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

ABSTRACT
Mobile device use while walking, or *eyes-busy mobile interaction*, is a leading cause of life-threatening pedestrian collisions. We introduce CrashAlert, a system that augments mobile devices with a depth camera, to provide distance and location visual cues of obstacles on the user's path. In a realistic environment outside the lab, CrashAlert users improve their handling of potential collisions, dodging and slowing down for simple ones while lifting their head in more complex situations. Qualitative results outline the value of extending users' peripheral alertness in eyes-busy mobile interaction through non-intrusive depth cues, as used in CrashAlert. We present the design features of our system and lessons learned from our evaluation.

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, web browsing) while walking. Researchers have introduced walking user interfaces (WUIs) [3] to improve mobile usage *efficiency* with tasks that require significant visual attention or *eyes-busy mobile interaction*. These interfaces include audio feedback [1], enlarged soft buttons [3], two-handed chorded keyboard input [10], and adaptive methods to compensate for extraneous movement [2].

WUIs primarily focus on task efficiency instead of user safety. Eyes-busy mobile interactions limit much of the user's peripheral vision, resulting in users tripping on curbs, walking into traffic or deviating from their intended path [4,5,9]. The year 2008 registered a twofold increase from the previous year in eyes-busy interaction-related accidents [7,6]. 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 at large unexplored.

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). The depth camera's field-of-view is orthogonal to that of the eyes-busy operator for increased peripheral awareness. Unlike navigation aids for the visually-impaired which rely on audio or vibro-tactile cues [11,12], CrashAlert displays a small slice of the depth camera's image as a minimal-footprint display on the mobile's screen. With an extended field-of-view, users can take simpler and early corrective actions upon noticing a potential collision. The display also alerts users of obstacles immediately in front of the user through a red alert, prompting users to immediately stop or lift their heads.

Figure 1 – (left) CrashAlert senses obstacles through a depth camera and informs the user of their positions using various visual transformations. (right) CrashAlert uses limited space on the display and shows nearby obstacles with a red alert.

We studied CrashAlert's impact in a collision-prone environment 'in the wild'. Eight participants walked through a cafeteria during busy hours while playing a game on a mobile device. Our participants easily interpreted our minimal-footprint display and used it to dodge or slow down for obstacles and raised their head in more collision-prone situations. These results were obtained at no cost to game performance. We also observed unexpected usage strategies such as (a) 'sheltering behind' a pedestrian walking in the same direction, as identified by the red alert, and (b) walking toward the 'dark' visual bands. Finally, participants reported that they felt safer using CrashAlert.

To the best of our knowledge CrashAlert is one of the first explorations of a safety-aware WUI. Our contribution is threefold: (1) a prototype implementation of CrashAlert, a system designed for safer eyes-busy interaction, (2) a set of visualizations aimed at minimizing screen real-estate and optimizing information about obstacles outside the user's field-of-view, and (3) a study of CrashAlert showing improved handling of potential collisions and an increased perception of safety, without loss of task performance.

Under review for:

CHI'13.

INFORMATIVE FIELD OBSERVATIONS

CrashAlert's design emerged from informal observations of pedestrians walking and interacting with their mobiles in a university cafeteria. We noted the holding angle of the device, the number of hands used, the number of steps taken before users lift their heads (to detect on-comers and obstacles), the type of obstacles commonly avoided, patterns in walking speed and how many steps users took while typing.

We noticed that when walking, people handle potential collisions with varying degrees of safety 'cost': from slowing down to dodging obstacles, then lifting their heads and/or ultimately coming to a full stop to avoid a crash. Users rely on their peripheral awareness to notice obstacles early on and to take simpler corrective actions (slowdown/dodge). As their walking continues, the obstacle is reevaluated and, if needed, further corrective actions are taken (heads-up). Limited peripheral vision means that obstacles are noticed later on, restricting the suitable corrective actions to higher cost ones (full stop or a crash). These observations led to the following design requirements (R). A WUI supporting safer walking should therefore prompt users to take simpler corrective actions early on by **encouraging dodges (R1)**, and **alerting on imminent collisions (R2)**.

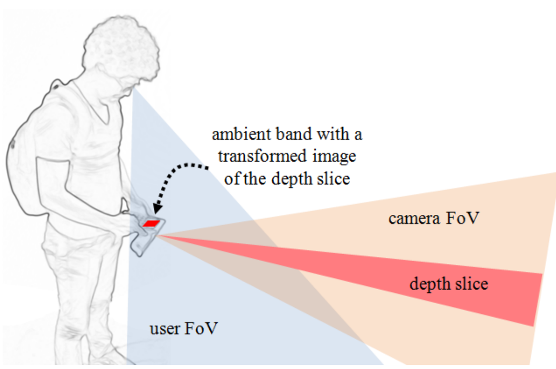

Figure 2 – A visualization based on depth-camera information extends the users' limited peripheral alertness.

CRASHALERT DESIGN

We designed CrashAlert (Figure 2) to let users act safely in eyes-busy mobile interaction. CrashAlert has two main components: an ambient visual band and visual alerts for near-by objects. Our system uses both a depth and a regular camera to capture the region in front of the user but outside their eyes-busy field-of-view (FoV). We extract only a small slice of the camera's image and process this to present obstacle positions and distances on the small footprint ambient band. The band conveys a glance-able representation of the elements in front of the user and outside their FoV (R1). Visual alerts are generated from the depth image for objects that are 2 meters away or less from the user. Their appearance (a bright red square in the position of collision, Fig 3) is quite salient, prompting the user to raise her head to better cope with obstacles (R2). We explicitly excluded other feedback modalities such as tactile or auditory to isolate the effects of our visual approach on safety.

We generated different visualizations for the ambient band through a design workshop with eight participants who interacted with 11 different visualizations. Figure 3 presents the three most preferred ambient bands: a) color image, b) depth image where closer objects are brighter (black at >5m), and c) the color image masked using the depth data. The *color* image is a slice of the picture taken with the color camera (figure 3a). The *depth* image is obtained by applying a binary threshold to the depth capture for a fixed distance (5 meters) and assigning the max value of each column to all of its pixels (figure 3b). The *masked* image uses the depth image (from figure 3b) as a mask on the color image; this way it shows the full color version of the closest objects on a black background (figure 3c). All bands presented a red alert when the obstacle was <2 m away.

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

Implementation Details

The CrashAlert prototype operates on an Acer A100 7'' tablet computer, a laptop computer, and a Microsoft Kinect (Figure 4). The laptop is carried in a backpack together with a 12 volt battery to power 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 approximately 10-11 frames per second. The application is written in C#.NET. It interfaces with the Kinect, processes the images with OpenCV, and communicates them over Bluetooth. The tablet software is an Android 2.3.3 application.

Figure 4 - CrashAlert prototype: Acer A100 + Kinect.

EVALUATION AND USER FEEDBACK

We conducted our experiment to observe participants' safety behaviors using CrashAlert. We recruited eight university students, from various disciplines, who habitually text and walk (6 male, 2 female, mean age of 25.5 years). All participants text while walking, but agreed that such practice is dangerous. On average, our participants reported having a dozen collisions over the last year. We designed a within-subjects experiment in which participants were exposed to four conditions: (1) No feedback (None), (2) Camera Alone (CA), (3) Depth Image (DI) and (4) Image with Mask (IM). Conditions were counter-balanced with an incomplete Latin-square design. The camera was fixed at a 0º angle (Figure 4) and participants were asked to hold the tablet in a natural way. The depth slice covered the middle-low 2/5 of the camera image.

Task and Procedure

We asked participants to play a whack-the-mole game while walking through the university cafeteria. Each trip (or trial) consisted of starting the walk at the near-by bookstore and looping around the entire food court (180 meters). 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 do one of the following: cut the participants' path orthogonally, would stop right in front of them, would come toward them at a fast pace, or would walk beside them but then immediately swerve in their lane. None of our participants suspected the presence of the 'actor'. Participants also faced 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,' or a 'crash'. Each experiment lasted roughly 30 minutes. Each condition was done twice, resulting in 8 participants×4 conditions×2 trials = 64 trials in total. We also interviewed the participants between trials and had a longer debriefing at the end of the experiment to collect data (5-step Likert-like scale) about their perceived safety, efficiency, alertness, walking speed, understandability and glance-ability of each condition. If a mole was not hit within 2.5s we recorded an error and the mole was shown as being destroyed.

Participants wore a backpack containing the battery pack to which the Kinect was connected. We first explained the task and 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.

Results and Discussion

Each trial lasted 130.5 seconds on average (*sd* = 20.8) with an average number of 246.25 moles whacked per trial (*sd* = 42.4), and an average error rate of 0.64% (moles missed, *sd* = 1.37%). Participants whacked moles at an average rate of 1.91 moles/second (*sd* = 0.04). There were no significant differences between conditions on the number of moles hit, error rate or completion time. For the total 64 trials, we registered 721 instances of possible collisions with an average of 11.26 (*sd* = 2.96) per trial (only 4 per trial were caused by the actor) and 180.25 (*sd* = 8.84) per condition.

We used the univariate ANOVA test and the Bonferroni correction for post-hoc pair-wise tests for our analysis. 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. The results showed a main effect of feedback style on the number of dodge/slow-downs ($F_{3,21}$ = 3.694, p < 0.03) and head ups ($F_{3,21}$ = 10.553, p < 0.01). Post-hoc analysis showed differences only between the no-feedback (None) condition and all the others, but not between the various visualizations. These results show that with CrashAlert participants avoided more obstacles by dodging and slowing down, rather than by heads up. Moreover, this better handling came at no cost in playing the game (no significant difference in error rate – figure 5-left-bottom or completion time $F_{3,21}$ = 0.7, p = 0.4). These results show that CrashAlert induced simpler corrective actions (i.e. dodging and slowing down) to avert possible collisions, providing users additional time and space for other more complex corrections (i.e. heads-up and full stops), and thus leading to safer walking.

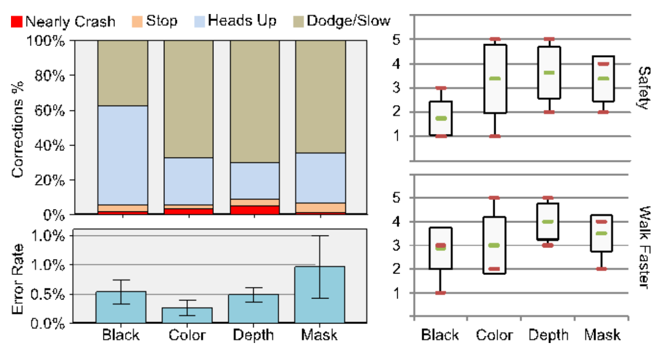

Figure 5 – (left-top) User corrective actions broken down into % of the total actions taken. (left-bottom) No significant difference in error rates among techniques. (right) User rankings for their perception of Safety and Walking Speed.

For subjective ratings we used the Friedman χ^2 test. Results (figure 5-right) showed that with CrashAlert, participants felt safer ($\chi^2(3)$ = 9, p = 0.029) and had a perception of walking faster ($\chi^2(3)$ = 10.385, p = 0.016). There were no main effects on the other factors.

We coded their answers (19 tags) into 3 topics: *abstraction*, *navigation*, *alerts.* In terms of the *abstraction* level, participants said that even though the color and the masked images provide higher levels of detail, they were 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 depth images were found easier to read "at a glance"; for example P7 indicated that "*[with the depth image] I can see the [thin] veranda which I couldn't in the color image*". Moreover, participants reported depth images as falling into the background to the point where some were convinced they had used them unconsciously.

Participants talked about the different ways that CrashAlert enhances their *navigational* senses (sound, peripheral view and knowledge of the environment) beyond simply alerting about obstacles and potential collisions, by: (1) allowing participants to walk within the dark regions shown on the ambient band, and (2) by interpreting the alert in unforeseen ways. 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 noted that a system based only on depth *alerts* (just the red box with no color, masked, or depth images) would be a marked advantage over current systems. Moreover, participants indicated the need for different *alert types*. One such type are 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 significant 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*".

Lessons Learned

We summarize three key benefits of CrashAlert:

- Depth and color images orthogonal to the user's FoV can facilitate safe navigation (dodging, slowing down and heads-up if necessary);
- Only a slice of the camera's image is needed to observe a benefit in extending users' peripheral alertness;
- Visual alerts based on depth information can support safer walking when interacting with a mobile device.

Limitations and Future Work

This initial exploration was limited by a low image rate (10-11 fps), a bulky hardware set-up, and naïve detection of obstacles (distance-based). However limited, our system demonstrated the value of considering safety in WUIs. Future work should investigate alternative visualizations, varying alert styles, such as a growing or shrinking boxes based on distance and speed, other feedback modalities, impact on complex tasks, dynamic selection of the image slice, scene analysis and object recognition (type, speed).

CONCLUSIONS

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 potential obstacles. The information is displayed on a minimal footprint ambient band on top of the device's display. Study results show that users took simpler corrective actions early on in their path upon noticing an obstacle, felt safer with our system and use it in unexpected ways to help navigate around the environment. This improvement came with no negative impact on performance, showing that even minimal environment information outside the user's periphery can provide for safer usage of mobiles while walking.

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 Accommodate Situational Impairments in Mobile Touch Screen Text Entry. 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.
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: September 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 (2011), Friedman, M. http://newsfeed.time.com/2011/01/27/can-lamakers-ban-texting-while-walking/, last accessed: September 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.