EyeRing: An Eye on a Finger

Suranga Nanayakkara[1, 2]
suranga@sutd.edu.sg

Roy Shilkrot[1]
roys@media.mit.edu

Patricia Maes[1]
pattie@media.mit.edu

[1] MIT Media Laboratory
75 Amherst Street
Cambridge, MA, 02142

[2] Singapore University of Technology and Design
20 Dover Drive
Singapore, 138682

CHI'12, May 5–10, 2012, Austin, Texas, USA.
ACM 978-1-4503-1016-1/12/05.

Abstract
Finger-worn devices are a greatly underutilized form of interaction with the surrounding world. By putting a camera on a finger we show that many visual analysis applications, for visually impaired people as well as the sighted, prove seamless and easy. We present EyeRing, a ring mounted camera, to enable applications such as identifying currency and navigating, as well as helping sighted people to tour an unknown city or intuitively translate signage. The ring apparatus is autonomous, however our system also includes a mobile phone or computation device to which it connects wirelessly, and an earpiece for information retrieval. Finally, we will discuss how different finger worn sensors may be extended and applied to other domains.

Author Keywords
Pointing-based Interaction; Wearable Assistive Device; Intuitive Interfaces

ACM Classification Keywords
H.5.2 User Interfaces; H.5.1 Multimedia Information Systems; I.4.8 Scene Analysis; K.4.2 Social Issues: Assistive technologies for persons with disabilities

General Terms
Algorithms; Design; Human Factors

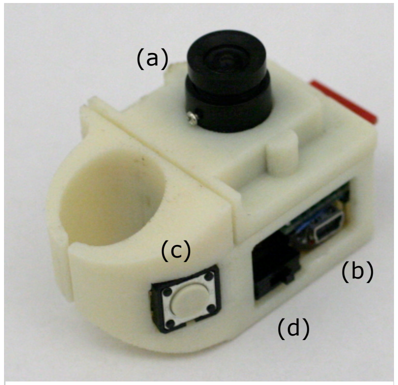


Figure 1: EyeRing prototype. The case was 3D printed with ABS nylon, and the electronics fit completely inside the skeletal compartment. (a) VGA mini camera, (b) Mini-USB port for recharging and reprogramming, (c) Trigger button, (d) On/Off switch.

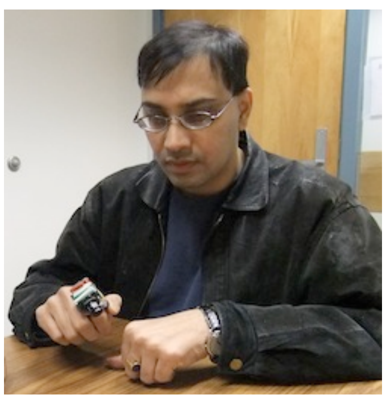


Figure 2: A blind user wearing EyeRing System. We are in the stages of developing a more slick design, which will have a center of gravity closer to the finger and overall smaller size.

Introduction

Despite the attention finger-worn interaction devices have received over the years, there is still much room for innovative design. Earlier explorations of finger-worn interaction devices may be divided into a few subspaces according to how they are operated: Pointing [1]; Tapping [2]; Gesturing [3]; Pressing/Clicking On-Device [4]. Our system is based on performing Free-air Pointing (FP) gestures, as well as Touch Pointing (TP) gestures. TP gestures utilize the natural touch sense, however the action trigger is not based on touch sensitivity of the surface, rather on an external sensor. Recently Chi et al presented *Seeing With Your Hand* [5], a glove apparatus that uses TP gestures. The haptic element of TP gestures is interesting especially in the case of assistive technologies for the visually impaired. This enabled them to get additional feedback on the object they want to interact with.

FP gestures on the other hand, are rooted in human behavior and natural gestural language. This was shown to be true by examining gestural languages of different cultures [6]. Usually FP gestures are used for showing a place or a thing in space -a passive action. However, augmenting FP for information retrieval is an interesting extension. Previous academic work in the field of FP gestures utilize a specialized sensor, usually an infrared connection, between the pointing finger and the target. This implies the environment to be rigged especially for such interaction. We chose to use a generalized approach by using a general-purpose camera. This choice breaks the bonds of dimensionality of a single signal source or sensor, as well as utilizing the ubiquitous visible light micro-cameras. Most of the work around FP and some TP gestures (e.g. the Optical Finger Mouse) are aimed towards sighted people. At the initial stage of this project, we choose to focus on a more compelling aspect - exploring how visually impaired people may benefit from finger worn devices.

EyeRing – A Finger-worn Assistant

The desire to replace an impaired human visual sense or augment a healthy one had a strong influence on the design and rationale behind EyeRing. To that end, we propose a system composed of a finger-worn device with an embedded camera, a computing element embodied as a mobile phone, and an earpiece for audio feedback. The finger-worn device is autonomous and wireless, and includes a single button to initiate the interaction. Information from the device is transferred to the computation element where it is processed, and the results are transmitted to the headset for the user to hear.

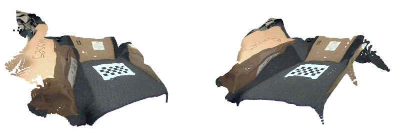
Typically, a user would single click the pushbutton switch on the side of the ring using his thumb (Figure 1). At that moment, a snapshot is taken from the camera and the image is transferred via bluetooth to the mobile phone. An Android application on the mobile phone then analyzes the image using our computer vision engine. Upon analyzing the image data, the Android application uses a Text-to-Speech module to read out the information though a hands-free head set. Users could change the preset mode by double-clicking the pushbutton and giving the system a brief verbal commands such as “distance”, “color”, “currency”, etc (Figure 2).

Applications For The Visually Impaired

The task of replacing the optical and nervous system of the human visual sense is an enormous undertaking. Thus, we choose to concentrate on learning the

(a) The user takes two photos pointing at the ground in front.

(b) A sample pair of photos



(c) A rendering of the reconstructed 3D scene.

(d) The estimated floor model.

Figure 3: Virtual Walking Cane. Estimating free walking space using a single EyeRing.

possible interaction mechanics for three specific scenarios outlined in this section.

1. Virtual Walking Cane: Compared to a steel cane, a finger worn device used for navigation is certainly less obtrusive, as well as fashionable and appealing. The essence of this application is to provide an approximate estimation of the clear walking space in front of the holder of the ring. User needs to user FP gesture to take two pictures of the space in front of him/her by pointing the camera and clicking, with some motion between the shots. The system clearly notifies the approximate free space in front. For this application, we employ the concept of Structure from Motion (SfM) from computer vision. Upon receiving the two images, an algorithm to recover the depth is performed. The general outline of the algorithm is as follows: (a) the two images are scanned for salient feature points, (b) the features in both images are matched into pairs, (c) for each points pair we estimate the motion it underwent, and from that – the approximate 3D position [7], (d) we use a robust method to fit a model of a floor to the sparse data, and return the distance of the clear walking path between objects on it. By repeatedly taking photos with motion, equivalent of moving a steel cane, we check the recovered 3D mapping of the floor and objects for any obstacles in the way of the user. Figure 3 outlines this process.

2. Currency Detector: This application is intended to help the user to identify currency bills (1$, 5$, 10$, 20$, 100$) to aid with payments. The interaction process is simple; a user would simply touch point index finger to a currency note (TP gesture) and click the button. The system will voice out what the note is (Figure 4). A detection algorithm based on a Bag of Visual Words (BoVW) approach [8] scans the image and makes a decision on the type of note it sees. We use Opponent Space [9] SURF features to retain color information, for notes detection. Our vocabulary was trained to be of 1000 features long, and we use a 1-vs.-all SVM approach for classifying the types of notes.

3. Color Detector: This application of EyeRing aids a visually impaired person to understand the color of an object. Again, the user interaction is simple; the user simply touch point (TP gesture) to an object and click the button to deliver an image for processing (Figure 5). The system analyses the image and returns the average color via audio feedback. We use a calibration step to help the system adjust to different lighting conditions. A sheet of paper with various colored boxes is printed, and a picture of it is taken. We rectify the region in the image so that it aligns with the colored boxes, and then extract a sample of the pixels covering each box. For predictions we use a normal distribution, set to the maximum likelihood of the perceived color.

User Reactions

We consulted a visually impaired user to get some initial feedback on our prototype design. This user was born sighted, but he lost his eyesight due to Retinitis Pigmentosa about 10 yeas ago. He uses a white cane to get around and relies on speech technology for reading and using a computer.

He was given the EyeRing system and explained how he could use it. His first reaction was "*this is neat*". He commented that his white cane often falls down when he leaves it by his chair and he has to find it on his own or ask someone else to pick it up for him. He favors the idea of having a ring on his finger that helps him to

Figure 4: Identifying a currency note

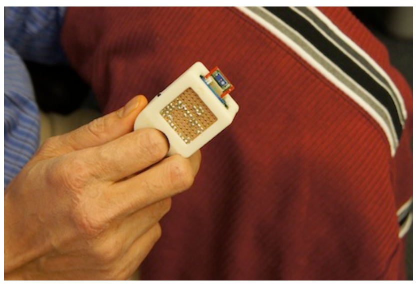


Figure 5: Identifying a color

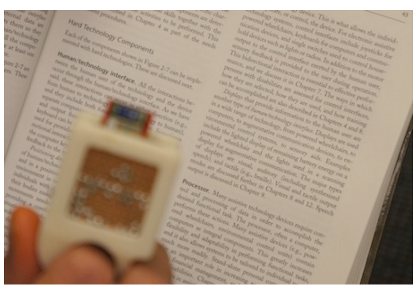


Figure 6: Concept of using EyeRing to read printed text.

navigate. He also added that his regular activities include choosing a shirt with a specific color (at the beginning of the day) and making payments with cash. For both color identification and currency identification he has to find his iPod touch, unlock screen, browse to the application, open it, and take a picture before he hear the result. With EyeRing, he said, "*Can't get any easier*". He also added that we could improve the color detection and currency detection algorithms. Overall, he was satisfied with the system as the interactions are seamless and he could use a single device to do three things by changing the mode.

Conclusion and Future Work

EyeRing suggests a novel interaction method for both visually impaired and sighted people. We choose to base the interaction on a human gesture that is ubiquitous in any language and culture – pointing with the index finger. This has determined the nature and design of the ring apparatus, location of the camera and trigger. Preliminary feedback received from a visually impaired user supports that EyeRing assistive applications are intuitive and seamless.

Future applications will include reading non-Braille (Figure 6) and we plan to extend EyeRing applications to domains beyond assistive technology. We believe that adding more hardware such as a microphone, an infrared light source or a laser module, a second camera, a depth sensor or inertial sensors, will open up a multitude of new uses for this specific wearable design.

Acknowledgements

We would like to thank the Fluid Interfaces group for their insightful remarks, Amit Zoran for the design of the prototype casing, and Savithru Jayasinghe for his work on the Android application development.

References

[1] Merrill, D. and Maes, P. Augmenting looking, pointing and reaching gestures to enhance the searching and browsing of physical objects. *In Proc. Pervasive'07*, (2007) 1–18.

[2] Lee, J., Lim, S. H., Yoo, J. W., Park, K. W., Choi, H. J. and Park, K. H. A ubiquitous fashionable computer with an i-Throw device on a location-based service environment. *In Proc. AINAW'07*, 2 (2007), 59-65.

[3] Rekimoto, J. Gesturewrist and gesturepad: Unobtrusive wearable interaction devices. In Proc. ISWC '01, (2001), 21-27.

[4] Chatterjee R. and Matsuno, F. Design of A Touch Sensor Based Single Finger Operated Wearable User-Interface Terminal. *In Proc. SICE-ICASE'06,* (2006), 4142-4147.

[5] Chi, L. Y., Ryskamp, R. A., Gomez, L. R. P., Ho, H. and Brin, S. Seeing with your hand. Google Patents, 2011.

[6] McNeill, D. *Language and gesture*, Cambridge University Press, 2000.

[7] Hartley, R. and Zisserman, A. *Multiple view geometry in computer vision*, Cambridge University Press, 2000.

[8] Csurka, G., Dance, C.., Fan, L., Willamowski, J. and Bray, C. Visual categorization with bags of keypoints. *In Proc. ECCV'04*, (2004), 59-74.

[9] Van de Sande, K., Gevers, T. and Snoek, C.. Evaluation of color descriptors for object and scene recognition. *In Proc. CVPR'08,* (2008), 1-8.