Obstacle Detection and Alert System for Smartphone AR Users

HyeongYeop Kang
siamiz88@gmail.com
Kangwon National University
Samcheok, Korea

Geonsun Lee
g_sun@korea.ac.kr
Korea University
Seoul, Korea

JungHyun Han
jhan@korea.ac.kr
Korea University
Seoul, Korea

ABSTRACT

This paper presents an obstacle detection and alert system for the pedestrians who use smartphone AR applications. The system analyzes the input camera image to extract feature points and determines whether the feature points come from obstacles ahead in the path. With the obstacle detector, two experiments were made. The first investigated the obstacle alert interfaces, and the second investigated the orientation guide interfaces that instruct users to hold their smartphones with some angles/orientations appropriate to capture the environment. Then, the best interfaces identified from the experiments were integrated and tested to examine their usability and user experiences.

CCS CONCEPTS

• **Human-centered computing → Mixed / augmented reality**; *Empirical studies in HCI*.

KEYWORDS

augmented reality, pedestrian safety, alert interface

ACM Reference Format:

HyeongYeop Kang, Geonsun Lee, and JungHyun Han. 2019. Obstacle Detection and Alert System for Smartphone AR Users. In *25th ACM Symposium on Virtual Reality Software and Technology (VRST '19), November 12–15, 2019, Parramatta, NSW, Australia.* ACM, New York, NY, USA, 11 pages. https://doi.org/10.1145/3359996.3364256

1 INTRODUCTION

Thanks to the wide spread of smartphones equipped with good quality cameras, augmented reality (AR) has been penetrating deeper into our daily lives. However, incautious use of AR applications (such as AR games) may lead to dangerous accidents when the users walk. For example, the users may run into obstacles ahead in the path. A way to prevent such accidents is to detect the obstacles and alert the users.

In this paper, we first propose a simple but effective obstacle detection method. It adopts a visual-inertial odometry technique, which utilizes the smartphone's built-in camera and inertial measurement unit. In such a method, it is important for the users to hold the smartphone with appropriate orientations since the camera's field of view is determined by the orientations. For this, an appropriate orientation guide should be provided for the users. It is a user interface issue. Once obstacles are detected, alerts should be delivered to the users in the most effective manners. This leads to another user interface issue. We conducted two experiments - one for the orientation guide interfaces and the other for the obstacle alert interfaces. This paper reports the results. By integrating the results obtained from the experiments, we developed an obstacle detection and alert system for pedestrians. This paper also presents the user test with the system.

A few previous works were reported on the safety issue of AR users. To the best of our knowledge, however, our work is the first to design and test the obstacle-related interfaces for smartphone AR users. Both the interfaces and the prototype system were tested in the real-world environments. We believe that our study will benefit industry as well as academia.

VRST '19, November 12–15, 2019, Parramatta, NSW, Australia

ACM ISBN 978-1-4503-7001-1/19/11…$15.00
https://doi.org/10.1145/3359996.3364256

2 RELATED WORK

These days AR applications are generally used on portable devices such as smart glasses and smartphones. Users are hence allowed to walk while experiencing AR. However, pedestrians are at serious danger when their attention is distracted by the devices [Nasar and Troyer 2013; Pizzamiglio et al. 2017]. Even though the real-world environment is displayed on the device screen, people often fail to perceive the danger as their attention is focused on the contents. This is mainly due to the limit of concurrent multitask processing [Ophir et al. 2009].

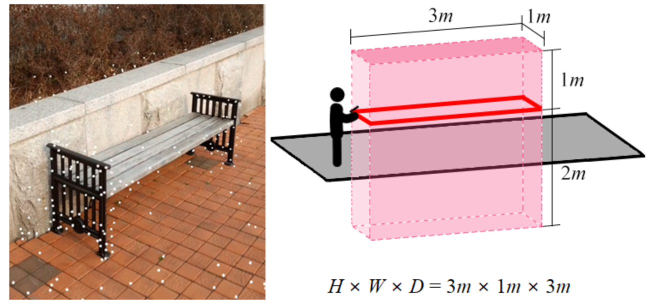
As a technical solution to ensure pedestrians' safety, researchers proposed systems that analyze the environment and alert the user if a situation is determined as dangerous. To examine an environment, various sensors such as ultrasound and infrared sensors are usually exploited. Shin et al. [Shin and Lim 2007] implemented a wearable system using ultrasound sensors to detect obstacles and determine the direction of avoidance. CrashAlert [Hincapié-Ramos and Irani 2013] is a system that attaches a depth camera to a mobile device and displays obstacles beyond the user's peripheral view. LookUp [Jain et al. 2015] used a shoe-mounted inertial sensor to detect users' transitions from a sidewalk into the road. The sensor measurements were relayed to a smartphone, and then the step pattern was extracted. UltraSee [Wen et al. 2015] and Infrasee [Liu et al. 2017] each mounted an ultrasonic sensor and an infrared sensor on the phone to detect changes in the ground such as stairs or metro station platforms. However, these sensors are not often included in portable devices and cost additional expenses.

To address the issue, there have been many attempts to ensure pedestrians' safety using portable devices without external sensors. WalkSafe [Wang et al. 2012] is an android application that detects vehicles approaching the user. For this purpose, it analyses a captured image from the rear camera of a phone based on the machine learning technique. The system detects vehicles only when users are on a phone call. SpareEye [Foerster et al. 2014] is also an android application that warns the user to avoid obstacles ahead. It finds image blobs and separates potential obstacles from the background, assuming that an area that touches the bottom line of the image is regarded as a part of the background. Zhou and Zhengjuan [Zhou 2015] analyzed data from sensors embedded in phones such as accelerometer and gyroscope to determine whether a user is walking or not. Based on this, the phone screen is locked when a user is walking. Tang et al. [Tang et al. 2016] detected tactile paver from a given camera image to distinguish the safe sidewalks from dangerous roads for the pedestrian. An alert was given to users when they stepped out of the safe sidewalk. Uchida et al. [Uchida et al. 2017] suggested an alert system that predicts a collision between pedestrian and vehicles using the wireless signal and sensor data of mobile phones.

With the increase in AR usage, several studies have focused on the safety of a pedestrian AR user. Jain et al. [Jain et al. 2014] used the inertial sensor of a mobile phone and GPS to determine whether pedestrians were at risk of a traffic accident. Using a convolutional neural network (CNN) model, Jung et al. [Jung et al. 2018] estimated the 3D position between a user and a vehicle. Then, they identified the most efficient method to visualize detected vehicles to AR user. However, the user test of the system was done with a virtual simulation on a desktop rather than a mobile environment. Gruenefeld et al. [Gruenefeld et al. 2018] developed a prototype of peripheral AR glasses to support pedestrians in critical traffic encounters and to evaluate three different light stimuli for shifting user's attention. Kanamori et al. [Kanamori et al. 2019, 2018] proposed obstacle avoidance methods for VR users. Their main idea was to superimpose the real-space information on the VR environment.

As an alert system for a pedestrian AR user, alert interfaces are also a crucial part of the system. Patterson [Patterson 1989] referred that the alert needs to be detectable, reliable, and has to lead users to an appropriate behavioral response. If an alert has not been detected by users, they cannot use it. If an alert is unreliable due to the high ratio of missed alert and false alert, it can become annoying and lead to more accidents [Baber 1995; Bliss and Acton 2003; Dingus et al. 1997]. Even highly detectable and reliable alerts can lead users to respond inappropriately. For example, Edworthy [Edworthy 1994] found that very loud warnings can cause startle reactions of users.

In the context of alert modalities, numerous studies have investigated which sensory modality is best for the alerts. Several studies conducted under simulated conditions with experienced aircraft pilots have found that auditory warnings produce faster response times than the visual warnings presented on panel indicators. [Reinecke 1976, 1981; Wheale 1981, 1983]. Kiefer et al. [Kiefer et al. 1999] conducted a user test with various alert modalities to find the best assistance method for drivers in avoiding a rear-end crash. Straughn et al. [Straughn et al. 2009] examined the change in reaction time when stimulus-response compatible warnings were given to drivers. It was found that tactile warnings elicit faster reaction time than auditory warnings. Braun et al. [Braun et al. 2018] investigated various warning modalities on a smartphone user facing an obstacle and found that auditive modalities lead significantly faster user's response than visual modalities.



(a) (b)

Figure 1: Features and region of interest (ROI) (a) Features are depicted as white dots. (b) ROI is depicted as a pink box and the walking path is in gray.

3 OBSTACLE DETECTION

Our system first detects obstacles within the field of view of a smartphone's camera. Section 3.1 presents how to build the obstacle detector, and Section 3.2 reports the results of a test made to validate its performances with real-world obstacles.

3.1 Design and Implementation

Visual-inertial Odometry. Our obstacle detector is built upon the Android platform of Samsung Galaxy S8+. The screen resolution is 2960×1440, and the rear camera's field of view is $77°$. The current implementation uses the visual-inertial odometry platform of Google ARCore [Google 2018]. Our obstacle detector takes two kinds of information it returns: a set of visually distinct 3D *feature points* (or simply *features*) located within the camera's field of view (Figure 1-(a)), and the 6DOF camera pose.

Region of Interest. Consider the box illustrated in Figure 1-(b). It represents a space located in front of the user's smartphone and is named the *region of interest* (ROI). The feature points that are out of the ROI are ignored. The ROI dimensions are $3m \times 1m \times 3m$. They were determined through the preliminary tests conducted indoor and outdoor with seven volunteers (one female, six males) with heights between $164cm$ to $180cm$ ($\mu = 172.5$, $\sigma = 5.78$).

- The ROI's height (H) is $3m$. It is divided into the upward height, $1m$, above the smartphone and the downward height, $2m$. When the upward height was too tall (e.g., $2m$), the corridor ceiling was often detected as an obstacle. When it was too short, our system often failed to detect the obstacles above the smartphone, such as tree branches. On the other hand, the downward height was set to $2m$ considering the smartphone's elevation from the indoor floor or outdoor ground. If the downward height is too short, the obstacles under the smartphone, such as downward staircase, may not be detected.
- The ROI's width (W) is $1m$. When W was too large, the corridor walls on the sides of a user were often detected as obstacles, falsely alerting the user who was safely walking down the corridor.
- The ROI's depth (D) is $3m$. It is the most important dimension as the obstacle detector's effectiveness highly depends on it. If D were too large, users would be unnecessarily alerted to the

far-away obstacles. On the other hand, if D were too small, users would be alerted too late, making them unable to avoid the obstacles.

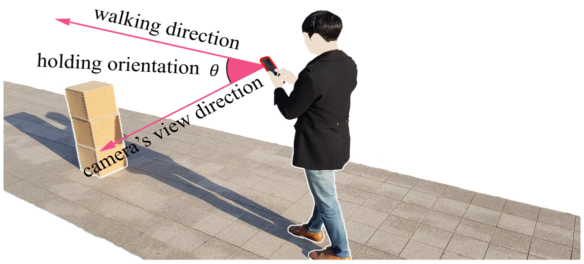


Figure 2: The holding orientation is defined as the angle between the user's walking direction and the camera's view direction.

Keyframes. Once the feature points within the ROI are selected, we determine whether to take the current frame as a *keyframe*. The obstacles are detected only from the keyframes. A keyframe should satisfy the following three criteria:

- The number of feature points within the ROI is at least 20.
- The *holding orientation*, defined in Figure 2, is in $[30°, 60°]$.
- The interval from the previous keyframe is at least $0.1s$.

These criteria were set also through the preliminary tests. The first criterion is to ensure the reliability of obstacle detection. An alert based on an insufficient number of features might be false. On the other hand, if the lower limit of the feature count is too high, dangerous obstacles may not be detected, thus skipping necessary alerts. The preliminary tests on the Android and Google ARCore platforms recommended to choose the minimum number of features from the range of $[10, 30]$ depending on the users' walking environments.

The second criterion is also to ensure the obstacle detector's reliability. When the holding orientation was greater than $60°$, the user's legs were often captured by the camera, occluding the features on the floor/ground. When it was smaller than $30°$, the small objects near the user's feet were hardly detectable.

The third criterion prevents our detector from repeating unnecessary computations. When the interval was too narrow, the computational load increased with little improvement on the obstacle detector's performances. When the interval was too wide, however, users would be alerted too late.

Feature Classification. Given a keyframe, random sample consensus (RANSAC) [Fischler and Bolles 1981] is applied to the feature points to identify a *reference plane* with two constraints: (1) it should be perpendicular to the vertical axis (y-axis) of the world space, and (2) it should be under the smartphone. In most cases, the reference plane corresponds to the floor/ground.

Then, the vertical distance between each feature point and the reference plane is computed. If it is greater than a threshold, α, the feature point is classified as an obstacle's. In the current implementation, α is $10cm$. If the number of the obstacle features is greater than another threshold, β, our system takes the front area as unsafe and issues an alert. In the current implementation, β is 3. Based on the subjects' feedbacks obtained in the preliminary tests, α and β were empirically determined. Many subjects complained about unnecessary alerts when α was smaller than $10cm$ since uneven road surfaces were often detected as obstacles. When β was 1 or 2, some noises were often misinterpreted as obstacles and unnecessary alerts were issued. This way, our obstacle detector identifies evident protrusions or depressions on the floor/ground.

3.2 Validation Test

An experiment was conducted to *validate* the obstacle detector in the real-world environments. The same seven volunteers were re-recruited from the preliminary tests. They were asked to walk with a speed of $1m/s$ along a straight path of $12m$, at the end of which an obstacle was placed. The subjects were instructed to continuously walk towards the obstacle even after an alert was issued. A safety guard near the obstacle prevented collisions. The subjects were advised to maintain the holding orientation of $45°$. All alerts were delivered as a beep sound. (Throughout this paper, we will use 'alert' and 'alarm' interchangeably.)

We selected eight kinds of obstacles, shown in Figure 3, which represent the obstacles frequently encountered in everyday life. They can be categorized into four classes:

- *Obstruction*: We used three instances of obstruction, named *1 cube*, *2 cubes* and *3 cubes*. They were realized by stacking $40cm \times 40cm \times 40cm$ sized cubes, which were made of soft material, and therefore they did not hurt the users even when collided.
- *Wall*: This was $8m$ high and made the $12m$-long path a dead end.
- *Staircases*: For both *upward staircase* and *downward staircase*, a step was $26cm$ high and $150cm$ wide, and the tread depth was $40cm$.
- *Curbs*: For both *upward curb* and *downward curb*, the height difference from the ground was $30cm$.

A volunteer tried each obstacle four times, and we collected 224 (= 7 volunteers $\times$ 8 obstacles $\times$ 4 repetitions) trials. They were analyzed to show that volunteers walked at the average speed of $1.04m/s$, the average holding orientation was $47.41°$, and our system ran at about 60fps.

Suppose that an obstacle enters the ROI and the alarm is issued. When it lasts longer than $0.3s$, we consider it as a valid alarm. Then, the user-obstacle distance at the time of alarm initiation is taken as the *first detection distance* (FDD). Obviously, the largest possible FDD is the ROI depth, $3m$, but the alarm was not always issued as soon as the obstacle entered the ROI. In the experiments, we measured the FDDs for all obstacle types. Table 1 lists their means and standard deviations. The FDDs for *downward staircase* and *downward curb* were about $2m$ while those for other obstacles were about $2.8m$. Such a difference is not unexpected because *downward staircase* and *downward curb* are hardly visible from the camera at far distances.

If the alarm at the FDD halted and there has been no alarm since then, we call it *missed alarm*. On the other hand, we have several kinds of *false alarm*. If an alarm is issued when no obstacle is in the ROI, it is false. If the first alarm issued after the obstacle's entering

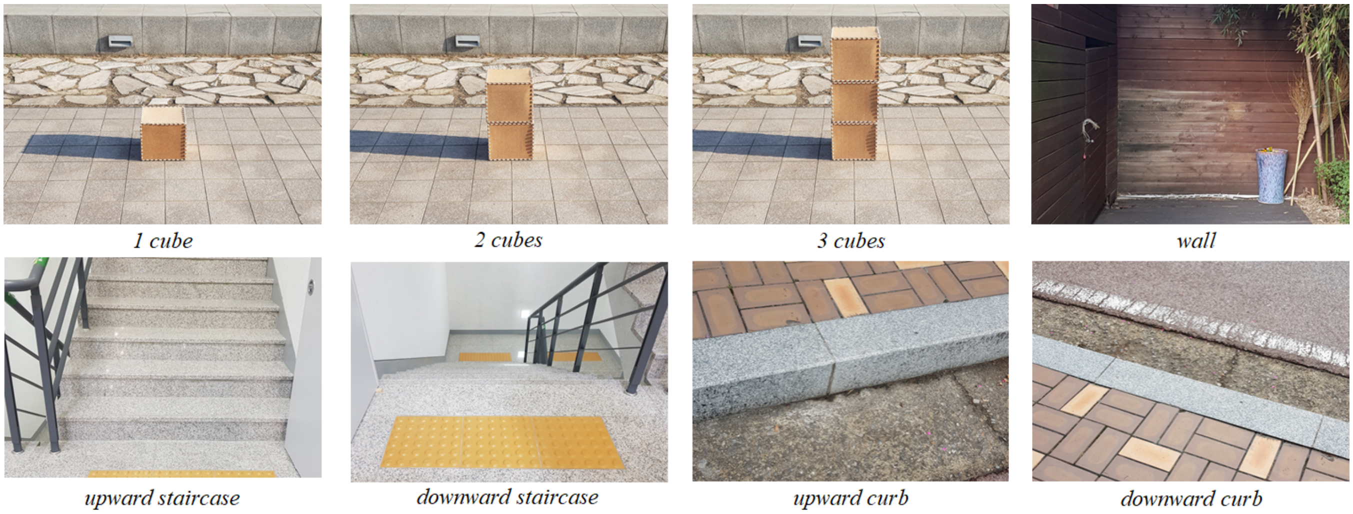


Figure 3: Eight obstacles used in the validation test.

the ROI lasts just for 0.2*s*, for example, it is also taken as a false alarm. We counted the missed and false alarms for 224 trials, but there were neither missed alarms nor false alarms at all. These perfect counts and the FDD statistics reported in Table 1 validated our obstacle detector.

Table 1: First detection distances (FDDs) for eight obstacles, where μ and σ denote the mean and standard deviation, respectively.

obstacle		*1 cube*	*2 cubes*	*3 cubes*	*wall*
FDD	μ	2.82m	2.78m	2.77m	2.80m
	σ	0.07	0.06	0.05	0.06
obstacle		*upward staircase*	*downward staircase*	*upward curb*	*downward curb*
FDD	μ	2.84m	1.93m	2.79m	2.08m
	σ	0.06	0.23	0.06	0.08

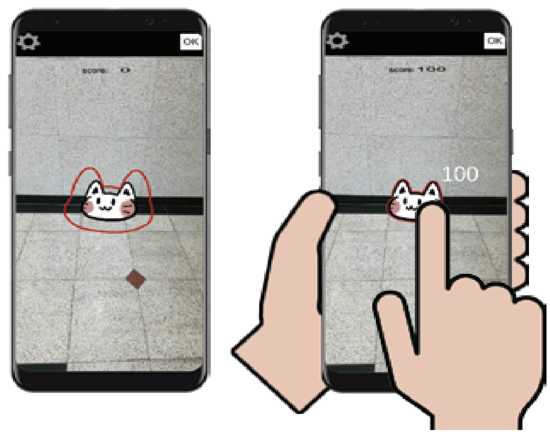
4 EXPERIMENT 1: OBSTACLE ALERT INTERFACES

We made two experiments (henceforth, called EXP1 and EXP2) with the obstacle detector. The goal of EXP1 was to find the interfaces that can most effectively alert the user while least disturbing the use of AR applications.

4.1 AR Game

It was reported that playing mobile games distracted the pedestrians' attention significantly [Haga et al. 2015]. In our experiments, the participants were instructed to play an AR game while walking.

We developed a prototype game shown in Figure 4, where the player touches a target character repeatedly appearing on the screen. A red outline appears simultaneously with the target. Initially, the outline's size is twice the target's. For 0.6*s*, the outline keeps shrinking to the target's boundary, and then for another 0.6*s*, it continues to shrink into a dot. If the target is not touched for the 1.2*s* duration, we call it 'miss.' As the interaction posture, the two-handed index finger input was used [Azenkot and Zhai 2012; Hoober 2013], as shown in Figure 4-(b). For each touch, a score is given. The player gets the highest score, 100 points, if the target is touched when the shrinking outline exactly matches the target's. The accumulated scores are displayed on the screen to make the user focus on the game, but the scores were not analysed after the experiments.



(a) (b)

Figure 4: AR game: (a) The target (cat face) and the initial outline. (b) When the target is touched, a score is displayed.

The screen is divided into 3×3 rectangular cells, and the targets appear in a counterbalanced order between the cells. Within a cell, the target's position is randomly determined. The target sizes vary randomly between small, medium and large. The game features two *difficulty levels*, which we name LV1 and LV2. In LV1, the interval between two consecutive targets' appearances is randomly determined in the range of [0.45*s*, 0.55*s*]. In LV2, it is in [0.3*s*, 0.37*s*].

For both EXP1 and EXP2 (to be presented in Section 5), a 90*s*-long version of the game was used. In contrast, 120*s*-long version was used for the application test (to be presented in Section 6).

4.2 Participants

Twenty subjects (8 females, 12 males) took part in EXP1. They aged between 21 to 30 ($\mu = 25.85$, $\sigma = 2.52$), and their heights ranged between $162cm$ and $184cm$ ($\mu = 172.09$, $\sigma = 6.35$). Ten subjects had experiences in mobile AR applications. We asked the subjects two questions: (Q1) how often they would use their smartphones while walking, and (Q2) how often they would play mobile games while walking. The answers in a 5-point Likert scale showed that $\mu = 4.35$ and $\sigma = 1.06$ for Q1 and $\mu = 3$ and $\sigma = 1.48$ for Q2. All subjects were right-handed, and each was paid 10 USD for their participation.

4.3 Method and Procedure

We used three *common-modality* alert methods and "no alert."

- *Visual*: An alert icon appeared on the screen's upper bar. It was a yellow triangle with an exclamation mark, which most users would be familiar with. See the right inset. In addition, the screen turned into translucent yellow.

- *Auditory*: We used an abstract sound, a siren, which is known to induce faster reaction from users compared to auditory icons [Braun et al. 2018].
- *Tactile*: A vibration of about 200Hz was delivered via the smartphone.
- *No alert*: No alerts were presented at all. (Henceforth, denoted simply as *No*.)

In EXP1, subjects were first asked to fill out a demographic survey and were provided with a brief introduction of the experiment. Prior to the main experiment, they were given 10 minutes to practice the AR game.

In the main experiment, the subjects played the game while walking a $90m$-long outdoor straight path. As was done in the previous works [Ng et al. 2014a,b], an experimenter performed a pacesetter's role to keep each subject's walking speed at about $1m/s$. We advised subjects to maintain the holding orientation of 45°.

A subject was provided with the four alert methods one by one. For each method, the subject went through two difficulty levels, LV1 and LV2. We used three *obstructions*, i.e., *1 cube*, *2 cubes* and *3 cubes* presented in Figure 3. For each trial, we selected one of three obstructions and placed it at random position on the path. In total, a subject made 24 trials (= 4 alert methods × 2 difficulty levels × 3 obstacles). The orders of the alert methods, difficulty levels, and obstacles were counterbalanced between subjects. As soon as the subjects found the obstruction ahead, they were instructed to promptly touch the OK button at the upper-right corner of the AR game screen, shown in Figure 4, and then walk around it.

After three obstacles were tested for a level, the subject was asked to fill out a subjective evaluation questionnaire shown in Table 2. In total, the questionnaire was answered eight times per subject. A trial took about $90s$ and the entire experiment consumed about an hour.

Table 2: The subjective evaluation questionnaire in EXP1. Every question was answered in a 5-point Likert scale (1: strongly disagree, 5: strongly agree).

category	question
immersion	Q1. I was immersed in the game.
distraction	Q2. I was distracted by the surrounding environment.
safety	Q3. I felt safe with the alert system.
effectiveness	Q4. The alert was effective for avoiding obstacles.
preference	Q5. I prefer the presented alert.

4.4 Result and Analysis

The subjects' average walking speed was $1.06m/s$ and the holding orientation was 41.34°. The analysis results will be presented in terms of detection time differences, miss ratios, and questionnaire responses.

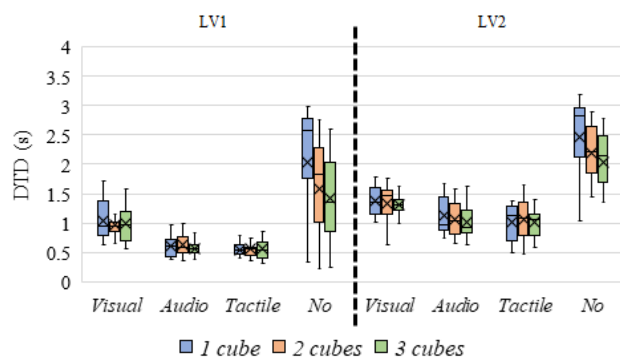


Figure 5: EXP1: Median, mean, interquartile ranges, and max/min values of DTDs (whiskers).

Detection Time Difference (DTD). Let t_s denote the time when our system detects obstacles and let t_u denote the time when the user touches the OK button. DTD is defined as $t_u - t_s$. No users detected obstacles before the system did, i.e., DTD was always positive.

Figure 5 illustrates the DTD statistics. Wilcoxon signed-rank test was conducted to find whether the DTDs were significantly different between the game levels for each alert method. The test revealed that there were significant differences for *Visual* ($Z = -5.253, p < 0.05$), *Audio* ($Z = -6.448, p < 0.05$), *Tactile* ($Z = -6.413, p < 0.05$), and *No* ($Z = -4.693, p < 0.05$).

We evaluated DTD differences among the four methods in LV1. A Friedman test, which is a non-parametric statistical test, was conducted. The test revealed that there was a significant difference ($X^2(3) = 98.629, p < 0.05$). Post hoc analysis using Wilcoxon signed-rank test was conducted with a Bonferroni correction applied, resulting in a significance level set at $p < 0.008$. The post hoc results are presented in Table 3.

We also investigated whether the obstacle sizes affected DTDs for each alert method. A Friedman test was conducted for each interface and found that DTD in *No* was significantly different

Table 3: EXP1: Post hoc test results of DTD differences between the alert methods in LV1.

test pairs (LV1)	*Visual -Audio*	*Visual -Tactile*	*Visual -No*	*Audio -Tactile*	*Audio -No*	*Tactile -No*
Z	-6.129	-6.652	-4.767	-1.981	-5.639	-5.761
p	< .008	< .008	< .008	> .008	< .008	< .008

between obstacle sizes ($X^2(2) = 30.700, p < 0.05$). Post hoc analysis found that there were significant DTD differences between all test pairs of obstacle sizes: *1 cube* and *2 cubes* ($Z = -3.584, p < 0.017$), *1 cube* and *3 cubes* ($Z = -3.173, p < 0.017$), and *2 cubes* and *3 cubes* ($Z = -3.061, p < 0.017$).

With LV2, we made the same analysis. We conducted a Friedman test to compare DTDs among the four alert methods. The test indicated that there was a significant difference ($X^2(3) = 102.005, p < 0.05$). The results of post hoc analysis are shown in Table 4.

Table 4: EXP1: Post hoc test results of DTD differences between the alert methods in LV2.

test pairs (LV2)	*Visual -Audio*	*Visual -Tactile*	*Visual -No*	*Audio -Tactile*	*Audio -No*	*Tactile -No*
Z	-4.546	-4.907	-5.856	-0.659	-6.368	-6.390
p	< .008	< .008	< .008	> .008	< .008	< .008

We conducted a Friedman test to find the correlation between obstacle sizes and DTDs within each alert method. The test result showed that DTDs in *No* were significantly different between obstacle sizes ($X^2(2) = 22.500, p < 0.05$). Post hoc analysis found that there were significant DTD differences between all pairs of obstacle sizes: *1 cube* and *2 cubes* ($Z = -2.688, p < 0.017$), *1 cube* and *3 cubes* ($Z = -3.360, p < 0.017$), and *2 cubes* and *3 cubes* ($Z = -3.024, p < 0.017$).

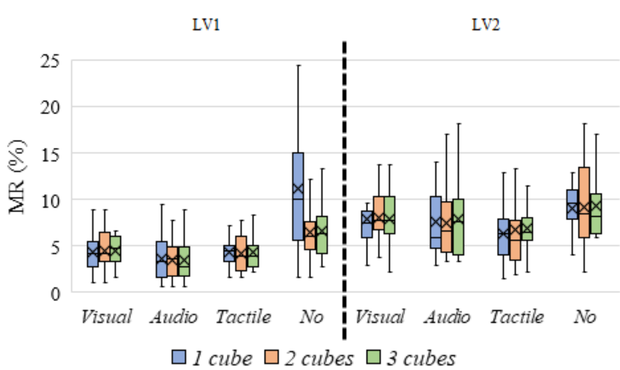


Figure 6: EXP1: Median, mean, interquartile ranges, and max/min values of MRs (whiskers).

Miss Ratio (MR). Recall that, if the target in the AR game is not touched for the 1.2*s* duration, we call it 'miss.' Figure 6 illustrates the MR statistics. We conducted Wilcoxon signed-rank test to find whether MRs were significantly different between game levels for each alert method. The test found that there were significant differences for *Visual* ($Z = -6.716, p < 0.05$), *Audio* ($Z = -6.738, p < 0.05$), *Tactile* ($Z = -6.513, p < 0.05$), and *No* ($Z = -5.465, p < 0.05$).

Similar to the analysis of DTDs, we first investigated the MR differences among four alert methods in LV1. A Friedman test showed that there was a significant difference ($X^2(3) = 72.005, p < 0.05$). Post hoc analysis results are shown in Table 5.

Table 5: EXP1: Post hoc test results of MR differences between the alert methods in LV1.

test pairs (LV1)	*Visual -Audio*	*Visual -Tactile*	*Visual -No*	*Audio -Tactile*	*Audio -No*	*Tactile -No*
Z	-3.030	-0.422	-5.285	-3.100	-6.504	-3.100
p	< .008	> .008	< .008	< .008	< .008	< .008

We investigated whether the obstacle sizes affected MRs in LV1. A Friedman test showed that MRs in *No* were significantly different between obstacle sizes ($X^2(2) = 6.872, p < 0.05$). Post hoc analysis found that there was a significant MR difference between the trials with *1 cube* and *2 cubes* ($Z = -2.841, p < 0.017$).

With LV2, we made the same analysis. We conducted a Friedman test to compare the MRs among the four alert methods. The test revealed that there was a significant difference ($X^2(3) = 34.074, p < 0.05$). The results of post hoc analysis are presented in Table 6.

Table 6: EXP1: Post hoc test results of MR differences between the alert methods in LV2.

test pairs (LV2)	*Visual -Audio*	*Visual -Tactile*	*Visual -No*	*Audio -Tactile*	*Audio -No*	*Tactile -No*
Z	-1.016	-3.166	-2.929	-1.961	-3.666	-5.352
p	> .008	< .008	< .008	> .008	< .008	< .008

We conducted a Friedman test to find the correlation between obstacle sizes and MRs within each alert method. The test result showed that there was no significant difference in MRs between obstacle sizes.

Questionnaire Response. Figure 7 illustrates the statistics of the questionnaire responses for four alert methods with two game levels. We tested whether the responses were significantly different between game levels. Wilcoxon signed-rank test found that there was a significant difference between the responses to Q5 in LV1 and LV2 ($Z = -2.104, p < 0.05$) whereas there were no significant differences in Q1, Q2, Q3 and Q4.

We investigated whether there were significant differences in responses between four alert methods for each game level. A Friedman test and post hoc results are presented in Table 7 and Table 8, respectively.

4.5 Discussion

For the trials in both LV1 and LV2, *No* had the largest DTD. This implies that the common-modality alerts contributed to improving

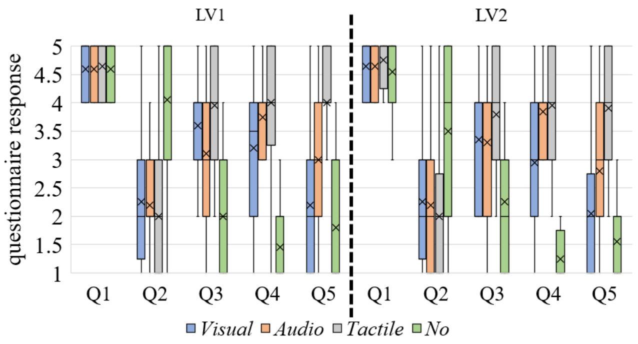


Figure 7: EXP1: Median, mean, interquartile ranges, and max/min values of the questionnaire responses (whiskers).

Table 7: EXP1: A Friedman test results for the subjective evaluation questionnaire.

	Q1		Q2		Q3		Q4		Q5	
	LV1	LV2	LV1	LV2	LV1	LV2	LV1	LV2	LV1	LV2
$X^2(3)$	0.2	1.2	21.4	14.1	30.4	15.9	37.8	35.1	24.7	26.0
p	>.05	>.05	<.05	<.05	<.05	<.05	<.05	<.05	<.05	<.05

Table 8: EXP1: Post hoc test results of the differences in subjective evaluation questionnaire responses between the alert methods.

test pairs		*Visual -Audio*	*Visual -Tactile*	*Visual -No*	*Audio -Tactile*	*Audio -No*	*Tactile -No*
Q2	Z	-0.188	-1.167	-3.014	-1.208	-3.309	-3.067
(LV1)	p	> .013	> .013	< .013	> .013	< .013	< .013
Q2	Z	-0.832	-1.508	-2.196	-0.586	-2.480	-2.388
(LV2)	p	> .013	> .013	> .013	> .013	< .013	< .013
Q3	Z	-2.066	-1.706	-3.474	-3.397	-2.652	-3.477
(LV1)	p	> .013	> .013	< .013	< .013	< .013	< .013
Q3	Z	-0.157	-2.264	-2.669	-2.214	-2.441	-3.093
(LV2)	p	> .013	> .013	< .013	> .013	< .013	< .013
Q4	Z	-1.713	-3.087	-3.431	-0.865	-3.965	-3.767
(LV1)	p	> .013	< .013	< .013	> .013	< .013	< .013
Q4	Z	-2.481	-3.078	-3.318	-0.090	-3.985	-3.677
(LV2)	p	< .013	< .013	< .013	> .013	< .013	< .013
Q5	Z	-2.231	-3.436	-0.992	-2.185	-3.225	-3.412
(LV1)	p	> .013	< .013	> .013	> .013	< .013	< .013
Q5	Z	-2.208	-3.470	-1.428	-2.193	-3.497	-3.583
(LV2)	p	> .013	< .013	> .013	> .013	< .013	< .013

users' reactions to obstacles. Furthermore, all common-modality alerts had lower MRs than *No*. Many subjects commented that they could concentrate more on the game thanks to the obstacle detectors. Only *No* was significantly different between obstacle sizes, and the DTD was larger when the obstacle was smaller (when the number of the stacked cubes is fewer). This is mainly because human eyes have difficulties in detecting small obstacles. These findings prove the usefulness of our obstacle alert system.

Among the three common-modality alert methods, *Visual* showed the largest DTD in both LV1 and LV2. However, there was no significant difference between *Auditory* and *Tactile*. Several subjects commented that *Visual* did not deliver as much sense of danger as *Auditory* and *Tactile*. The subjective questionnaire responses also show that users preferred *Auditory* and *Tactile* the most.

5 EXPERIMENT 2: ORIENTATION GUIDE INTERFACES

As discussed earlier, our system's obstacle detection performance is dependent on the camera's *holding orientation*. The goal of EXP2 was to find the interfaces that can most effectively guide users to hold the smartphones with desired orientations. For this, we designed a set of orientation guide interfaces and conducted user tests. The same subjects were re-recruited from EXP1, and each was paid 10 USD for their participation.

5.1 Method and Procedure

The desired range of holding orientations was set to $[40°, 50°]$. Once the orientation went out of this range, an instruction was provided. In EXP2, we used two visual guides, two auditory guides, a vibrational tactile guide, and "no guide." The first five methods belong to the *common-modality* interfaces.

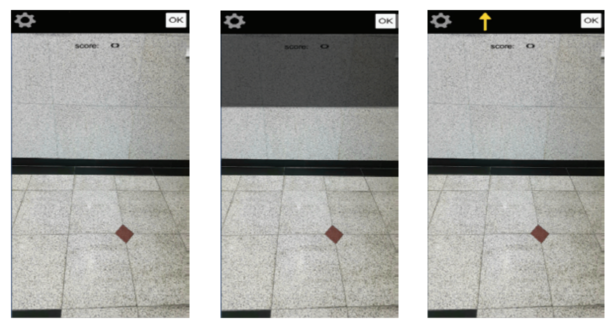


Figure 8: Brightness and Pointing Guides in EXP2: (a) Initial orientation in $[40°, 50°]$. (b) Brightness Guide. (c) Pointing Guide.

- *Brightness Guide* (*BG*): This interface guides users how to tilt the phones by changing the brightness of a subarea of the screen. If the angle goes over 50°, the upper part of the screen fades to black, i.e., the screen in Figure 8-(a) changes to that in Figure 8-(b). In order to restore the part, users have to tilt the phones upwards. By the same token, if the angle goes below 40°, the lower part fades to black and users have to tilt the phones downwards.
- *Pointing Guide* (*PG*): Arrows appear on the screen's upper bar to guide the users: ↑ and ↓ for upward and downward tilting, respectively. Observe that Figure 8-(b) and -(c) provide the same guide.
- *Abstract Auditory Guide* (*AAG*): A siren is issued when the holding orientation is out of range. This guide brings no directional information.
- *Specific Auditory Guide* (*SAG*): A recorded sound of "Up" or "Down" is played back according to the same metaphor as *BG* and *PG*.
- *Tactile Guide* (*TG*): A vibration of 200Hz is delivered when the holding orientation is out of range. This guide brings no directional information.

Table 9: The subjective evaluation questionnaire in EXP2.

category	question
immersion	Q1. I was immersed in the game.
distraction	Q2. I was distracted from the game due to the task of maintaining the desired orientation.
convenience	Q3. The guide was convenient for me.
effectiveness	Q4. The guide was effective in maintaining the orientation.
preference	Q5. I prefer the guiding method.

- *No Guide* (*NG*): At the beginning of the experiment, the users are instructed to maintain the desired holding orientations. However, no guides are given while walking.

As in EXP1, subjects walked down the $90m$-long path while playing the AR game at LV2. In EXP2, however, no obstacles were in the path. An experimenter performed a pacesetter's role to keep each subject's walking speed at about $1m/s$. A subject walked the path six times, each with a distinct orientation guide. The order of the orientation guides was counterbalanced between subjects. The holding orientations were measured every frame, and a subjective evaluation questionnaire was filled after each trial. The questions are listed in Table 9. The experiment took about 20 minutes.

5.2 Result and Analysis

In EXP2, we collected three kinds of data: (1) the average holding orientation, (2) the total amount of time, denoted as T_o, when the holding orientation was out of the desired range, and (3) the miss ratio (MR). Figure 9 illustrates their statistics.

A Friedman test found that there was no significant differences of the average holding orientations between the six guide methods ($X^2(5) = 8.400, p > 0.05$). In contrast, there were significant differences between the guide methods in terms of T_o ($X^2(5) = 24.814, p < 0.05$) and MR ($X^2(5) = 12.092, p < 0.05$). Post hoc analysis using Wilcoxon signed-rank test was conducted with a Bonferroni correction applied, resulting in a significance level set at $p < 0.003$. The post hoc test for T_o found that there were significant differences between *BG* and *NG* ($Z = -2.987, p < 0.003$), between *PG* and *NG* ($Z = -3.248, p < 0.003$) and between *SAG* and *NG* ($Z = -3.173, p < 0.003$). The post hoc test for MRs showed that there was a significant difference between *BG* and *NG* ($Z = -3.073, p < 0.003$).

Figure 10 illustrates the statistics of the questionnaire responses for the orientation guides. A Friedman test found that there were significant differences between the responses to the six methods in Q2 ($X^2(5) = 26.786, p < 0.05$), Q3 ($X^2(5) = 70.943, p < 0.05$), Q4 ($X^2(5) = 66.765, p < 0.05$), and Q5 ($X^2(5) = 58.670, p < 0.05$). As a post hoc test, we conducted Wilcoxon signed-rank test. The post hoc results are listed in Table 10.

5.3 Discussion

NG's T_o was significantly larger than *BG*'s, *PG*'s, and *SAG*'s. This implies that the orientation guides "with directional information" were effective in general. Several subjects complained that they could not figure out how to correct the holding orientation with *AAG* and *TG*.

Table 10: EXP2: Post hoc test results of the differences in subjective evaluation questionnaire responses between orientation guide methods. Only the test pairs with statistically significant differences are listed.

test pairs with significant differences					
	test pairs	Z		test pairs	Z
Q2	*BG-AAG*	−3.054	Q4	*BG-AAG*	−3.787
	PG-AAG	−3.100		*BG-SAG*	−3.471
	PG-TG	−3.002		*BG-TG*	−3.555
	AAG-SAG	−3.239		*BG-NG*	−3.364
	SAG-TG	−3.135		*PG-AAG*	−3.457
	test pairs	Z		*PG-SAG*	−3.114
Q3	*BG-AAG*	−3.294		*PG-TG*	−3.460
	BG-SAG	−3.450		*AAG-SAG*	−4.027
	BG-TG	−3.700		*SAG-TG*	−3.903
	BG-NG	−3.382		*SAG-NG*	−3.796
	PG-AAG	−3.783		test pairs	Z
	PG-SAG	−3.209	Q5	*BG-SAG*	−3.556
	PG-TG	−3.800		*BG-TG*	−3.166
	PG-NG	−3.299		*PG-SAG*	−3.863
	AAG-SAG	−3.988		*PG-TG*	−3.471
	SAG-TG	−3.994		*AAG-SAG*	−3.985
	SAG-NG	−3.978		*SAG-TG*	−4.010
				SAG-NG	−3.872

BG's MR was significantly higher than *NG*'s. Six subjects commented "It was hard to see the targets with *BG*. It was quite annoying."

With respect to the subjective evaluation questionnaire, *SAG* was significantly higher than the others in terms of convenience, effectiveness, and preference. On the other hand, *BG* and *PG* were significantly higher than *NG* and *AAG* in terms of convenience and effectiveness. However, *BG* had a significantly higher MR than *NG*.

One notable finding is that the interfaces "without directional information" (*AAG* and *TG*) had as bad scores as *NG* in terms of distraction, convenience and effectiveness. Several subjects commented that, with *AAG* and *TG*, they had a hard time figuring out which direction they should change their orientation to.

6 APPLICATION TEST

In both EXP1 and EXP2, we investigated visual, auditory and tactile modalities. In EXP1, *Auditory* (using an abstract sound, which is a siren) and *Tactile* were the best. In EXP2, *Pointing Guide* (*PG*; visual) and *Specific Auditory Guide* (*SAG*) were the best. There were no modality overlaps between the results of EXP1 and EXP2. This inspired us to combine the best interfaces.

Let us abbreviate *Auditory* and *Tactile* in EXP1 to *A* and *T*, respectively. With *PG* and *SAG* from EXP2, we defined four interface combinations: (1) *A*+*PG*, (2) *A*+*SAG*, (3) *T*+*PG*, and (4) *T*+*SAG*. Then, we developed a prototyping system by integrating these combinations of interfaces with our obstacle detector. We name the system **SafeAR**. (The name contains **AR** because the interfaces were tested with an AR game, but the functionalities of **SafeAR** can be extended to general mobile content through an appropriate user study.) This section presents a user test conducted to investigate the usability of each interface combination of **SafeAR**.

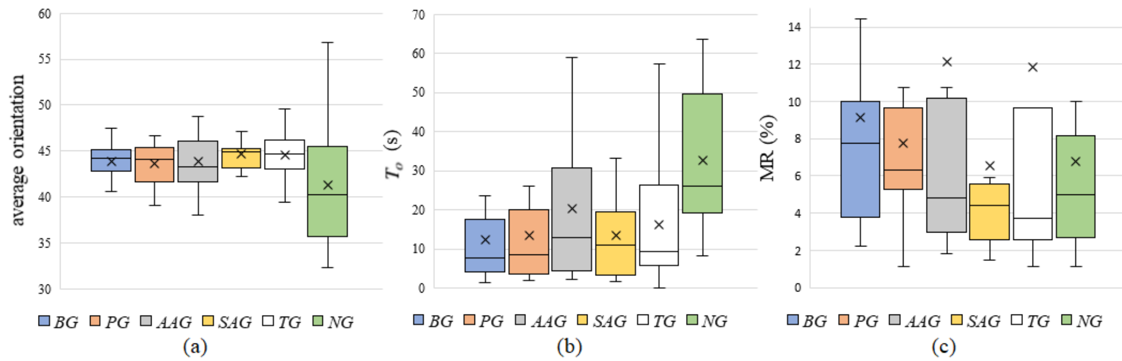


Figure 9: EXP2 statistics: (a) The average holding orientation. (b) T_o. (c) Miss ratio (MR).

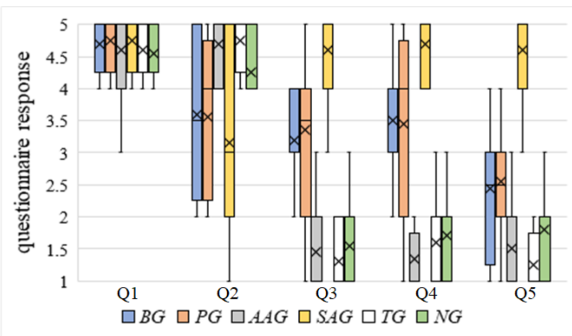


Figure 10: EXP2: Median, mean, interquartile ranges, and max/min values of the questionnaire responses (whiskers).

6.1 Participants

Thirty subjects (13 females) were newly recruited for test. They aged between 18 and 32 (μ = 24, σ = 3.17), and the heights ranged between 155cm and 185cm (μ = 168.56, σ = 7.26). Sixteen subjects had experiences with AR: Seven in AR headsets and fourteen in mobile AR applications. We asked the subjects two questions: (Q1) how often they would use their smartphones while walking, and (Q2) how often they would play mobile games while walking. The answers in a 5-point Likert scale showed that μ = 4.64 and σ = 0.48 for Q1 and μ = 2.92 and σ = 1.44 for Q2. All subjects were right-handed, and each was paid 10 USD for their participation.

6.2 Method and Procedure

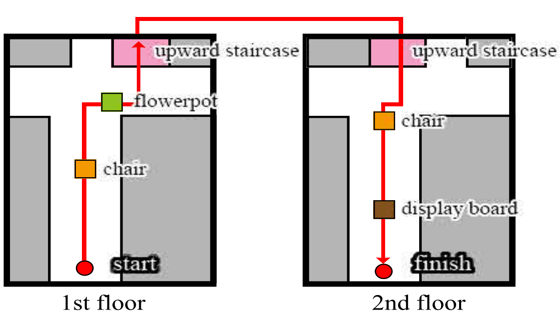


Figure 11: The 120m-long indoor path used for SafeAR test.

Table 11: The subjective evaluation questionnaire used in the application test.

category	question
immersion	Q1. I was immersed in the game.
distraction-angle	Q2. I was distracted from the game due to the task of maintaining the desired orientation.
distraction-warning	Q3. I was distracted by the surrounding environment.
distraction-overall	Q4. The system distracted me from the game
preference	Q5. I prefer the system.

SafeAR was tested in the 120m-long indoor path shown in Figure 11. The path had six obstacles: two swivel chairs, a flowerpot, two upward staircases, and a display board. First of all, subjects were informed about **SafeAR**'s functionalities and the test procedure. They were then asked to answer the *intention-to-use* question of "I will use an obstacle alert system when I play an AR game" in a 5-point Likert scale. For practice, the subjects were asked to walk the path shown in Figure 11 without holding smartphones in order to get familiar with the environment. The walking speed of 1m/s was recommended. They also practiced the AR game (in LV2) for ten minutes.

In the main test, the subjects walked the path while playing the AR game at the walking speed of 1m/s, together with an experimenter performing the roles of both pacesetter and safety guard. A subject performed a trial for a combination, resulting in four trials in total. The order of combinations was counterbalanced between subjects. The test took about 20 minutes. After each trial, subjects filled out the System Usability Scale (SUS) [Brooke et al. 1996] and a subjective evaluation questionnaire shown in Table 11. After completing the entire test, the subjects were asked to rank the combinations. Finally, they answered the *intention-to-use* question again.

6.3 Result and Analysis

Figure 12 illustrates the SUS analysis results. A Friedman test found that a significant difference of SUS scores exists between four combinations ($X^2(5)$ = 15.547, p < 0.05). Post hoc analysis using Wilcoxon signed-rank test was conducted with a Bonferroni correction applied, resulting in a significance level set at p < 0.008. There were significant differences between the SUS scores of T+SAG

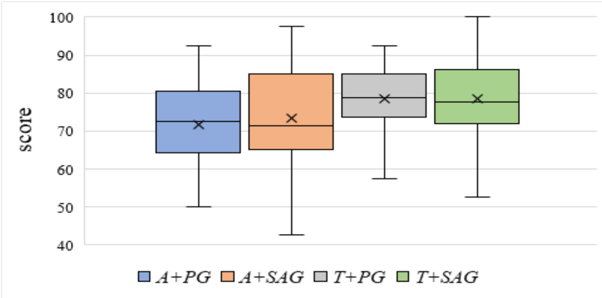


Figure 12: SUS analysis results.

and *A+PG* ($Z = -2.817, p < 0.008$) and between the SUS scores of *T+PG* and *A+PG* ($Z = -3.035, p < 0.008$).

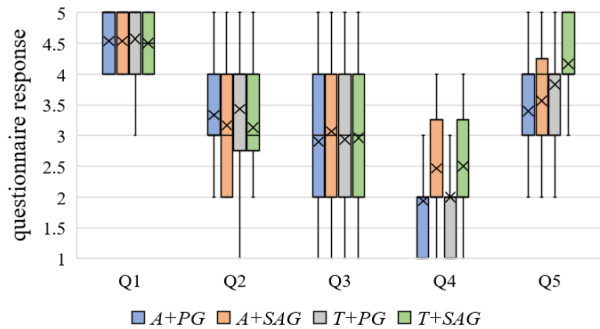


Figure 13: Application: Median, mean, interquartile ranges, and max/min values of the questionnaire responses (whiskers).

Figure 13 illustrates the statistics of the questionnaire responses. A Friedman test found that there was a significant difference between four combinations in Q5 ($X^2(5) = 9.081, p < 0.05$) whereas there were no significant differences between four combinations in Q1, Q2, Q3 and Q4. Post hoc analysis with Wilcoxon signed-rank test found that there was a significant difference between *T+SAG* and *A+PG* ($Z = -2.723, p < 0.008$).

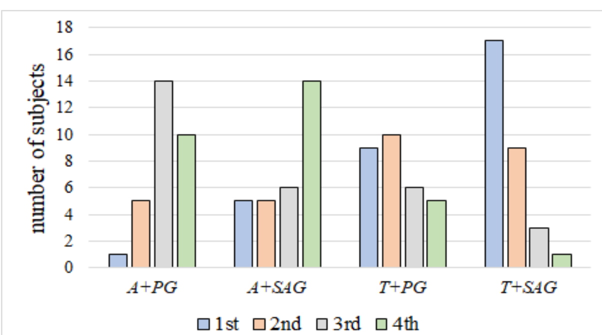


Figure 14: Preference (Q5) analysis results.

Figure 14 shows the users' preferences among the four combinations. The preferences were in the order of *T+SAG*, *T+PG*, *A+SAG*, and *A+PG*. We used Wilcoxon signed-rank test to investigate whether the users' responses to *intention-to-use* changed through the test. It was found that there was a significant difference before and after the test ($Z = -2.464, p < 0.05$).

6.4 Discussion

The responses to the subjective evaluation questionnaire revealed that four combinations had no significant differences in terms of immersion and distraction. The preferences increased in the order of *A+PG*, *A+SAG*, *T+PG* and *T+SAG*. *Tactile*(T) was evidently favored. However, four subjects commented "The tactile alarm was confused with notifications from text messages. A unique vibrational pattern should be developed." The majority of the subjects, 23 users, pointed out that *Auditory*(A) might not be an appropriate alert method for practical use. Some users commented "It is unlikely that I will be putting on headphones every time." and "I doubt that this will be useful in a noisy environment."

A significant difference was found between the users' responses to *intention-to-use* before and after the test, indicating that users became more willing to use the alert system once they experienced it. Some users gave valuable comments for future improvements. For example, twelve users pointed out that alerts should not be issued while they were going up or down the stairs. A user recommended to adopt a range of alert intensities which convey the different levels of danger.

7 CONCLUSION

We developed an obstacle detector and validated its performances using real-world obstacles. Through two experiments, we found that auditory and tactile interfaces were the most effective for issuing alerts whereas specific audio or arrow pointing was the most preferred for guiding the users to hold the smartphone with desired orientations. These interfaces were combined to **SafeAR**, with which we conducted a user test to examine the usability and user experiences. The users expressed their intentions to use **SafeAR**.

As argued in Section 6, the name, **SafeAR**, contains **AR** because the interfaces were tested with an AR game. However, the functionalities of **SafeAR** can be extended to non-AR applications such as general mobile content or even mobile VR games if appropriate user studies are designed and conducted in those applications. These efforts will establish the general method of securing safe walk while enjoying mobile content. We also envision that **SafeAR** can be ported to non-entertainment areas. For example, it can be adopted for a danger detection system for the visually impaired. We are planning to extend **SafeAR** along the directions.

Our obstacle detector uses the monocular camera and consequently reveals some limitations. For example, feature points are not extracted well in a textureless environment, and abrupt change in illumination may also cause an error in visual-inertial odometry process. In order to tackle such problems without attaching additional sensors to the smartphone, a deep learning-based approach [DeTone et al. 2018] will be worth investigating.

ACKNOWLEDGMENTS

This work was supported by the National Research Foundation of Korea (NRF) Grant funded by the Korea government (MSIT) (No. NRF-2017M3C4A7066316 and No. NRF-2016R1A2B3014319). The preliminary version of this work was presented as a poster at ISMAR 2019.

REFERENCES

Shiri Azenkot and Shumin Zhai. 2012. Touch behavior with different postures on soft smartphone keyboards. In *Proceedings of the 14th international conference on Human-computer interaction with mobile devices and services*. ACM, 251–260.

C Baber. 1995. Psychological aspects of conventional in-car warning devices. In *Human factors in alarm design*. Taylor & Francis, Inc., 193–205.

James P Bliss and Sarah A Acton. 2003. Alarm mistrust in automobiles: how collision alarm reliability affects driving. *Applied ergonomics* 34, 6 (2003), 499–509.

Melinda C Braun, Sandra Beuck, Matthias Wölfel, and Alexander Scheurer. 2018. Investigating Multimodal Warnings for Distracted Smartphone Users on the Move in Potentially Dangerous Situations. In *Transactions on Computational Science XXXII*. Springer, 1–14.

John Brooke et al. 1996. SUS-A quick and dirty usability scale. *Usability evaluation in industry* 189, 194 (1996), 4–7.

Daniel DeTone, Tomasz Malisiewicz, and Andrew Rabinovich. 2018. Superpoint: Self-supervised interest point detection and description. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*. 224–236.

Thomas A Dingus, Daniel V McGehee, Natarajan Manakkal, Steven K Jahns, Cher Carney, and Jonathan M Hankey. 1997. Human factors field evaluation of automotive headway maintenance/collision warning devices. *Human factors* 39, 2 (1997), 216–229.

Judy Edworthy. 1994. The design and implementation of non-verbal auditory warnings. *Applied ergonomics* 25, 4 (1994), 202–210.

Martin A Fischler and Robert C Bolles. 1981. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Commun. ACM* 24, 6 (1981), 381–395.

Klaus-Tycho Foerster, Alex Gross, Nino Hail, Jara Uitto, and Roger Wattenhofer. 2014. Spareeye: enhancing the safety of inattentionally blind smartphone users. In *Proceedings of the 13th international conference on mobile and ubiquitous multimedia*. ACM, 68–72.

Google. 2018. ARCore - Google Developers Documentation. https://developers.google.com/ar.

Uwe Gruenefeld, Tim Claudius Stratmann, Jinki Jung, Hyeopwoo Lee, Jeehye Choi, Abhilasha Nanda, and Wilko Heuten. 2018. Guiding Smombies: Augmenting Peripheral Vision with Low-Cost Glasses to Shift the Attention of Smartphone Users. In *2018 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*. IEEE, 127–131.

Shigeru Haga, Ayaka Sano, Yuri Sekine, Hideka Sato, Saki Yamaguchi, and Kosuke Masuda. 2015. Effects of using a smart phone on pedestrians' attention and walking. *Procedia Manufacturing* 3 (2015), 2574–2580.

Juan David Hincapié-Ramos and Pourang Irani. 2013. CrashAlert: enhancing peripheral alertness for eyes-busy mobile interaction while walking. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 3385–3388.

Steven Hoober. 2013. How do users really hold mobile devices. *Tillgänglig http://www.uxmatters. com/mt/archives/2013/02/how-do-users-really-hold-mobile-devices. php (Hämtad 2015-02-11)* (2013).

Shubham Jain, Carlo Borgiattino, Yanzhi Ren, Marco Gruteser, and Yingying Chen. 2014. On the limits of positioning-based pedestrian risk awareness. In *Proceedings of the 2014 workshop on Mobile augmented reality and robotic technology-based systems*. ACM, 23–28.

Shubham Jain, Carlo Borgiattino, Yanzhi Ren, Marco Gruteser, Yingying Chen, and Carla Fabiana Chiasserini. 2015. Lookup: Enabling pedestrian safety services via shoe sensing. In *Proceedings of the 13th Annual International Conference on Mobile Systems, Applications, and Services*. ACM, 257–271.

Jinki Jung, Hyeopwoo Lee, Jeehye Choi, Abhilasha Nanda, Uwe Gruenefeld, Tim Stratmann, and Wilko Heuten. 2018. Ensuring Safety in Augmented Reality from Trade-off Between Immersion and Situation Awareness. In *2018 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE, 70–79.

Kohei Kanamori, Nobuchika Sakata, Yoshinori Hijikata, Kensuke Harada, Kiyoshi Kiyokawa, et al. 2019. Walking Assist Method for VR Zombie. In *2019 12th Asia Pacific Workshop on Mixed and Augmented Reality (APMAR)*. IEEE, 1–7.

Kohei Kanamori, Nobuchika Sakata, Tomu Tominaga, Yoshinori Hijikata, Kensuke Harada, and Kiyoshi Kiyokawa. 2018. Obstacle avoidance method in real space for virtual reality immersion. In *2018 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE, 80–89.

RJ Kiefer, D LeBlanc, MD Palmer, J Salinger, Richard K Deering, Mike Shulman, et al. 1999. *Development and validation of functional definitions and evaluation procedures for collision warning/avoidance systems*. Technical Report. United States. Department of Transportation. National Highway Traffic Safety ….

Xuefeng Liu, Jiannong Cao, Jiaqi Wen, and Shaojie Tang. 2017. Infrasee: An unobtrusive alertness system for pedestrian mobile phone users. *IEEE Transactions on Mobile Computing* 16, 2 (2017), 394–407.

Jack L Nasar and Derek Troyer. 2013. Pedestrian injuries due to mobile phone use in public places. *Accident Analysis & Prevention* 57 (2013), 91–95.

Alexander Ng, Stephen A Brewster, and John H Williamson. 2014a. Investigating the effects of encumbrance on one-and two-handed interactions with mobile devices. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 1981–1990.

Alexander Ng, John H Williamson, and Stephen A Brewster. 2014b. Comparing evaluation methods for encumbrance and walking on interaction with touchscreen mobile devices. In *Proceedings of the 16th international conference on Human-computer interaction with mobile devices & services*. ACM, 23–32.

Eyal Ophir, Clifford Nass, and Anthony D Wagner. 2009. Cognitive control in media multitaskers. *Proceedings of the National Academy of Sciences* 106, 37 (2009), 15583–15587.

Roy D Patterson. 1989. Guidelines for the design of auditory warning sounds. *Proc. of the Institute of Acoustics* 11, 5 (1989), 17–25.

Sara Pizzamiglio, Usman Naeem, Shafiq Ur Réhman, Mhd Saeed Sharif, Hassan Abdalla, and Duncan L Turner. 2017. A mutlimodal approach to measure the distraction levels of pedestrians using mobile sensing. *Procedia computer science* 113 (2017), 89–96.

M Reinecke. 1976. Experimentelle erkenntnisse zur eignung von stimmwarnsystemen in flugzeugen [Experimental cognition for qualification of voice warning systems in aircraft], Report No. *DGLR PAPER* (1976), 76–211.

Michael Reinecke. 1981. Voice warning systems: Some experimental evidence concerning application. *7-E I 3 3 5 ADIOYG RUm E R AE E EAC ND EVE O MET mhhhhEEE* (1981).

Byeong-Seok Shin and Cheol-Su Lim. 2007. Obstacle detection and avoidance system for visually impaired people. In *International Workshop on Haptic and Audio Interaction Design*. Springer, 78–85.

Stacie M Straughn, Rob Gray, and Hong Z Tan. 2009. To go or not to go: Stimulus-response compatibility for tactile and auditory pedestrian collision warnings. *IEEE Transactions on Haptics* 2, 2 (2009), 111–117.

Maozhi Tang, Cam-Tu Nguyen, Xiaoliang Wang, and Sanglu Lu. 2016. An efficient walking safety service for distracted mobile users. In *2016 IEEE 13th International Conference on Mobile Ad Hoc and Sensor Systems (MASS)*. IEEE, 84–91.

Noriki Uchida, Shoma Takeuchi, Tomoyuki Ishida, and Yoshitaka Shibata. 2017. Mobile Traffic Accident Prevention System based on Chronological Changes of Wireless Signals and Sensors. *JoWUA* 8, 3 (2017), 57–66.

Tianyu Wang, Giuseppe Cardone, Antonio Corradi, Lorenzo Torresani, and Andrew T Campbell. 2012. WalkSafe: a pedestrian safety app for mobile phone users who walk and talk while crossing roads. In *Proceedings of the twelfth workshop on mobile computing systems & applications*. ACM, 5.

Jiaqi Wen, Jiannong Cao, and Xuefeng Liu. 2015. We help you watch your steps: Unobtrusive alertness system for pedestrian mobile phone users. In *2015 IEEE International Conference on Pervasive Computing and Communications (PerCom)*. IEEE, 105–113.

John L Wheale. 1981. The speed of response to synthesized voice messages. *British Journal of Audiology* 15, 3 (1981), 205–212.

John L Wheale. 1983. Evaluation of an experimental central warning system with a synthesized voice component. *Aviation, space, and environmental medicine* (1983).

Zhengjuan Zhou. 2015. Headsup: keeping pedestrian phone addicts from dangers using mobile phone sensors. *International Journal of Distributed Sensor Networks* 11, 5 (2015), 279846.