Driver Behavior and Performance with Augmented Reality Pedestrian Collision Warning: An Outdoor User Study

Hyungil Kim, Joseph L. Gabbard, *Member, IEEE*, Alexandre Miranda Anon, and Teruhisa Misu, *Member, IEEE*

Abstract— This article investigates the effects of visual warning presentation methods on human performance in augmented reality (AR) driving. An experimental user study was conducted in a parking lot where participants drove a test vehicle while braking for any cross traffic with assistance from AR visual warnings presented on a monoscopic and volumetric head-up display (HUD). Results showed that monoscopic displays can be as effective as volumetric displays for human performance in AR braking tasks. The experiment also demonstrated the benefits of conformal graphics, which are tightly integrated into the real world, such as their ability to guide drivers' attention and their positive consequences on driver behavior and performance. These findings suggest that conformal graphics presented via monoscopic HUDs can enhance driver performance by leveraging the effectiveness of monocular depth cues. The proposed approaches and methods can be used and further developed by future researchers and practitioners to better understand driver performance in AR as well as inform usability evaluation of future automotive AR applications.

Index Terms— Augmented reality, human performance, automotive, depth cues, three-dimensional displays

1 Introduction

Augmented reality (AR) supplements the real world, rather than replacing it with a virtual world, whereby computation is embodied as perceived virtual elements in physical environments. Given the inherently spatial nature of many promising AR application interfaces, effective interaction with co-existing virtual and physical objects, and thus task performance often relies on a user's understanding of the spatial layout of augmented environments [1], [2]. In most cases, users need to correctly perceive three-dimensional (3D) objects, their spatial arrangements, and egocentric distances to reach, grasp and manipulate objects while navigating through the environment. Ideally, viewing AR graphics through AR displays should be much like viewing the real world such that virtual objects are (1) correctly perceived at intended locations in a 3D space, and, (2) produce physiological responses in the human visual system consistent with real-world viewing experiences. However, in many cases, current AR display technologies are not capable of visualizing virtual objects with all the visual attributes normally afforded by real-world objects [3]-[5]. Therefore, for optimal design of AR applications, an important open question is which depth cues should be available to enable sufficient human spatial perception and task performance in AR.

Driving is one of many promising spatiotemporal tasks that can benefit from AR where computer-generated graphics presented on windshield head-up displays (HUD) can guide drivers' attention to relevant task-related elements and potential hazards [1]. Through HUDs, computer graphics can be either directly overlaid atop real-world referents (*world-fixed* or *conformal*), fixed to certain areas within a display field of view (*screen-fixed*), or associated with but not directly superimposed on the referents (*world-associated*) [1]. Therefore, existing knowledge about human depth perception in AR can inform HUD interface design. However, for optimal design of AR driver interfaces, there are still open questions such as; What are the most appropriate display technologies and visual cue presentation methods for specific driving use-cases? What are the most important or effective depth cues to ensure driver performance in AR? And, do we need stereoscopic head-up displays in the car?

As an initial attempt to address these questions, this work considered two extreme display conditions (in terms of affordable depth cues by current display technologies, see section 2.2 for details) for pedestrian collision warning (PCW) and compared their effects on driver performance in a pedestrian hazard situation. This work extends results from a preliminary study [6], using a *monoscopic display* that conveyed a set of perspective depth cues including linear perspective, relative size and height in the visual field, while a *volumetric display* delivered an additional set of depth cues including binocular disparity, convergence, motion parallax and accommodative focus. This work also aimed to investigate the consequences of conformal AR graphics on driver experience. For this purpose, we examined driver behavioral changes and performance gains associated with visual warnings presented in a conformal manner, as compared to a current PCW driver interface which presents visual warnings in a screen-fixed manner, and a no warning scenario (i.e., no PCW driver interface).

2 Related Work

To better understand the problem space, we briefly examine existing knowledge about (1) human depth perception in natural viewing conditions, (2) capabilities of current display technologies, (3) depth perception in AR viewing conditions, and, (4) effects of available depth cues on human performance in various tasks.

2.1 Human Depth Perception

In natural viewing conditions, humans estimate the 3D structure of the world by combining multiple sources of information, known as *cue integration* through which the brain weights different depth cues based on how reliable or informative they are in a given viewing instance [5], [7]-[9]. Cutting and Vishton [9] present the relative strength of depth cues according to viewing distance, segmenting the space around an observer into three classes: personal (<1.5m), action (1.5~30m) and vista (>30m). In personal space, binocular depth cues (only perceivable by both eyes including binocular disparity and convergence) are very strong and effective but the strength of these cues decreases with distance. In action space, occlusion is the strongest depth cue followed by relative height in the visual field, relative size of the same object at different depths, motion parallax, binocular disparity, convergence and accommodative focus. In vista space, a handful of monocular depth cues (perceivable by one eye, such as occlusion, relative size, relative height and atmospheric haze)

- *Hyungil Kim is with Virginia Tech. E-mail: hci.kim@vt.edu.*
- *Joseph L. Gabbard is with Virginia Tech. E-mail: jgabbard@vt.edu.*
- *Alexandre Miranda Anon is with Honda Research Institute, USA. E-mail: amiranda@honda-ri.com.*
- *Teruhisa Misu is with Honda Research Institute, USA. E-mail: tmisu@honda-ri.com.*

Manuscript received 11 Sept. 2017; accepted 8 Jan. 2018.
Date of publication 19 Jan. 2018; date of current version 18 Mar. 2018.
For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org, and reference the Digital Object Identifier below.
Digital Object Identifier no. 10.1109/TVCG.2018.2793680

are dominant. This suggests that within each space, some cues are stronger and more effective than others in human perception of depth.

2.2 Display Technologies and Affordable Depth Cues

Based on the knowledge about the human visual system, visual displays can be classified into four different families, by their capability of presenting different depth cues (for more details, see [4], [5]). For the consistency of terminology, we will use monocular /binocular for depth cues and monoscopic/stereoscopic for types of displays. Further, we slightly adapted the classification from [5] to include terminology commonly used in the virtual reality community.

Monoscopic displays (family 1), confusingly and commonly known as 2D displays, can convey 3D depth information by perspective projection of 3D structures on to a single display surface. They are capable of providing a group of monocular depth cues, also known as pictorial or *perspective cues* such as occlusion, relative height, relative size, linear perspective, shading and texture gradient [4], [7].

Stereoscopic displays (family 2) can additionally provide binocular depth cues (i.e., *binocular disparity and convergence*) by delivering a slightly different image to each eye, taken from a single, fixed viewpoint [7], [8]. This family of displays presents images onto a fixed focal plane and requires special glasses or headgear to convey separate views to each eye (e.g., 3D TVs).

Stereoscopic with multiple views displays (family 3) also present images on a single focal plane, but provide an additional depth cue afforded by head movements, *motion parallax*, by either presenting many views taken from multiple viewpoints or tracking an observer's head [3], [4]. Typical examples of this family of displays include multi-view displays and VR head-worn displays (e.g., Oculus). When viewing through these displays, moving one's head from side to side produces a different viewpoint of the same scene, leveraging motion parallax to provide more natural viewing experiences.

The last family of displays, *volumetric displays* (family 4) can naturally support *accommodative focus*. They mainly differ from all other families of displays in the use of voxels, 3D equivalents of pixels in fixed-focal-plane displays [4], [10]. These displays illuminate voxels to create virtual objects within a 3D volume, for example, by using either multiple static focal planes (spatial multiplexing or static volume technique) or a moving image plane (temporal multiplexing or swept volume technique) in a 3D space. Since voxels physically occupy a 3D space, many depth cues available in physical objects are also available in virtual objects are satisfied naturally. Binocular disparity and motion parallax exist naturally rather than simulated by displays, and accommodation and convergence are not in conflict as is typical in common VR and AR displays [4], [8], [11]. However, this family of displays are usually expensive and require complex optical mechanisms, resulting in bulky hardware form factors with small field-of-views [3], [4], [12].

2.3 Human Depth Perception in AR

Human observers see augmented environments through AR displays where both physical and virtual objects can affect perceived depth of each other. Empirical studies have shown that depths of physical objects are more accurately judged than those of virtual objects in both personal and action space [13]-[15]. In personal space, Ellis and Menges [16] reported degraded depth judgment performance with a monoscopic display as compared to a stereoscopic one. Swan et al. [17] compared egocentric depth judgment performance in action space with a stereoscopic optical see-through head-worn display. They found that egocentric depth to virtual objects is underestimated in AR but even more underestimated in VR. Spatial relationships between virtual and physical objects can also affect perceived depth to virtual objects. Kirkley [18] found that placing virtual objects on the ground plane improved depth judgments. In a study on automotive AR, Tonnis and Klinker [19] found that the direction of a virtual arrow is better perceived when it is attached to a car body using a virtual pole rather than hovering over the car. Virtual objects can also affect perceived depth to physical objects. Smith et al. [20] reported that distance to a pedestrian was underestimated when augmented and viewed through a HUD that overlays a virtual box atop the pedestrian. However, all of the aforementioned user studies have used fixed-focal-plane displays which have inherent limitations in conveying depth cues.

For the automotive application of AR, researchers have conducted empirical studies on depth perception of virtual objects in action space, using prototypes of optical see-through HUDs which encompass all four families of displays discussed earlier: monoscopic, stereoscopic, multi-view and volumetric display. Hotta et al. [21] improved depth perception of monoscopic AR HUDs by using an animation effect of AR cues that move toward the real world target. Broy et al. [22] examined design factors for stereoscopic AR HUDs suggesting that *5~8m* of focus distance would be a good trade-off for fixed-focal-plane stereoscopic displays in consideration of both observers' depth judgment accuracy and visual comfort. However, they argued that with fixed-focus-plane displays, AR graphics in a depth layer can hide objects in other depth layers. Takaki et al. [23] prototyped an autostereoscopic, multi-view windshield display that supports not only accurate binocular disparity but also motion parallax. In their experiment using a perceptual matching task within a range of 50m, participants showed significant improvement in depth judgments with the multi-view display as compared to a stereoscopic display, suggesting that motion parallax is one of the most dominant and effective depth cues in action space. Finally, Bark et al. [24] prototyped a 3D volumetric AR HUD and conducted a depth judgment study in an outdoor setting where observers were asked to make a forced-choice among physical targets as to which was perceived to be closest to virtual objects placed at *9~26m*. Depth judgment accuracy with the volumetric display was higher (97%) than that with a monoscopic display (32%). In sum, these studies suggest that the more depth cues are available, the better egocentric depth perception is expected. It is noticeable that most studies quantified depth judgment capability of stationary observers, which might differ from that of fast-moving observers such as vehicle drivers.

2.4 Human Performance in AR / VR

Despite known benefits of advanced display technologies in human depth perception, their effects on human performance in higher level spatial tasks are modulated by several factors. A recent comprehensive review [25] on stereoscopic displays and human performance in various tasks reports that stereoscopic displays are beneficial to spatial tasks; out of 184 experiments, 60% reported benefits of stereoscopic displays over monoscopic displays, 15% reported marginal benefits only in some performance measures, and 25% reported no benefit. The review also shows differential benefits of stereoscopic displays depending upon the type of spatial tasks: distance judgment (57%), visual search (65%), spatial understanding (52%), object manipulation (67%) and navigation (42%). In addition to the task-type, an in-depth analysis revealed other modulating factors such as (1) salience of monocular depth cues, (2) task difficulty, (3) viewing distance, (4) user expertise, and, (5) movement. The review suggests that stereoscopic displays are less beneficial when strong monocular depth cues are available for easy tasks requiring far-field interactions done by experienced moving operators. In spite of the authors' comprehensive efforts, most experiments reviewed (98%) were conducted in either VR or video see-through AR, where depth-cue-rich real-world views were replaced by either virtual worlds or video feeds which offer different depth cues and thus consequences on human performance, as compared to optical see-through AR [2].

3 Experiment

To explore the effects of visual warning presentation methods on driver performance and behavior in pedestrian hazard situations, we ran a user study in an outdoor setting, aiming to answer specific research questions;

- What are consequences of visual warnings presented via conformal graphics on driver behavior and performance?
- Do visual warnings presented on volumetric HUDs have benefits over monoscopic HUDs?

We invited participants to a large parking lot where they needed to manage the actual demands of driving in a controlled but realistic driving scenario. Participants drove a test vehicle while braking for any cross traffic with assistance from visual warnings on a HUD. Drivers' behavior and performance were recorded by eye tracking glasses, in-vehicle cameras and a global positioning system (GPS).

3.1 Participants

Sixteen licensed drivers with normal or corrected-to-normal vision participated in the study. On average, they were 42 years old (SD=8), had 23 years (SD=9) of driving experience, and drove 11,200 miles (SD=4,854) per year. Three participants had some minimal experience with head-worn AR (e.g., Google Glass), but none had experience in driving a car with AR HUDs.

3.2 Apparatus

3.2.1 Test Vehicle

We equipped a 2009 Honda Odyssey test vehicle with various devices to record driver behavior and performance (Fig. 1a). A high accuracy real-time kinematic GPS (OxTS RT4003 with smaller than *20cm* localization error) was used to record the test vehicle's position, velocity, and acceleration at *200Hz*. Two cameras (GoPro Hero3+) recorded drivers' foot behavior and the external scene at *24Hz*. Eye tracking glasses (SMI ETG with *80°×60°* tracking range) recorded participants' gaze behavior at *30Hz*.

3.2.2 Head-Up Display

In this experiment, the test vehicle also contained an in-vehicle prototype of an optical see-through HUD (Fig. 1b). It is a projection-based volumetric display with a swept-volume technique [11] using fast switching image planes within a range of focal distance between *8m* and infinity (*0.125D ~ 0D*); affording flicker-free appearances of virtual objects in the 3D space with about *17°* circular field of view. This volumetric HUD is capable of providing not only perspective depth cues available in monocular displays but also additional depth cues such as binocular disparity, motion parallax, convergence, and accommodation. Furthermore, by presenting AR graphics at the exact same focal distance as their real-world referents, observers are not forced to switch focus between virtual and physical objects. Therefore, it helps attenuate perceptual consequences (e.g., visual fatigue, discomfort and distorted depth perception) of incomplete or conflicting depth cues [8], [26]-[28].

Fig. 1. (a) The test vehicle equipped with a GPS, eye tracking glasses, cameras, and, (b) an in-vehicle volumetric AR head-up display.

3.2.3 Parking Lot

This experiment was conducted in a large three-sided parking lot (150m × 100m, Fig. 2a) which was filled with many parked vehicles to provide participants with a realistic driving environment. A one-way driveway passes through the parking area with a 15mph speed limit. The parking lot consisted of three zones and one of them was designated for the participants to park the test vehicle as a part of an experimental scenario. Note that we chose a real roadway to better understand human performance in AR, which cannot be replaced by driving simulators where most of the real-world depth cues are confounded or not available.

3.3 Driving Tasks

For the better ecological validity of the study, we gave participants a realistic parking lot scenario similar to their everyday life experiences. Participants were asked to drive in the parking lot to find an available spot within the reserved parking zone (Fig. 2a). First, participants were asked to approach the entrance of the parking lot which was a starting line for each driving trial. They were asked to drive a constant speed of 15mph (except during turns where they could slow down) until they arrived at the reserved parking zone. While driving, they were asked to brake for any cross traffic (e.g., backing-up vehicles or stepping-out pedestrians, Fig. 2b). Participants were instructed not to swerve or detour but to make a complete stop for any cross traffic. No visual warnings were given in the control condition, while visual warnings on the HUD were given for experimental conditions. The driving task imposed actual driving demands (visual, cognitive and manual) on participants as they drove and reacted to real road events.

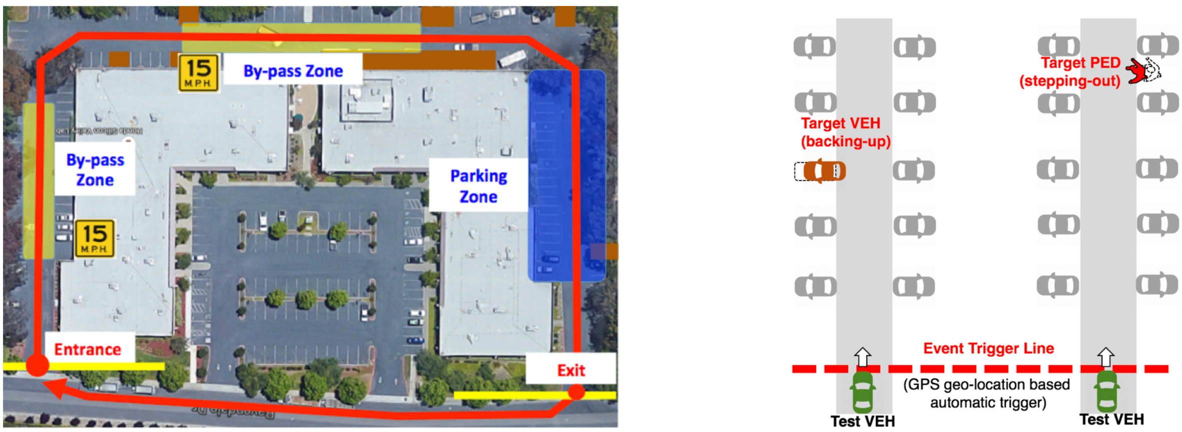

Fig. 2. (a) A large parking lot (three-sided 150m X 100m) filled with parked vehicles, was used for the test site. Participants were asked to drive the test vehicle to find an available spot in the reserved parking zone. (b) Participants were asked to drive constant speed of 15mph, and brake for any cross traffic such as backing-up vehicles or stepping-out pedestrians. Visual warnings on the HUD were automatically activated using real-time GPS and predefined geolocation of the trigger line.

3.4 Procedure

The experimental user study consisted of four sessions; (1) pre-test survey, (2) practice trials, (3) experimental trials, and, (4) post-test interview. Upon participants' arriving at a preparation room, an experimenter introduced the study briefly and surveyed participants' demographic information and driving experience. Then, the experimenter guided participants to the test vehicle in the parking lot. During the practice session, participants were given an in-car orientation of the overall procedure and driving tasks. Eye tracking glasses were calibrated for each participant inside the test vehicle. Two practice trials were given without any measurement so that participants could get familiar with driving the test vehicle, the layout of the test site and the presence of AR HUD technology. In the experimental trials, participants were asked to drive the test vehicle to perform the parking task following a pre-defined route (the red line in Fig. 2a). During the driving trials, an experimenter sat in the backseat and traffic was controlled by experimenters. After completing all driving trials, participants gave comments on their experience. The experiment took about an hour for each participant.

We simulated V2X communication technology [29]-[31] for vehicle and pedestrian localization in accordance with road events (i.e., backing-up vehicles and stepping-out pedestrians). For this purpose, we pre-defined the location of road events and associated trigger points. When the test vehicle passed trigger points, the experimenter sent signals to the backing-up vehicles or pedestrians via a walkie-talkie to activate the road events. These events were automatically transmitted to the HUD by the GPS to trigger visual warnings (Fig. 2b).

3.5 Experimental Design

A two-factor repeated measures experiment was conducted, where each participant experienced all experimental conditions; 4 levels of visual warnings and 2 levels of distance to the pedestrian. We counterbalanced the presentation order of experimental conditions to account for learning and ordering effects. To reduce drivers' anticipation of pedestrians, no-event trials were randomly added and the side of road events (left or right) was randomized. We also introduced backing-up vehicle events during the practice trials.

3.5.1 Visual Warning (no, current, monoscopic, volumetric)

The real-world pedestrian events were augmented by visual warnings with four levels, which provided different sets of depth cues. In the *control condition* (i.e., no warning), the visual stimuli available to the driver were only those associated with the real pedestrians. In *warning conditions*, both a visual warning and a pedestrian appeared at the same time. The *current warning interface* condition was inspired by currently available PCW on the market. Specifically, a "BRAKE" indicator (text) was shown at the center of the HUD (Fig. 3a), to notify drivers of the presence of a pedestrian in the vehicle's path. This cue contained no depth information about the real pedestrian. Both *monoscopic* and *volumetric warning* display conditions presented a virtual shadow [32], [33], to inform drivers of the direction and distance to an approaching pedestrian using conformal graphics (Fig. 3b). Note the term *conformal* is used hereafter to refer to both the monoscopic and volumetric virtual shadow. A virtual shadow is a dome-shaped conformal virtual object combined with a tether that appears on the ground at the location of the real pedestrian. The absolute size (diameter) of the dome was *1.0m* but it's apparent angular size from the driver's viewpoint varied between $2°$ and $8°$ depending upon the viewing distance as the driver moved (Fig. 4). In the monoscopic display condition, the HUD presented the virtual shadow by a perspective projection of the 3D virtual object onto the nearest (*8m*) focal plane of the HUD. Therefore, it provided a set of perspective depth cues including linear perspective, relative size, and relative height in the visual field. In contrast, the volumetric display condition directly generated a 3D virtual object on the ground at the same distance as the pedestrian, providing a set of additional depth cues including binocular disparity, convergence, motion parallax, and accommodative focus.

Fig. 3. Visual warnings presented on the HUD. (a) The current warning condition shows "BRAKE" sign to inform the presence of a pedestrian, and (b) the monoscopic and volumetric display conditions show a "virtual shadow" to inform the distance to and direction of an approaching pedestrian.

3.5.2 Distance to Pedestrian (near, far)

The real-world road events happened at two different distances which might affect drivers' depth judgment and risk perception yielding different responses. The specific distances to the pedestrian were chosen to correspond with the time to collision (TTC); one of many critical factors that influences drivers' braking responses [34]. The near target pedestrians stepped out when TTC was *2.5 second* (*16.7m* at *15mph*) which represents an urgent situation that might instigate last second hard braking responses [34]-[36]. The far target pedestrians appeared at *5.0 second* TTC (*33.5m* at *15mph*) which represents a normal situation that might induce drivers' timed-normal braking responses.

3.6 Measures

Human behavior and performance in braking tasks can be evaluated by various measures that show the driver's capability of detecting, reacting to, and finally stopping apart from a road hazard. Measures can be either time-based (e.g., gaze reaction time, pedal reaction time and time to stop) or distance-based (e.g., perception, reaction, and stopping distance) [34]-[38]. In this study, we used distance-based measures from the GPS log which was synchronized with eye-tracking glasses and in-vehicle cameras. The distance-based measures allowed us to examine the process of braking as referenced to the initial distance gap to the pedestrian. All measures were corrected by an

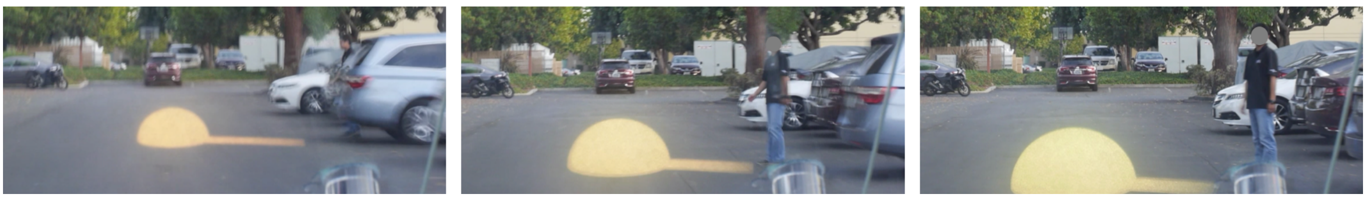

Fig. 4. The virtual shadow in action. The shadow is integrated into the real world based on the geolocation of the pedestrian (GPS coordinate) so that its *apparent* size and location (relative size and height in the driver's visual field) change as the driver approaches the target pedestrian, similar to the real shadow of the pedestrian.

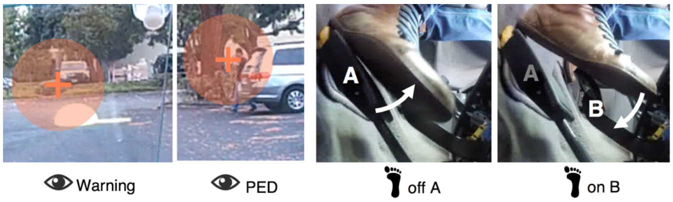

Fig. 5. Drivers (a) gaze and (b) foot behaviors are analyzed to capture critical events including gaze-on warning, gaze-on pedestrian, foot-off accelerator pedal, and foot-on brake pedal.

equation recommended by SAE J299 [39] to account for the discrepancy between actual and target speed of the test vehicle (i.e., *15mph* for this experiment).

$$d_{corrected} = d_{measured} \times (v_{target} / v_{actual})^2$$

where, $d_{corrected}$ = corrected value of dependent variable, $d_{measured}$ = measured value of dependent variable, v_{target} = target entrance speed, v_{actual} = actual entrance speed.

3.6.1 Driver Behavior

We analyzed drivers' gaze and foot video, timestamping drivers' behavioral events to obtain corresponding distance-based measures from synchronized GPS logs. Driver behavior was evaluated by four dependent variables: (1) *gaze-on visual warning*, (2) *gaze-on pedestrian*, (3) *foot-off accelerator pedal*, and, (4) *foot-on brake pedal*. Each dependent measure was calculated as a distance (in meters) between the trigger point and corresponding location in which the event occurred (e.g., *foot off accelerator pedal* represents the distance traveled between the trigger point and the moment in which a driver took their foot off the accelerator pedal). By considering four measures together, we also defined a categorical dependent variable,

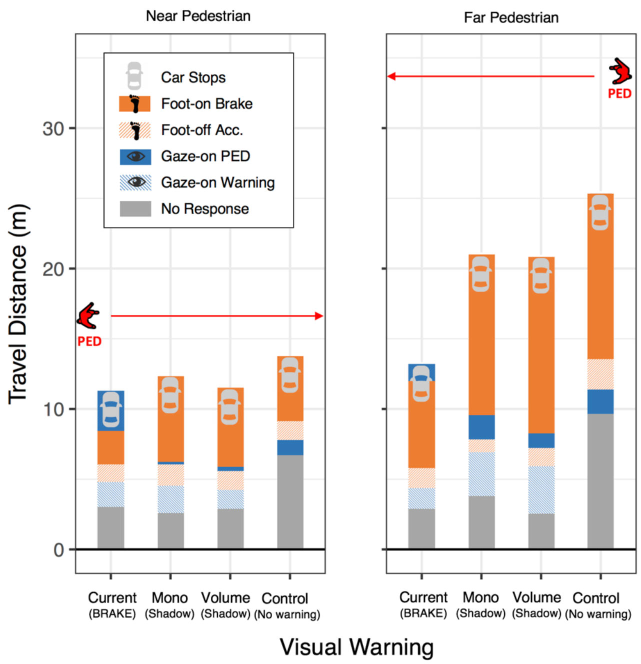

Fig. 6. The entire process of braking is visualized by distance measures referenced to the initial gap from the pedestrian. The test vehicle's mean travel distances in experimental conditions are plotted showing driver behavioral events such as gaze-on warning, gaze-on pedestrian, foot-off accelerator pedal, foot-on brake pedal and finally distance to test vehicle coming to a stop.

(5) *behavioral pattern*, based on the sequence of drivers' discrete actions that characterize drivers' overall response behavior.

Additionally, drivers' gaze behavior was classified into fixations, saccades, and smooth pursuits (for more details, see [40]). In a driving context, smooth pursuits are important measures to address because all objects in driving scenes (even stationary objects) are continuously moving from the driver's moving viewpoint. Therefore, for the remainder of this paper, we will refer to *gaze on objects* as the smooth pursuit of objects. Since most gaze analysis software on the market do not support automatic analysis of smooth pursuits, we manually identified smooth pursuits based on a set of criteria. Specifically, we set threshold values on the location and time duration for a gaze to be considered a smooth pursuit. For the location threshold, we used the circular area within the radius of *2.5°* around each gaze point to represent human foveal vision [41]. For the duration threshold, *120 ms* was used which is the same threshold used in standard eye tracking software systems (e.g., SMI ETG) use to identify fixations [42]. With these thresholds in place, we identified the set of *gaze on visual warning* and *gaze on pedestrian* events that occurred when drivers' gaze resided on the area around the AR visual warnings and actual pedestrians, respectively (Fig. 5a).

Processing and coding video of drivers' footwell allowed us to determine when drivers' pedal maneuvers began and ended. *Foot-off accelerator pedal* response was timestamped when drivers lifted their foot off the accelerator pedal, while *foot-on brake pedal* response was timestamped when drivers started pressing the brake pedal (Fig. 5b).

As mentioned, we classified *behavioral patterns* by the sequence of drivers' responses. For example, pattern A was defined as a series of the following driver reactions: gaze-on warning followed by gaze-on pedestrian, foot-off accelerator pedal, then foot-on brake pedal. Pattern B was defined as gaze-on warning followed by foot-off accelerator pedal, gaze-on pedestrian, then foot-on brake pedal. Pattern C was defined as gaze-on warning, followed by foot-off accelerator pedal, foot-on brake pedal, and gaze-on pedestrian.

3.6.2 Task Performance

We also analyzed the test vehicle's deceleration profiles (changes in position, velocity, and acceleration over time) during each braking maneuver to extract values for dependent variables such as (1) *stopping distance*, and, (2) *peak deceleration* [34], [35], [37]. The *stopping distance* was defined as the test vehicle's total travel distance between passing the trigger point and completely stopped. It was used as a measure of "effectiveness of braking" to account for the risk of forward collision. The *peak deceleration* was derived from the acceleration profile logged during braking and was used as a measure of "smoothness of braking" to account for the risk of rear-end collision (by a closely following "tailgating" vehicle) that might be caused when drivers urgently brake.

3.7 Data Analysis

To identify potential main and interaction effects of visual warning and distance to the pedestrian on quantitative measures of driver behavior and performance (i.e., travel distances at gaze and foot events, stopping distance and peak deceleration), we performed two-way repeated measures analysis of variances (ANOVA). For the categorical variable, behavioral pattern, we conducted a Durbin's chi-square test, which is a non-parametric equivalent of the repeated measures ANOVA.

4 Results

We analyzed data from fourteen participants excluding two drivers who completely ignored the visual warnings and relied on their own driving skill. Eye-tracking data from two participants was also excluded from further analysis due to poor quality of data (e.g., scattered or lost gaze data due to reflected ambient light). In the no warning condition, we observed two instances where the driver stopped after passing the pedestrian, which were included in the analysis. Fig. 6 visualizes the entire process of braking with distance-

Table 1. Driver behavioral changes and performance gains by visual warning, as compared to the no warning condition. Effect sizes are reported by % differences, followed by mean differences and 95% confidence intervals of the mean differences. Blue-downward arrows indicate improvement (e.g., reduced stopping distance or lower peak deceleration), while red-upward arrows indicate deterioration of performance. Only statistically significant results are reported with details.

Exp. Conditions / Dependent Variables	PED @ near			PED @ far		
	Current (BRAKE)	Mono (Shadow)	Volume (Shadow)	Current (BRAKE)	Mono (Shadow)	Volume (Shadow)
Gaze-on PED travel dist.	25.32%	-9.79%	-13.17%	▲ 24.02% 2.32m [0.48, 4.16]	▼ -19.31% -1.86m [-3.71, -0.02]	▼ -25.21% -2.43m [-4.28, -0.59]
Foot-off A pedal travel dist.	▼ -38.57% -3.40m [-5.82, -0.97]	▼ -41.32% -3.42m [-5.84, -1.00]	▼ -46.22% -3.79m [-6.16, -1.43]	▼ -61.47% -7.03m [-9.34, -4.72]	▼ -39.19% -4.98m [-7.29, -2.66]	▼ -47.64% -5.63m [-7.99, -3.27]
Foot-on B pedal travel dist.	▼ -34.01% -3.61m [-6.01, -1.20]	▼ -32.44% -3.02m [-5.42, -0.62]	▼ -36.15% -3.48m [-5.84, -1.13]	▼ -57.21% -7.82m [-10.14, -5.50]	▼ -29.42% -4.50m [-6.82, -2.18]	▼ -38.75% -5.45m [-7.81, -3.10]
Stopping Distance	▼ -17.86% -2.33m [-4.60, -0.06]	-10.91%	-16.93%	▼ -49.91% -12.18m [-14.36, -9.96]	▼ -17.15% -4.67 [-6.87, -2.47]	▼ -17.79% -4.52m [-6.72, -2.32]
Peak Deceleration	▼ -19.86% -0.08g [-0.13, -0.04]	▼ -20.34% -0.10g [-0.15, -0.05]	▼ -14.67% -0.06g [-0.11, -0.02]	▲ 34.46% 0.08g [0.04, 0.12]	3.89%	-2.58%

based measures as referenced from the trigger point (*0.0m*) to the pedestrian position (*16.7m* for near condition and *33.5m* for far condition). Descriptive statistics of drivers' mean responses are reported in Appendix A. Repeated-measures ANOVA tests revealed significant main and interaction effects of visual warning and distance to the pedestrian on all 6 quantitative measures (Appendix B). Therefore, we performed post hoc contrast tests for planned comparisons among experimental conditions with Tukey's adjustment for multiple comparisons [43]. For better practical interpretation of the effect size, we report % differences between experimental conditions normalized by mean responses in the control condition that indicate performance gains associated with the visual warnings. Therefore, changes in drivers' responses are reported in the form of ES (d [CI_{lower}, CI_{upper}]), where ES = effect size, d = mean difference between experimental conditions and the control condition, and [CI_{lower}, CI_{upper}] = lower and upper limits of 95% confidence interval on the mean difference. Only statistically significant differences are reported herein with details such as mean differences, 95% confidence intervals, and effect sizes, based on guidelines from Cumming [44]. Table 1 summarizes driver behavioral changes and performance gains resulted from visual warnings, as compared to the control (i.e., no warning) condition.

4.1 Behavioral Changes

Overall, drivers showed a different *behavioral pattern* when warnings were presented via conformal graphics, as compared to both no warning and current warning conditions (Fig. 7); A Durbin's $\chi^2(3) = 34.87, p < 0.001$. Post hoc pairwise comparisons revealed differences among all conditions except for between the monoscopic and the volumetric condition. With conformal graphics, most drivers (62% in volumetric; 41% in monoscopic display conditions) exhibited behavioral pattern B such that once they gazed upon the visual warning, they began taking their foot off the accelerator pedal, looked at the real pedestrian, and finally started pressing the brake pedal. If we combine pattern A and pattern B, most drivers (76% in volumetric, 77% in monoscopic display conditions) looked at the pedestrian before they began pressing the brake pedal when visual warnings were presented via AR conformal graphics. With the current warning, all drivers reacted to the visual warning and looked for the pedestrian *after* pressing the brake pedal (pattern C).

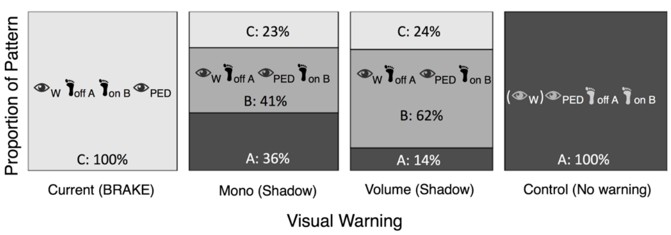

Fig. 7. When visual warnings were presented via conformal AR graphics (i.e., in both monoscopic and volumetric display conditions), most drivers (77% in monoscopic, 76% in volumetric condition; pattern A + B) looked at the real pedestrian before pressing the brake pedal. In the current warning condition, all drivers reacted to the visual warning and looked for the pedestrian afterward.

The *gaze-on pedestrian* distances decreased when warnings were presented via AR conformal graphics in the far pedestrian condition, as compared to no warning condition (Table 1 and Fig. 8b); in monoscopic display condition by 19.31% and volumetric display condition by 25.21%. With the current warning, drivers traveled even farther (24.02%) before looking at the pedestrian, as compared to no warning condition. The same tendency of gaze behavior was observed in the near pedestrian condition as well but was not found to be statistically significant. No statistically significant differences were found between monoscopic versus volumetric warning conditions.

The *foot-on-brake* distances also decreased when drivers viewed AR conformal graphics as compared to no warning condition (Table 1 and Fig. 8d). In the near pedestrian condition, distances decreased in the monoscopic display condition by 32.44% and the volumetric display condition by 36.15%. No differences were found among visual warning conditions. In the far pedestrian condition, the foot-on-brake distances decreased in the monoscopic display condition by 29.42%, the volumetric display condition by 38.75%, and the current warning condition by 57.21%. Participants' foot responses were not significantly different when using monoscopic and volumetric displays.

4.2 Driver Performance Gains

Drivers stopped in shorter distances when visual warnings were given by conformal AR graphics in the far pedestrian situation, as compared to no warning condition at the same distance (Table 1 and Fig. 8e). The volumetric display condition showed a reduction in stopping distance by 17.79%, followed by monoscopic display condition by 17.15%. The current warning condition (i.e., "BRAKE" sign) showed even more reduction by 49.91%. In the near pedestrian condition, warning by conformal AR graphics did not show any reduction in stopping distance, as compared to both no warning and the current warning condition (Table 1 and Fig. 8e). Differences between monoscopic and volumetric warning conditions were not statistically significant.

All visual warnings reduced *peak deceleration* of the vehicle in the near pedestrian condition, as compared to no warning condition (Table 1 and Fig. 8f). The peak deceleration decreased in the monoscopic display condition by 20.34%, the volumetric condition by 14.67%, and

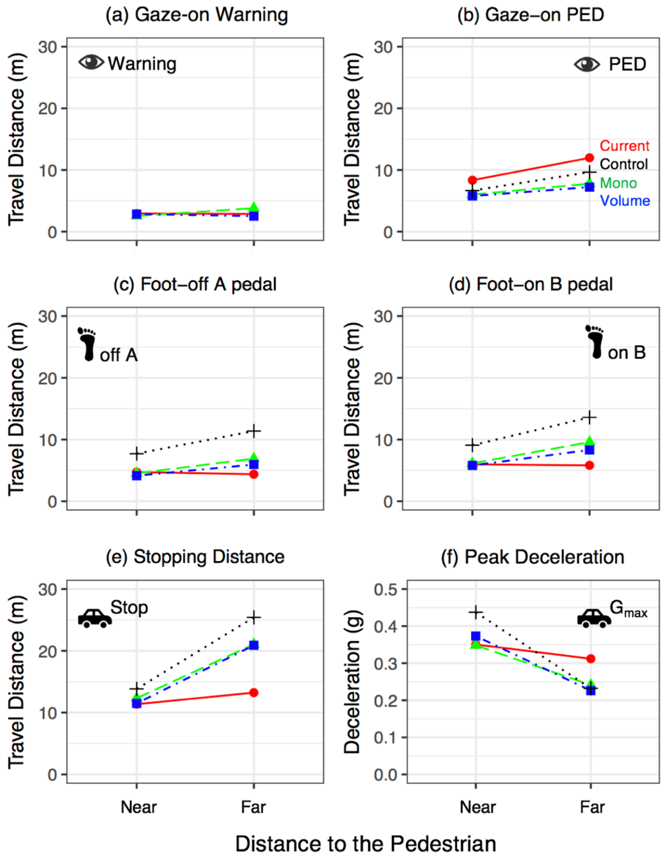

Fig. 8. Distance-based measures on drivers' mean responses; gaze-on warning, gaze-on pedestrian, foot-off accelerator pedal, foot-on brake pedal and total stopping distance for the braking task. Peak deceleration was measured in *g* relative to gravity (1 $g = 9.807\ m/s^2$). To avoid visual clutter, standard errors of the means were reported in a table (see Appendix A).

the current warning condition by 19.86%. In the far pedestrian condition (Table 1 and Fig. 8f), warning by conformal AR graphics did not show any reduction in peak deceleration, while the current warning resulted in even higher peak deceleration by 34.46%, as compared to the no warning condition. Post-hoc contrast tests did not find any differences in peak deceleration between monoscopic and volumetric warning conditions for both near and far pedestrians.

5 Discussion

The key findings from the empirical study are that (1) visual warnings presented via conformal AR graphics were associated with qualitatively different driver behavior resulting in improved braking performance as measured by stopping distance and peak deceleration, and, (2) the monoscopic HUD was as effective as the volumetric HUD for braking performance. We further discuss implications of these findings on warning interface design, followed by limitations of this study for generalization of the results.

5.1 Benefits of Conformal Presentation

An integrated analysis of drivers' behavioral patterns and performance revealed the advantage of conformal AR graphics in guiding drivers' attention and its positive consequences in driver perception, localization, and reaction to road hazards. The resulting behavioral pattern (Fig. 7) and reductions in distance measures (Table 1) suggest that the conformal AR graphics warning (i.e., the virtual shadow presented via either monoscopic or volumetric displays) quickly demanded drivers' attention (gaze-on warning, Fig. 8a), helped them start reacting to a hazard (foot-off accelerator pedal), guided their attention to the hazard (gaze-on pedestrian) before they executed a confirmative reaction (foot-on-brake), and resulted in appropriate responses to the hazard (stopping distances and peak decelerations) which were modulated by the distance to the pedestrian (Fig. 9).

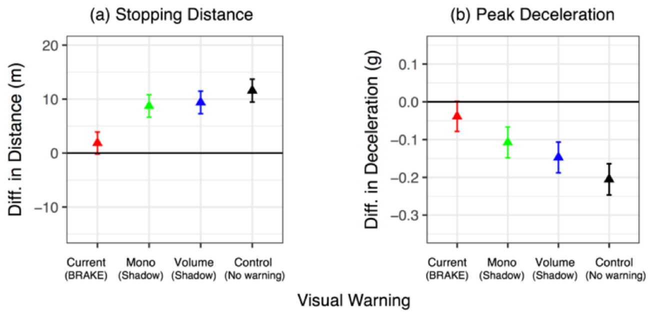

Fig. 9. The driver's responses were modulated by the distance to the pedestrian when the warnings were given by conformal AR graphics (in both monoscopic and volumetric display conditions). Mean differences in (a) stopping distance and (b) peak deceleration between near and far pedestrian conditions are plotted with 95% confidence intervals.

The benefits of conformal visual warnings are even more obvious when compared to driver's responses in the current warning condition. Table 1 shows an overview of driver performance gains by visual warnings, relative to the no warning condition. These results suggest that the "BRAKE" sign induced earlier braking (foot-on-brake) and resulted in improvement of one performance measure (i.e., stopping distance) at the expense of another performance measure (i.e., peak deceleration). In fact, with current warnings, drivers showed similar responses regardless of the distance to the pedestrian (The 95% CIs intersect with zero in Fig. 9a and 9b). Especially for the far pedestrian, the current warning's lack of attention-guiding cues resulted in *unnecessary* hard braking which increases the risk of rear-end crashes by following vehicles. Post-test interviews of participants also revealed less preference for the "BRAKE" sign mainly because participants felt they were (1) unnecessarily surprised by the warning for the far pedestrian, and, (2) not comfortable or confident with braking without locating pedestrians.

Conversely, the conformal visual warnings were associated with balanced driver performance across all measures, while still showing performance gains as compared to the no warning condition (Table 1). The visual warning presented via conformal AR graphics enabled drivers to perceive the threat earlier, brake *after* locating the hazard which resulted in smoother normal braking while ensuring sufficient gaps from the pedestrian.

5.2 Efficacy of Monoscopic Displays

The results also suggest that monoscopic displays can provide sufficient depth information to ensure human performance in AR driving conditions such as those studied herein. This finding further inspires the authors to reconsider unique characteristics of driving contexts that might affect human depth perception and account for no statistically significant differences between the monoscopic and volumetric display conditions.

One possible explanation for the results is the reliability or *effectiveness of depth cues in the space* those tested herein. The monoscopic HUD presented AR graphics at 8m (0.125D) regardless of the distance to pedestrians while the volumetric HUD presented visuals at 16.7m (0.6D) or 33.5m (0.3D) which are the same distance as actual pedestrians. The maximum difference in focal depth among conditions (0.095D) is smaller than human eyes' sensitivity (depth of focus 0.25~0.3D [7]). This suggests that additional depth cues provided by the volumetric HUD (e.g., oculomotor cues such as convergence and accommodation) might not be strong enough to make significant differences in drivers' depth perception at these distances. It is also noticeable that, in typical driving context, drivers need to be informed in advance to make appropriate navigational choices or avoid collision (e.g., 16.7m correspond with 2.5 second time to collision at 15mph) so in this far action space, monocular perspective depth cues might be as effective.

Further, the *environmental context* might account for no statistically significant differences between monoscopic and volumetric display conditions observed in this driving context. In fact, there is evidence that shows environmental contexts do effect human egocentric depth judgment. For example, Lappin et al. [45] compared human performance in judging the midpoint to a familiar object in different environments (e.g., an open large field, a hallway, and a lobby) and found that the environmental context affects human depth judgment. The large parking lot in our experiment might affect drivers' depth perception by providing natural depth cues (e.g., linear perspective in a roadway; relative sizes and heights of many parked vehicles). It is also possible that presenting the virtual shadow on the ground plane could affect perceived depth to the shadow, as supported by another empirical study [18]. Therefore, in the driving context, monoscopic displays might be able to deliver sufficient depth information by leveraging the depth-cue-rich real-world roadway environment.

Finally, drivers' *motion-in-depth* at speed might account for the no difference between two display conditions, since the resulting optical flow from motion is known to yield additional depth cues [46]-[48]. For example, Regan and Beverley compared the effectiveness of monocular and binocular cues from motion-in-depth and argued that precise judgment of motion-in-depth can be made even without binocular vision in some viewing conditions such as a pilot's landing task [49]. There may be more work needed to better understand effects of motion-in-depth on human depth perception in AR.

5.3 Limitations and Open Research Questions

The proposed experimental design, environment, and task scenario have the advantage of better ecological validity (as compared to laboratory or simulator studies) but also have limitations for generalization of our findings to other situations.

The effectiveness of a set of monocular depth cues affordable by monoscopic HUDs on driver performance in AR was supported by our experimental results. However, the possible reasons for this finding (i.e., the effectiveness of individual depth cue on driver depth perception in AR) were not directly supported by this experiment, even though we provided some promising evidence from the literature about modulating factors such as action space, environmental context, and motion-in-depth. Systematic investigations of these factors are required to better understand human depth perception in a moving, dynamic AR driving context.

The effectiveness of our monoscopic HUD (focal distance = *8m*) for driver performance may not be observed in HUDs currently on the market which typically have shorter focal distances (usually between *2* and *3m* [1]). As suggested by an empirical study with a stereoscopic HUD [50], a monoscopic HUD's focal distance may affect driver depth judgment, behavior and performance. Future studies could address the scalability of our findings by systematic investigation of different AR HUD virtual image distances.

Another limitation of this study is the use of different visual stimuli across experimental conditions, especially for the current warning interface condition. The comparison between monoscopic and volumetric conditions does not have any concern. However, the use of "BRAKE" sign might limit the validity of our comparison with visual warnings presented via conformal AR graphics, even though such comparison has a practical meaning to answer a simple question; How much better is an AR-based warning interface than the current one? Lastly, we note that the "BRAKE" sign might further affect drivers' responses, as Lorenz et al showed the different consequences of command display versus status display in driving contexts [51].

The depth-cue rich parking lot was an ecologically valid setting but the results from this environment may not apply to other situations such as nighttime-driving which is a more challenging environment in terms of pedestrian collision avoidance. Moreover, depth-based information available in daytime-driving would be attenuated greatly at night including the actual pedestrians, the real-world referents of the virtual shadow, and other critical sources of information such as ground plane and optical flow. Future studies should address the scalability of these findings by performing similar tests in various environmental and lighting contexts.

Additionally, the studied task scenario included only parts of task demands of driving and the results from this study might not be applied to other types of driving tasks. Specifically, the collision avoidance task requires significant bottom-up, stimulus-driven information processing whereby unexpected road events initiate drivers' information processing for appropriate reaction [52]. Other types of driving task, such as navigation, may require a top-down, goal-driven information processing whereby a specific goal in mind will direct drivers' information processing. Future work is required to examine the effects of AR visual cue presentation methods on different types of driving tasks.

Participants' expectancy of road events should be further addressed in future work. We added one no-event trial and backing-up vehicle event per participant to reduce drivers' expectancy of pedestrians. Counterbalancing the presentation order helped evenly distribute learning and ordering effects across experimental conditions, but expectancy could still be an issue to further explore. Future work could introduce more no-event trials or various other events to address this problem. Additionally, future work could examine driver behavior and performance considering the limitations of current PCW systems, as a recent review [53] shows that even most advanced pedestrian detection algorithms exhibit high false-positive (false alarm) and false-negative (miss pedestrians) rates. Lastly, to address drivers' adaptation or acceptance of new AR HUD technology, a longitudinal observational study could be conducted to examine changes in drivers' confidence, behavior, and performance while using AR pedestrian collision warning systems over time in various situations.

6 Conclusion

This work furthers our collective understanding about human performance in AR by providing empirical evidence supporting that monoscopic displays can be as effective as advanced stereoscopic volumetric displays in AR braking tasks in far action space by taking advantage of conformal graphics. A practical implication of this research on HUD interface design is that visual warnings presented via conformal AR graphics (via either monoscopic or volumetric displays) can have considerably positive consequences on driver behavior and performance by guiding drivers' attention to relevant real-world objects. By understanding how to design AR driver interfaces that can effectively guide drivers' attention at critical moments (as opposed to divide attention or distract), we can begin to inform the design of automotive AR applications.

Despite some limitations, this work is one of the first empirical studies that address human performance with volumetric AR HUDs while driving on real roadways. We also examined a range of human depth perception cues affordable by current AR HUD technologies by using an in-vehicle prototype of volumetric HUD capable of rendering both monoscopic and volumetric views. The proposed approaches and ecologically valid methods presented can be leveraged and further developed by practitioners and future researchers for better understanding of human performance in AR driving and usability evaluation of automotive AR applications. In particular, the analysis of the effects of AR warning interfaces on driver behavioral patterns and braking performance could assist others in determining design and safety tradeoffs.

Finally, this work provides empirical evidence for the relative effectiveness of monoscopic displays in driving contexts. Substantial future work is required to fully understand the effectiveness of depth cues afforded by diverse display technologies for human performance in optical see-through AR; especially when moving through space.

ACKNOWLEDGMENTS

The experimental user study was conducted when the first author worked at Honda Research Institute USA as a research intern. A comprehensive data analysis was supported by Honda, completed at Virginia Tech and reported herein. The authors thank Nanxian Li, Ashish Tawari, and Kikuo Fujimura for insightful feedback on the user study, supporting the recruitment of participants, and data collection.

REFERENCES

[1] J. L. Gabbard, G. M. Fitch, and H. Kim, "Behind the Glass: Driver Challenges and Opportunities for AR Automotive Applications," *Proc. IEEE*, vol. 102, no. 2, pp. 124–136, 2014.

[2] J. P. Rolland and H. Fuchs, "Optical Versus Video See-Through Head-Mounted Displays in Medical Visualization," *Presence*, vol. 9, no. 3, pp. 287–309, Jun. 2000.

[3] J. Chen, W. Cranton, and M. Fihn, *Handbook of visual display technology*. Springer Berlin Heidelberg, 2012.

[4] S. Reichelt, R. Häussler, G. Fütterer, and N. Leister, "Depth cues in human visual perception and their realization in 3D displays," *Proc SPIE*, 2010.

[5] B. G. Blundell, *3D Displays and Spatial Interaction: From Perception to Technology Volume I Exploring the Science, Art, Evolution and Use of 3D Technologies*. Walker & Wood Limited, 2011.

[6] H. Kim, A. M. Anon, T. Misu, N. Li, A. Tawari, and K. Fujimura, "Look at Me: Augmented Reality Pedestrian Warning System Using an In-Vehicle Volumetric Head Up Display.," presented at the 21st International Conference on Intelligent User Interfaces, Sonoma, CA USA, 2016, pp. 294–298.

[7] S. J. Watt and K. J. MacKenzie, "3D Media and the Human Visual System," in *Emerging Technologies for 3D Video Creation, Coding, Transmission and Rendering*, vol. 23, no. 4, John Wiley & Sons, Ltd, 2013, pp. 349–376.

[8] R. E. Patterson, *Human factors of stereoscopic 3D displays*. Springer, 2015.

[9] J. E. Cutting and P. M. Vishton, "Perceiving layout and knowing distances: The integration, relative potency, and contextual use of different information about depth," in *Handbook of perception and cognition*, vol. 5, Academic Press, 1995, pp. 69–117.

[10] B. G. Blundell and A. J. Schwarz, *Volumetric Three-Dimensional Display Systems*. Wiley-VCH, 2000.

[11] B. G. Blundell, "Volumetric 3D Displays," in *Handbook of Visual Display Technology*, no. 116, Springer Berlin Heidelberg, 2012, pp. 1917–1931.

[12] D. Dunn, C. Tippets, K. Torell, P. Kellnhofer, K. Aksit, P. Didyk, K. Myszkowski, D. Luebke, and H. Fuchs, "Wide Field Of View Varifocal Near-Eye Display Using See-Through Deformable Membrane Mirrors.," *IEEE Trans Vis Comput Graph*, vol. 23, no. 4, pp. 1322–1331, Apr. 2017.

[13] C. Jerome and B. Witmer, "The Perception and Estimation of Egocentric Distance in Real and Augmented Reality Environments," *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 49, no. 26, pp. 2249–2252, Nov. 2016.

[14] J. W. McCandless, S. R. Ellis, and B. D. Adelstein, "Localization of a time-delayed, monocular virtual object superimposed on a real environment.," *Presence*, vol. 9, no. 1, pp. 15–24, Feb. 2000.

[15] J. P. Rolland, W. Gibson, and D. Ariely, "Towards Quantifying Depth and Size Perception in Virtual Environments," *Presence*, vol. 4, no. 1, pp. 24–49, 1995.

[16] S. R. Ellis and B. M. Menges, "Localization of Virtual Objects in the Near Visual Field," vol. 40, no. 3, pp. 415–431, Nov. 2016.

[17] J. E. Swan, A. Jones, E. Kolstad, M. A. Livingston, and H. S. Smallman, "Egocentric depth judgments in optical, see-through augmented reality.," *IEEE Trans Vis Comput Graph*, vol. 13, no. 3, pp. 429–442, May 2007.

[18] S. E. H. Kirkley Jr, "Augmented reality performance assessment battery (ARPAB): object recognition, distance estimation and size estimation using optical see-through head-worn displays," PhD Dissertation, Indiana University, 2003.

[19] M. Tonnis and G. Klinker, "Effective control of a car driver's attention for visual and acoustic guidance towards the direction of imminent dangers," presented at the 2006 IEEE/ACM International Symposium on Mixed and Augmented Reality, 2006, pp. 13–22.

[20] M. Smith, N. Doutcheva, J. L. Gabbard, and G. Burnett, "Optical see-through head up displays' effect on depth judgments of real world objects," presented at the 2015 IEEE Virtual Reality (VR), 2015, pp. 401–405.

[21] A. Hotta, T. Sasaki, and H. Okumura, *Depth perception effect of dynamic perspective method for monocular head-up display realizing augmented reality*. Proc. IDW '11, 2011.

[22] N. Broy, S. Höckh, A. Frederiksen, M. Gilowski, J. Eichhorn, F. Naser, H. Jung, J. Niemann, M. Schell, A. Schmid, and F. Alt, "Exploring Design Parameters for a 3D Head-Up Display." presented at the International Symposium on Pervasive Displays, 2014, pp. 38–43.

[23] Y. Takaki, Y. Urano, S. Kashiwada, H. Ando, and K. Nakamura, "Super multi-view windshield display for long-distance image information presentation.," *Opt Express*, vol. 19, no. 2, pp. 704–716, Jan. 2011.

[24] K. Bark, C. Tran, K. Fujimura, and V. Ng-Thow-Hing, "Personal Navi: Benefits of an Augmented Reality Navigational Aid Using a See-Thru 3D Volumetric HUD," presented at the 6th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, 2014, pp. 1–8.

[25] J. P. McIntire, P. R. Havig, and E. E. Geiselman, "Stereoscopic 3D displays and human performance: A comprehensive review," *Displays*, vol. 35, no. 1, pp. 18–26, Jan. 2014.

[26] E. Kruijff, J. E. Swan, and S. Feiner, "Perceptual issues in augmented reality revisited," presented at the 2010 9th IEEE International Symposium on Mixed and Augmented Reality (ISMAR), 2010, pp. 3–12.

[27] M. Lambooij, M. Fortuin, I. Heynderickx, and W. IJsselsteijn, "Visual Discomfort and Visual Fatigue of Stereoscopic Displays: A Review," *J. Imaging Sci. Technol.*, vol. 53, no. 3, p. 030201, 2009.

[28] D. Gupta, "An Empirical Study of the Effects of Context-Switch, Object Distance, and Focus Depth on Human Performance in Augmented Reality," Master's Thesis, Virginia Tech, 2004.

[29] Honda, "Honda to Showcase New Connected Car and Automated Driving Technologies at 2014 ITS World Congress in Detroit," [Online]. Available:http://world.honda.com/news/2014/4140905eng.html. [Accessed: 15-Aug-2017].

[30] A. Hussein, F. García, J. M. Armingol, and C. Olaverri-Monreal, "P2V and V2P communication for Pedestrian warning on the basis of Autonomous Vehicles," presented at the 2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC, 2016, pp. 2034–2039.

[31] C. Olaverri-Monreal and T. Jizba, "Human Factors in the Design of Human–Machine Interaction: An Overview Emphasizing V2X Communication," *IEEE Trans. Intell. Veh.*, vol. 1, no. 4, pp. 302–313, 2016.

[32] H. Kim, J. D. Isleib, and J. L. Gabbard, "Virtual Shadow: Making Cross Traffic Dynamics Visible through Augmented Reality Head Up Display," presented at the Human Factors and Ergonomics Society Annual Meeting, 2016, vol. 60, no. 1, pp. 2093–2097.

[33] H. Kim, J. D. Isleib, and J. L. Gabbard, "Casting shadows: Ecological interface design for augmented reality pedestrian collision warning," presented at the 2016 IEEE Virtual Reality (VR), 2016, pp. 205–206.

[34] G. M. Fitch, M. Blanco, J. F. Morgan, J. C. Rice, and A. E. Wharton, "Human performance evaluation of light vehicle brake assist systems," 2010.

[35] D. Fambro, R. Koppa, D. Picha, and K. Fitzpatrick, "Driver Braking Performance in Stopping Sight Distance Situations," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1701, pp. 9–16, Jan. 2007.

[36] R. J. Kiefer, D. J. LeBlanc, and C. A. Flannagan, "Developing an inverse time-to-collision crash alert timing approach based on drivers' last-second braking and steering judgments.," *Accid Anal Prev*, vol. 37, no. 2, pp. 295–303, Mar. 2005.

[37] G. Markkula, O. Benderius, K. Wolff, and M. Wahde, "A review of near-collision driver behavior models.," *Human factors*, vol. 54, no. 6, pp. 1117–1143, Dec. 2012.

[38] M. Langham and N. Moberly, "Pedestrian conspicuity research: a review," *Ergonomics*, vol. 46, no. 4, pp. 345–363, Nov. 2010.

[39] "Stopping Distance Test Procedure," SAE International, 2009.

[40] E. Kasneci, G. Kasneci, T. C. Kübler, and W. Rosenstiel, "Online Recognition of Fixations, Saccades, and Smooth Pursuits for Automated Analysis of Traffic Hazard Perception," in *Artificial Neural Networks*, vol. 4, no. 20, Cham: Springer, Cham, 2015, pp. 411–434.

[41] H. Strasburger, I. Rentschler, and M. Jüttner, "Peripheral vision and pattern recognition: a review.," *J Vis*, vol. 11, no. 5, pp. 13–13, Dec. 2011.

[42] SMI, "Begaze Version 3.4 Manual," *SensoMotoric Instruments*, 2014.

[43] F. Bretz, T. Hothorn, and P. Westfall, "Multiple comparisons using R," 2016.
[44] G. Cumming, "The New Statistics: Why and How," *Psychological Science*, vol. 25, no. 1, pp. 7–29, Jan. 2014.
[45] J. S. Lappin, A. L. Shelton, and J. J. Rieser, "Environmental context influences visually perceived distance.," *Percept Psychophys*, vol. 68, no. 4, pp. 571–581, May 2006.
[46] W. A. Simpson, "Optic flow and depth perception," *Spatial Vision*, vol. 7, no. 1, pp. 35–75, Jan. 1993.
[47] K. Nakayama and J. M. Loomis, "Optical Velocity Patterns, Velocity-Sensitive Neurons, and Space Perception: A Hypothesis," *Perception*, vol. 3, no. 1, pp. 63–80, Jun. 2016.
[48] M. T. Swanston and W. C. Gogel, "Perceived size and motion in depth from optical expansion.," *Percept Psychophys*, vol. 39, no. 5, pp. 309–326, May 1986.
[49] D. Regan and K. I. Beverley, "Binocular and monocular stimuli for motion in depth: changing-disparity and changing-size feed the same motion-in-depth stage.," *Vision Res.*, vol. 19, no. 12, pp. 1331–1342, 1979.
[50] N. Broy, S. Höckh, A. Frederiksen, M. Gilowski, J. Eichhorn, F. Naser, H. Jung, J. Niemann, M. Schell, A. Schmid, and F. Alt, "Exploring Design Parameters for a 3D Head-Up Display," presented at the International Symposium on Pervasive Displays, Copenhagen, Denmark, 2014, pp. 38–43.
[51] L. Lorenz, P. Kerschbaum, and J. Schumann, "Designing take over scenarios for automated driving," presented at the Human Factors and Ergonomics Society Annual Meeting, 2014, vol. 58, no. 1, pp. 1681–1685.
[52] M. R. Endsley, *Designing for Situation Awareness*. CRC press, 2106.
[53] S. Zhang, R. Benenson, M. Omran, J. Hosang, and B. Schiele, "How Far are We from Solving Pedestrian Detection?," presented at the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR, 2016, pp. 1259–1267.

Appendix A. Drivers mean responses (standard error of the mean, corrected for within-subject variability) in experimental conditions.

Exp. Conditions / Dependent Variables	PED @ near				PED @ far			
	Control (No warning)	Current (BRAKE)	Mono (Shadow)	Volume (Shadow)	Control (No warning)	Current (BRAKE)	Mono (Shadow)	Volume (Shadow)
travel dist. gaze-on warning (m)	N.A.	2.93 (0.62)	2.49 (0.67)	2.79 (0.67)	N.A.	2.86 (0.62)	3.81 (0.62)	2.51 (0.62)
travel dist. gaze-on pedestrian (m)	6.65 (0.77)	8.34 (0.77)	6.00 (0.83)	5.78 (0.83)	9.65 (0.77)	11.97 (0.77)	7.79 (0.77)	7.22 (0.77)
travel dist. foot-off gas (m)	7.71 (0.77)	4.74 (0.80)	4.53 (0.80)	4.15 (0.77)	11.37 (0.77)	4.38 (0.77)	6.91 (0.74)	5.95 (0.80)
travel dist. foot-on brake (m)	9.09 (0.74)	5.99 (0.76)	6.14 (0.76)	5.80 (0.74)	13.59 (0.74)	5.82 (0.74)	9.59 (0.71)	8.33 (0.76)
total stopping distance (m)	13.84 (0.92)	11.37 (0.87)	12.33 (0.92)	11.50 (0.89)	25.42 (0.90)	13.24 (0.90)	21.06 (0.87)	20.90 (0.90)
peak deceleration (g)	0.44 (0.02)	0.35 (0.02)	0.35 (0.02)	0.37 (0.02)	0.23 (0.02)	0.31 (0.02)	0.24 (0.02)	0.23 (0.02)

Appendix B. Main and interaction effects of the visual warnings and distance to the pedestrian on quantitative measures of driver responses. Results of ANOVA tests are reported with F statistics and significance of the effects. **p<.05, **p<.01, ***p<.001*

Dependent Variables	Warning	Distance	W X D
travel dist. gaze-on warning (m)	F(3, 43.61) = 29.47 ***	F(1, 42.73) = 6.22 *	F(3, 42.64) = 3.73 *
travel dist. gaze-on pedestrian (m)	F(3, 32.49) = 9.20 ***	F(1, 42.88) = 30.21 ***	F(3, 42.76) = 1.31
travel dist. foot-off gas (m)	F(3, 31.25) = 13.67 ***	F(1, 42.81) = 18.07 ***	F(3, 42.78) = 3.62 *
travel dist. foot-on brake (m)	F(3, 32.31) = 14.54 ***	F(1, 41.44) = 44.58 ***	F(3, 41.44) = 6.71 ***
total stopping distance (m)	F(3, 35.00) = 28.01 ***	F(1, 11.40) = 159.62 ***	F(3, 31.98) = 19.95 ***
peak deceleration (g)	F(3, 36.29) = 3.92*	F(1, 13.04) = 127.41***	F(3, 38.95) = 12.88 ***