CrashAlert: Enhancing Peripheral Alertness for Eyes-Busy Mobile Interaction while Walking

Anonymous 1

Anonymous 2

ABSTRACT

Divided attention when people multi-task while walking (e.g. texting, web browsing) can lead to dangerous outcomes, such as colliding into obstacles or other people. To avoid such risks we introduce CrashAlert, a novel system that augments a mobile device with a depth camera to extend the user's peripheral awareness and to provide visual cues of the environment in a minimal footprint display on the device. CrashAlert aims at enhancing peripheral alertness with eyes-busy mobile interaction by giving distance and location information of obstacles in a users' path. We present the design and early feedback of CrashAlert. An initial study outside the lab environment showed that with CrashAlert users improve their handling of potential collisions without any loss in task performance. Participants reported an increased perception of safety and alertness. Our results suggest that low fidelity-visualizations (i.e. reduced ambient information) are easily understandable and glance-able, lowering the cost of accessing such information during mobile eyes-busy interaction.

ACM Classification: H5.2 [Information interfaces and presentation]: User Interfaces. - Graphical user interfaces.

Keywords: Eyes-busy interaction, obstacle avoidance, texting and walking, walking user interfaces.

INTRODUCTION

Mobile device users habitually multi-task (e.g. texting, browsing the web) while walking. Unfortunately, tasks that require a high degree of visual attention or *eyes-busy interaction* limit much of the user's peripheral vision. This practice has resulted in users tripping onto curbs, walking into traffic or deviating from their intended path [4,5,6,9]. In 2008, eyes-busy interaction while walking led to a twofold increase from the previous year, in emergency related accidents [6]. This has forced municipalities to consider safety policies that ban mobile device usage while walking [9]. Policy making aside, technological support for safer walking and multi-tasking is relatively unexplored.

Walking user interfaces [3] has emerged with the goal of studying and exploring methods to assist users with eyes-busy tasks while walking, aiding users in being more efficient and accurate with cognitively demanding tasks such as text-entry when mobile. These explorations include audio feedback [1], enlarged soft buttons [3], two-handed chorded keyboard input with a stylus [10], and adaptive methods to compensate for extraneous movement [2]. However, these systems focused on task efficiency rather than user safety.

Under review for:

CHI'13.

In this paper, we introduce *CrashAlert*, a system aimed at improving safety while on the move. CrashAlert captures and displays information beyond the user's peripheral view using a depth camera attached to a mobile device (Figure 1). CrashAlert scans the environment for obstacles on the user's path, and alerts users about the position and distance of these obstacles in a minimal footprint display on the mobile device.

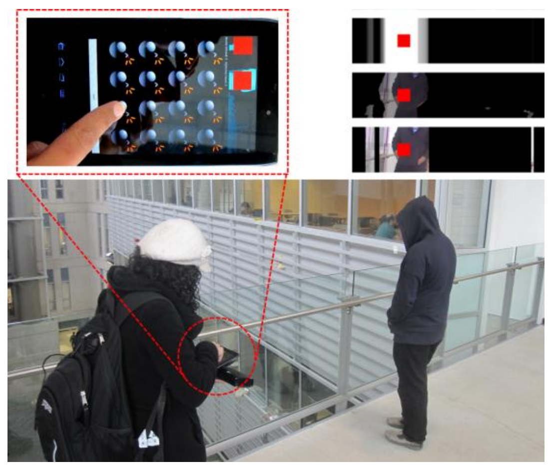

Figure 1 – CrashAlert senses environmental obstacles through a depth camera and informs the user of their positions. (bottom) User walking; (top-left) Peripheral objects displayed in an ambient display; (top-right) Various visual transformations available in CrashAlert with an alert in red.

We studied the impact of CrashAlert in a collision-prone environment in the 'wild'. Eight participants walked through a cafeteria during busy hours. We provoked potential collisions unknown to the participants by having an 'actor' moving head-on or cutting-off the user. Participants were able to easily interpret our visual alerts, felt safer using CrashAlert even after a short period of use, and their performance was not affected by the visual augmentations. We also found that with CrashAlert users were more proactive in avoiding collisions.

The main contribution of this paper is a demonstration that augmenting a mobile device with additional sensors, such as a depth sensing camera along with a minimal-footprint visual display of the sensor's data can assist users in eyes-busy interaction while walking. This contribution comes in three parts: (1) a mobile system, CrashAlert, augmented with a depth sensing camera, (2) a set of visual transformations aimed at minimizing screen space and optimizing relevant information, and (3) a study of CrashAlert showing improved handling of potential collisions with no negative impact on task performance.

RELATED WORK

Guidance systems for visually-impaired users

A number of commercially available (i.e. Sendero GPS, Trekker Breeze) and research-based systems [8] provide instructions to visually-impaired to assist with safe navigating and way-finding in common places. These systems rely on audio cues and recent results suggest that vibrotactile cues can also be mapped in a spatial manner to guide users in navigating through environments [11]. Prototypes closer in practice to ours process the mobile device's video feed from the environment and alert users about salient features, such as pedestrian crossings [12]. These systems rely on a fair amount of processing that may not be practical for real-time ambient feedback about out-of-periphery objects.

INFORMATIVE FIELD OBSERVATIONS

In order to inform our design we observed people walk while interacting with their mobiles in an indoor university cafeteria. We recorded the holding angle of the device, the number of hands used, the number of steps taken before lifting their heads (to detect on-comers and obstacles), the type of obstacles they most commonly avoided, patterns in walking speed and how many steps they took while typing. These observations led to the following considerations:

C1 – Collisions: People typically slowed down when reaching a crowded area or around obstacles on their route. In most cases the person would be the cause of a collision and not the other passersby. Therefore, the system should help users avoid collisions located on their path.

C2 – Fallback to vision: Navigating and dodging obstacles without lifting the head is a common practice. However, people fallback to vision in complex situations. Therefore, the system should prompt users so they lift their heads in order to sort out the situation.

CRASHALERT

Figure 2 presents CrashAlert, a system designed to provide users with ambient visualizations and visual alerts to help them avoid obstacles while on the move. CrashAlert uses both a depth and a normal camera attached to the device to capture a region in front of the user but outside their field-of-view. This slice of the image is processed and presented to the user on the mobile device.

CrashAlert has two main components: an ambient visual band and visual alerts for near-by objects. The ambient visualization seeks to convey a glance-able representation of the elements in front of the user, so that they gain an awareness of elements outside their field-of-view (C1). The visual alerts are generated from the depth image for objects that are as close as 2 meter away from the user. Their appearance (a bright red square in the position of collision, Fig 3) is easily noticeable by the user, causing them to raise their head to better cope with obstacles (C2).

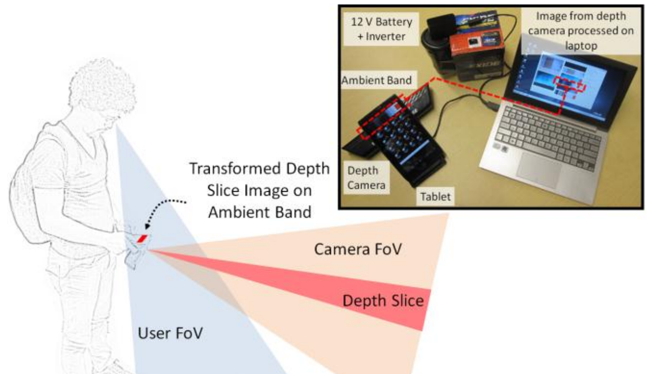

Figure 2 – (left) A depth sensing camera can assist in extending the user's limited peripheral view. Minimal processing on the slice of the depth image, or *Depth Slice*, can reveal salient information about the environment. Our implementation allows for an adjustable height and position for the Depth Slice. (inset) Hardware components in CrashAlert.

The ambient band displays a processed depth image and gives details in a minimalistic manner. Figure 3 presents three different ambient visualizations provided in our current implementation: a) color image, b) depth image, and c) masked image. The *color* image is simply a part of the color picture taken with the kinet color camera (see figure 3a). The *depth* image is obtained by applying a binary threshold to the depth capture from the Kinect, and calculating the average value for each column of pixels; this average is then assigned to all the pixels in the column (see figure 3b). The *masked* image uses the depth image as a mask on the image captured by the normal camera; this way it shows the full color version of the closest objects on a black background (see figure 3c).

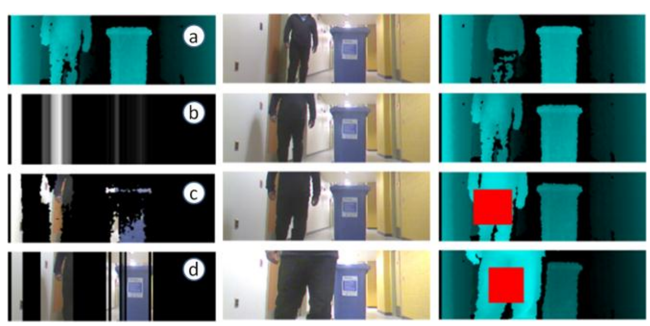

Figure 3 – (best seen in color) (top) Ambient band visualization based on given scene.

Implementation Details

Hardware. CrashAlert is a prototype built with an Acer A100 7'' tablet computer[1], a laptop computer, and a Mi-

[1] http://acer.us/ac/en/US/content/iconia-tab-a100

crosoft Kinect. The laptop is carried in a backpack together with a 12 volt battery to power up the Kinect in a mobile setting. The laptop receives images from the Kinect via USB, processes and transforms them, and sends them to the tablet via Bluetooth. The tablet receives images at a rate of approximately 10-11 frames per second. *Software.* The laptop runs a .NET C# application which interfaces with the Kinect through the OpenNI.net[2] stack, processes the images with OpenCV[3] through the Emgu.CV[4] wrapper, and communicates them over Bluetooth with the 32feet.Net network library[5]. The tablet application is an Android 2.3.3 application with normal Bluetooth permissions.

EVALUATION AND USER FEEDBACK

Our experiment was designed to observe the participants safety behaviors when using CrashAlert. We recruited university students who habitually text and walk. Eight volunteers (6 male, 2 female, mean age of 25.5 years) from various disciplines. All participants text while walking, although they agreed that such practice is dangerous, with an average of over a dozen collisions reported in the last year.

We designed a within-subjects experiment in which participants were exposed to four conditions: (1) None or No feedback (None), (2) Camera alone (CA), (3) Depth Image (DI) and (4) Image with Mask (IM). The camera was fixed at a 0º angle and participants were asked to keep the tablet horizontal. The depth slice covered the middle-low 2/5 of the depth and normal images.

Task and Procedure

We asked participants to play a whack-the-mole game, while walking through the university cafeteria. Each trip consisted of starting the walk at the near-by bookstore and looping around the entire food court. Each trial consisted of one loop around this path. Participants were asked to walk as normally as possible while playing the game. Their objective was to tap on as many moles as possible during their trajectory. Participants were asked to naturally avoid collisions with people and obstacles. We ensured that participants would face *at least* four collisions during each trial. This was achieved by asking an 'actor', unknown to the participant to provoke potential collisions. The 'actor' would cut the participants' path orthogonally, would stop right in front of them, would come toward them at a fast pace, or would walk with them but then immediately swerve in their lane. Participants also face obstacles from other people and objects in the cafeteria. The experimenter recorded participants' behavior during any potential collision.

We captured the user's total walking time, the number of moles they 'hit', as well as the number of times they performed a 'dodge/slow down', a 'heads-up', a 'full stop', and a 'bump'. Each experiment lasted roughly half an hour. Each condition was done twice, resulting in 8 participants x 4 conditions x 2 trials = 64 trials in total. We also interviewed the participants between trials and at the end of the experiment, and collected data (5-step likert scale) about their perceived safety, efficiency, alertness, walking speed, understandability and glance-ability of each of the conditions. If a mole was not hit within 2.5 secs we recorded an error and the mole was shown as being destroyed.

Participants wore the backpack containing the battery pack to which the Kinect was connected. They were first explained the task and were briefly explained the visualizations. We did not inform them of the planned collisions and asked them to behave naturally while trying to hit moles in the game as efficiently as possible. Participants walked through the cafeteria as per the assigned path. They completed 2 trips around the path or trials per condition.

Quantitative and Qualitative Results

Each trial lasted 130.5 seconds on average (stdev = 20.8) with an average number of 246.25 moles whacked per trial (stdev = 42.4), and an average error rate of 0.64% (moles missed, stdev = 1.37%). Moreover, participants whacked moles at an average rate of 1.91 moles/second (stdev = 0.04). There were no significant differences between techniques on number of moles hit or error rate. Therefore the wide spread in the number of moles per trial is explained by the wide spread in completion time. For the total 64 trials we registered 721 instances of possible collisions with an average of 180.25 instances per condition (stdev = 8.84).

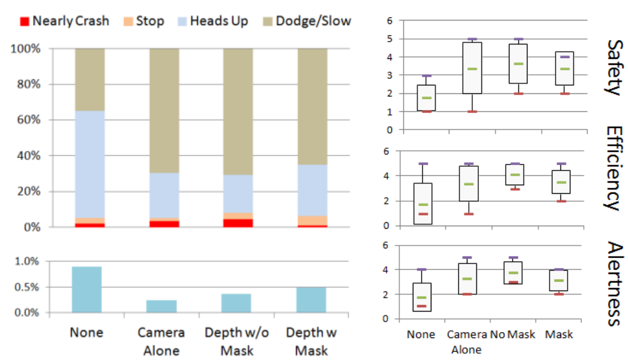

Figure 5 – (left-top) User navigation patterns broken down into % of total near crashes, full stops, heads-up and dodges. (left-bottom) No significant difference in error rates among techniques. (right) User rankings for their perception of Safety, Efficiency, and Alertness with each system.

We used the univariate ANOVA test and the Bonferroni correction for post-hoc pair-wise tests for our analysis. For the qualitative ratings we used the Friedman Chi-square test. Figure 5 shows an overview of the results. Figure 5-left-top shows the percentage distribution of collision handling maneuvers (dodge/slow-down, heads up, stop before crash, near crashes) for each condition. Results showed a main effect on the number of dodge/slow-downs ($F_{3,21}$ =

[2] http://www.openni.org/
[3] http://opencv.willowgarage.com/wiki/
[4] http://www.emgu.com
[5] http://inthehand.com/content/32feet.aspx

3.694, $p < 0.03$) and head ups ($F_{3,21} = 10.553$, $p < 0.01$). Post-hoc analysis showed difference only between the no-feedback condition and all the others, but not between them. These results show that participants avoided more obstacles by dodging rather than heads up as long as they used CrashAlert. Moreover, this better handling came at no cost in playing the game (no significant difference in error rate – figure 5.left-bottom).

The post-experiment survey results (see figure 5.right) showed that, when using CrashAlert, there was a trend for participants feeling more safety (Chi-squared(3) = 9, p = 0.029) and walking faster (Chi-squared(3) = 10.385, p = 0.016). There were no main effects on the other factors.

We interviewed each participant after each condition and had a longer debriefing after the whole experiment. We coded their answer into 19 tags and 4 topics as shown in table 1. In this section we present the results around three of the topics resulting from our coding: abstraction, navigation and alert. The fourth topic concerns technical improvements which we describe in our discussion.

In terms of the *abstraction* level of the visualizations, participants said that even though the full raw video and the depth mask images provide higher levels of detail they were nonetheless harder to read, requiring more attention and generating more stress when executing the task (even though we did not find any significant impact in performance); for example P8 said "*I have to check the [color] image much more and longer*". In contrast, the images in the depth condition were found easier to read "at a glance". Depth images communicate information that would have taken longer to read in the color or mask conditions; for example P7 indicated that "*[with the depth image] I can see the [thin] veranda which I couldn't in the normal video*". Moreover, participants reported depth images as falling into the background to the point that some users were not sure whether they were using them.

Participants talked about the different ways that CrashAlert enhances their current *navigational* senses (sound, peripheral view and knowledge of the environment) beyond simply noticing and alerting about obstacles and potential collisions: (1) allows participants to walk toward the dark regions shown on the ambient band and (2) react to the alert. Some participants found it useful to simply relax and follow the darker areas of the depth images, as they trusted that these areas would not have obstacles. In a different situation, when walking through a narrow and crowded corridor, a participant knew the person in front of her (shown with an alert due to proximity) was walking in the same direction and so she decided to follow the position of the alert to way-find through the crowd.

Finally, participants also noted that a system based only on alerts (no color, depth, or mask image) would already be an advantage compared to having nothing. Moreover, participants indicated the need for different *alert types*. One such type is alerts based on direction and speed; for example, participant 1 said "*I couldn't tell whether people where coming toward me or moving further away*". Another type of alert would be based on the type of object (static or moving object) and their related hazard estimation; for example P3 noted "*[I would like to see] a significance level indication of obstacles like how much danger if collision occurs*", and P5 said "*perhaps I could be alerted about different objects in different ways… moving people and static chairs require me to take action differently considering time and predictability*".

CONCLUSIONS AND FUTURE WORK

We presented CrashAlert, a mobile device augmented with a depth sensing camera that shows users out-of-periphery objects in their path while walking. CrashAlert shows salient information such as distance and position about object on the user's path. The information is displayed on a minimal footprint ambient band on top of the device's display. Study results show that users felt safer with our system and used it to help navigate around the environment. There was also no negative impact on performance with the ambient band, showing that even minimal environment information outside the user's periphery can assist with eyes-busy tasks in motion. We describe a number of features that can improve the effectiveness of CrashAlert and direction for further investigations.

REFERENCES

1. Brewster, S. (2002) Overcoming the lack of screen space on mobile computers. Personal and Ubiquitous Computing, 1 (2), 81-96.
2. Goel, M., Findlater, L., Wobbrock, J.O. (2012) WalkType: Using Accelerometer Data to Accomodate Situational Impairments in Mobile Touch Screen Text Entry, to appear CHI 2012.
3. Kane, S.K., Wobbrock, J.O. and Smith, I.E. (2008) Getting Off the Treadmill: Evaluating Walking User Interfaces for Mobile Devices in Public Spaces. Proc. MobileHCI'08, 109-118.
4. Nasar, J., Hecht, P., Wener, R. (2008) Mobile telephones, distracted attention, and pedestrian safety. Accident Analysis & Prevention 40, 69–75.
5. Neider, M.B., McCarley, J.S., Crowell, J.A., Kaczmarski, H., Kramer, A.F. (2010) Pedestrians, vehicles, and cell phones, Accident Analysis & Prevention, 42(2).
6. San Diego News (2012) How Risky Is Texting While Walking? http://www.10news.com/news/30530401/detail.html, last accessed: April 2012.
7. Stavrinos, D., Byington, K.W., Schwebel, D.C. (2011) Distracted walking: Cell phones increase injury risk for college pedestrians, Journal of Safety Research, 42(2).
8. Stewart, J., Bauman, S., Escobar, M., Hilden, J., Bihani, K., & Newman, M. W. (2008). Accessible contextual information for urban orientation. UbiComp '08.

9. TIME Magazine - Friedman, M. (2011) http://newsfeed.time.com/2011/01/27/can-lamakers-ban-texting-while-walking/, last accessed: April 2012.
10. Yatani, K. and Truong, K.N. 2009. An Evaluation of Stylus-based Text Entry Methods on Handheld Devices Studied in Different Mobility States. Pervasive and Mobile Computing, Vol. 5, No. 5, 496-506.
11. Yatani, K. and Truong, K.N. (2009) SemFeel: a user interface with semantic tactile feedback for mobile touch-screen devices. UIST 2009, 111-120.
12. Uddin, M.S. and Shioyama, T. (2005) Detection of pedestrian crossing and measurement of crossing length - an image-based navigational aid for blind people. Trans. on Intelligent Transportation Systems. IEEE, 2005.