.259	0.000	0.000	0.000
Vehicle	6693	0.104	0.177	0.906	0.790	0.686	0.791	0.578	0.548
Vehicle signal*	552	0.112	0.325	0.845	0.670	0.566	0.674	0.536	0.472
All classes	23084	0.423	0.322	0.558	0.337	0.324	0.221	0.133	0.133

Table 7: Per-category recognition performance of the outdoor model (*RTMDet-Ins-l-Street*) against the baseline (*RTMDet-Ins-l*) on the outdoor test set. *Instances* is the number of ground-truth instances of each class. The *false negative rate* (false negatives / ground truth) indicates the rate of unaugmented important objects, and the *false discovery rate* (false positives / detections) indicates the rate of wrongly augmented non-important objects. For each AP metric, the better model is shown in bold; the fine-tuned model outperforms the baseline on all classes except bench, bicycle, and pedestrian on AP75. A baseline score of 0.000 indicates a class absent from MS-COCO (e.g., sidewalk, curb, curb cut). Detection-level metrics use an IoU threshold of 0.3 and a confidence threshold of 0.45, matching the deployment parameters. *MS-COCO has a single *traffic light* class and does not distinguish between pedestrian light and vehicle traffic light, so the same set of baseline detections is evaluated against both classes.

D Demographic Information of Study I Participants

ID	Age/ Gender	Diagnosis	Legally Blind	Visual Acuity	Field of View	Other Visual Difficulties	Prior AR Experience
R1	63/F	Spinal meningitis	Y	L: 20/2200; R: 20/400	Cannot see lower half of vision	Sensitive to light	Y
R2	20/M	Stargardt disease	N	L: 20/200; R: 20/160	Blurry spots in central vision Full peripheral vision	N/A	N
R3	69/F	Macular dystrophy	Y	L: 20/125; R: 20/160	Right peripheral vision loss due to bleeding	Sensitive to light	N
R4	29/M	Stargardt disease	N	L: 20/100; R: 20/100	Blurry spots in central vision Full peripheral vision	N/A	Y
R5	73/M	Macular degeneration	Y	L: 20/160; R: 20/200	Full	N/A	N
R6	73/F	Cone dystrophy	Y	L: 20/100; R: 20/125	Full	Cannot distinguish between yellow and green; sensitive to light	N
R7	19/F	Stargardt disease	Y	L: 20/200; R: 20/200	Reduced central acuity Full peripheral vision	Sensitive to light	N
R8	54/F	Achromatopsia	Y	L: 20/200; R: 20/400	Full	Full color blind	N
R9	37/M	Partial achromatopsia	Y	L: 20/400; R: 20/400	Full	Difficulty differentiating colors	Y
R10	57/F	Macular degeneration	N	L: 20/125; R: 20/320	Dark spots in central vision	Sensitive to light	Y
R11	41/M	Central scotoma due to damage in optic nerve	Y	L: 20/300; R: 20/400	Central scotoma Full Peripheral vision	Sensitive to light	Y
R12	50/M	Macular degeneration (Best disease)	Y	L: 20/400 R: 20/600–20/800	Hazy spots in central vision Full Peripheral vision	Sensitive to light	Y

Table 8: Participant demographics and visual conditions in the kitchen observation study (Study I).

E Demographic Information of Study II Participants

ID	Age/ Gender	Diagnosis	Legally Blind	Visual Acuity	Field of View	Other Visual Difficulties	Prior AR Experience
T1	73/M	Genetic optic atrophy	N	L: 20/100; R: 20/80	Full	Cannot distinguish between dark colors	N
T2	57/F	Retinitis pigmentosa	Y	L: 20/70; R: 20/70	Both eyes <20°	Cannot distinguish between blue and green	N
T3	62/F	Spinal meningitis	Y	L: 20/2200; R: 20/400	Cannot see lower half of vision	Sensitive to light	Y
T4	64/M	Retinitis pigmentosa	Y	L: 20/40; R: 20/40	Peripheral vision loss Full central vision	Color blind	N
T5	72/F	Cone dystrophy	Y	L: 20/160; R: 20/160	Central vision loss Full peripheral vision	Cannot distinguish between dark colors	N
T6	58/M	Optic neuropathy	Y	L: 20/70; R: 20/300	Both eyes <19°	Sensitive to light	Y
T7	41/M	Retinitis pigmentosa	Y	L: 20/40; R: 20/40	Both eyes <15°	Sensitive to light	N
T8	28/M	Stargardt disease	N	L: 20/100; R: 20/100	Blurry spot in central vision Full peripheral vision	N/A	N
T9	58/F	Retinitis pigmentosa	Y	L: 20/50; R: 20/50	Peripheral vision loss Full central vision	Cannot distinguish between red, pink, and orange	Y
T10	53/F	Achromatopsia	Y	L: 20/200; R: 20/400	Full	Full color blind	N
T11	82/F	Macular degeneration	N	L: 20/50; R: 20/60	L: Central vision loss R: Full	Sensitive to light	N
T12	40/F	Glaucoma; injury on right eye	Y	L: light perception R: blurry vision	R: Peripheral vision loss, full central vision	Sensitive to light	N
T13	60/M	Acute asymmetric optic neuropathy	Y	L: <20/90; R: 20/90	Scotoma spread out across entire vision	Sensitive to light	N

Table 9: Participant demographics and visual conditions in the outdoor navigation study (Study II).