CamIO: a 3D Computer Vision System Enabling Audio/Haptic Interaction with Physical Objects by Blind Users

Huiying Shen, Owen Edwards, Joshua Miele and James M. Coughlan
The Smith-Kettlewell Eye Research Institute
2318 Fillmore St.
San Francisco, CA 94115 USA
{hshen, owen, jam, coughlan}@ski.org

ABSTRACT
CamIO (short for "Camera Input-Output") is a novel camera system designed to make physical objects (such as documents, maps, devices and 3D models) fully accessible to blind and visually impaired persons, by providing real-time audio feedback in response to the location on an object that the user is pointing to. The project will have wide ranging impact on access to graphics, tactile literacy, STEM education, independent travel and wayfinding, access to devices, and other applications to increase the independent functioning of blind, low vision and deaf-blind individuals. We describe our preliminary results with a prototype CamIO system consisting of the Microsoft Kinect camera connected to a laptop computer. An experiment with a blind user demonstrates the feasibility of the system, which issues Text-to-Speech (TTS) annotations whenever the user's fingers approach any pre-defined "hotspot" regions on the object.

Categories and Subject Descriptors
I.5.5 [**Pattern Recognition**]: Applications: Computer Vision

General Terms
Algorithm, Performance, Experimentation, Human Factors.

Keywords
Access, blindness, low vision.

1. INTRODUCTION
Major progress has been made towards accessibility in the digital realm for blind and visually impaired persons, with the advent of screenreader technology that works with a wide range of computers, smartphones and tablets, and the maturing framework of standards established by the World Wide Web Consortium (W3C) to promote website accessibility.

However, full access to physical objects such as printed documents, maps, devices and 3D models, which is very important for a variety of daily tasks at home, in the office and at school, is a challenging problem. Existing approaches to providing this access all have significant limitations. The simplest approach is to affix tactile (e.g., Braille) labels to an object, but such labeling is severely limited in terms of the amount of information it can encode (and the density with which labels can be placed), may interfere with the exploration and recognition of salient tactile features of the object, or may be inaccessible to those who can't read Braille.

It is also possible to modify objects by adding capacitive or touch sensors to "hotspot" locations of interest which can sense the presence of a person's finger near a hotspot, and thereby trigger the presentation of the hotspot annotation using TTS [6]. A convenient approach using smartpen-based talking graphics [4] is more promising since it leverages off-the-shelf pen technology and dot paper, rather than requiring custom-built technology. These last two approaches have the advantage of being able to encode information as densely as desired, and without requiring knowledge of Braille. Unfortunately, all three approaches require each object to be modified or augmented, typically by hand, which limits their widespread adoption.

Computer vision-based approaches have shown great promise for providing access to some types of objects in their original, unmodified form. Optical character recognition (OCR) is a well-established technology for reading printed documents, but [2] has limited ability to handle complex document layouts, and does not in itself facilitate interaction with a document; such interaction could be useful for navigating a complex document, or, for example, knowing where to sign a document.

Figure 1. (a) RGB-D camera (Kinect) used in current CamIO prototype. (b) Toy truck model used in experiment.

CamIO (Fig. 1) was conceived by Dr. Miele, a blind scientist with extensive experience in developing and evaluating audio/haptic interfaces for access to spatial information by people with visual disabilities. It builds on two computer vision-based projects for accessing physical objects: KnowWare [3] and Access Lens [1], both of which use a rigidly mounted camera to track a blind or visually impaired person's finger gestures as they explore a printed document or display screen. While these projects have provided the inspiration for CamIO, they are limited to the use of flat (2D) documents and display screens.

By contrast, CamIO is designed to provide access to general 3D objects, including relief maps, 3D models (e.g., anatomical models used in medicine or biology courses), appliances/devices used at home or work (e.g., microwave ovens, glucometers) as well as flat documents. The CamIO system requires a minimum of hardware (an inexpensive mounted camera and computer), and no modifications are needed to make an object accessible to it. In addition to providing audio or haptic feedback about hotspots that the user is pointing to, CamIO also has the potential to support a rich set of interactions with the object using multiple finger/hand gestures (e.g., circling a hotspot could trigger one type of action for a hotspot while double-clicking it would trigger another) and

provide audible guidance (as has been demonstrated with Access Lens) to guide the user's fingers to a hotspot of interest.

2. APPROACH

The CamIO system consists of an RGB-D camera (see Fig. 1a; Microsoft Kinect: http://www.microsoft.com/en-us/kinectforwindows/), rigidly mounted above the workspace area, which is connected to a laptop computer. The RGB-D camera acquires video images in which the RGB color and depth value (i.e., distance from the camera) is measured at each pixel. For the prototype CamIO system that we demonstrate in this paper, each object of interest is rigidly mounted on a flat base, which is a white surface marked with three bullseye fiducials in a specific configuration. These fiducials specify a 3D coordinate reference frame, which is rigidly bound to the object. However, we emphasize that future work on CamIO will harness the full 3D information provided by the RGB-D camera without needing any base to be mounted to the object.

The system is first launched with the object visible to the camera in the workspace area, while the user's hand is kept away from the workspace. During this time the depth image is sampled to acquire a background "reference" depth map; any deviations from this reference depth map, such as arise when the user places his/her hand on the object, signal the presence and location of the arm or hand. The reference depth map is automatically updated each time the object is moved significantly.

Given the pixel locations of the arm/hand, a simple procedure to extract the "skeleton" [5] of the arm/hand area is used to determine the presence and location of one or more outstretched fingers. The location of the fingertips are tracked over time (at the rate of approximately 9 frames per second), using motion continuity cues to filter out false positive fingertip detections.

Each object is annotated with one or more hotspot locations or regions, which are 3D locations/regions associated with specific text. For example, a 3D relief map may contain one hotspot for each topographic feature of interest in the map. Alternatively, a hotspot may define a specific region on the object, such as a lake on a map. Any time a fingertip approaches a hotspot, the system looks up the information associated with the hotspot and reads it aloud using TTS.

3. EXPERIMENTAL RESULTS

A simple experiment was conducted with a volunteer blind participant (male, age 44, with no usable vision) to test the performance of the system on four objects: a relief map, computer motherboard, toy truck (Fig. 1b) and flat plexiglass map (inscribed with tactile features). Four hotspot locations/regions were defined for each object. The participant was highly trained with the system and the objects tested, and practiced finding the four hotspots on each object before performing the experiment. In the experiment, for each object, the participant was requested to locate a specific hotspot, for a total of 12 trials per object (such that each hotspot was requested exactly three times in a random sequence). The experimenter noted whether the participant was successful in eliciting the annotation for the desired hotspot, or whether other hotspots were detected first.

For two objects (the truck and motherboard), all hotspots were successfully detected the first time in all trials. For the plexiglass map, one system error caused CamIO to detect an incorrect hotspot, which was immediately followed by the correct detection. The relief map was most challenging in that the hotspots were located close to one another (only a few inches apart, rather than the several-inch separation between hotspots on the other objects), and there were no tactile cues to identify some of the hotspots; as a result, the wrong hotspot was pointed to by the participant (and detected by the system) in five out of the 12 trials, but was immediately followed by the correct detection in all cases.

4. CONCLUSION

We have described the CamIO system for facilitating audio/haptic interaction with physical objects by blind or visually impaired users. An experiment with a blind participant using a prototype version of the system demonstrates its feasibility.

Future work on CamIO will focus on adding multiple finger gestures and guidance features (as in Access Lens). Object recognition will be implemented (to determine which object is currently in view from a library of objects stored in database), and greater use of 3D information will allow the finger location to be determined relative to the object without the use of a base plane or fiducial markings. Other sensors will also be investigated, such as the Leap Motion and PrimeSense sensors, which may offer higher spatial resolution, the ability to handle smaller objects and/or better ability to recognize gestures. Finally, blind and low vision persons will test the system on an ongoing basis to maximize the system's effectiveness and ease of use.

5. ACKNOWLEDGMENTS

This work was supported by a grant from the Department of Education, NIDRR grant number H133E110004. However, the contents do not necessarily represent the policy of the Department of Education, and you should not assume endorsement by the Federal Government.

6. REFERENCES

[1] Kane, S. K., Frey, B. and Wobbrock, J.O.. "Access lens: a gesture-based screen reader for real-world documents." Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. ACM, 2013.

[2] King, A. "Screenreaders, Magnifiers, and Other Ways of Using Computers." In Assistive Technology for Blindness and Low Vision. Manduchi, R. and Kurniawan, S. (eds.). CRC Press, a Taylor & Francis Group. Boca Raton, FL, 2012.

[3] Krueger, M.W. and Gilden, D. "Going Places with 'KnowWare': Virtual Reality Maps for Blind People." In: International Conference on Computers Helping People with Special Needs (ICCHP). 2002: 565-567.

[4] Miele, J.A. and Landau, S. "Audio-Tactile Interactive Computing with the Livescribe Pulse Pen 2." CSUN 25th Annual International Conference on Technology and Persons with Disabilities, San Diego, CA. 2010.

[5] Szeliski, R. Computer Vision: Algorithms and Applications. Springer, New York, 2010.

[6] Touch Graphics Research. Retrieved Jun. 20, 2013 from: http://touchgraphics.com/research.html