Using Computer Vision to Access Appliance Displays

Giovanni Fusco[1], Ender Tekin[1], Richard E. Ladner[2] and James M. Coughlan[1]
[1]The Smith-Kettlewell Eye Research Institute
2318 Fillmore St.
San Francisco, CA 94115 USA
{giofusco,ender,coughlan}@ski.org
[2]University of Washington
Computer Science & Engineering, DUB Group
Seattle, WA 98195 USA
ladner@cs.washington.edu

ABSTRACT
People who are blind or visually impaired face difficulties accessing a growing array of everyday appliances, needed to perform a variety of daily activities, because they are equipped with electronic displays. We are developing a "Display Reader" smartphone app, which uses computer vision to help a user acquire a usable image of a display, to address this problem. The current prototype analyzes video from the smartphone's camera, providing real-time feedback to guide the user until a satisfactory image is acquired, based on automatic estimates of image blur and glare. Formative studies were conducted with several blind and visually impaired participants, whose feedback is guiding the development of the user interface. The prototype software has been released as a Free and Open Source (FOSS) project.

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
People who are blind or visually impaired face severe difficulties performing a variety of daily activities due to the growing array of everyday appliances equipped with inaccessible electronic displays, often controlled with touch screens. Such appliances include microwave ovens, media players, digital blood pressure monitors, thermostats and vending machines. Available access technology to make printed documents accessible to persons with visual impairments is not equipped to handle the challenges of reading text on electronic displays, which is significantly different in appearance from printed text, and is often obscured by glare, reflections or poor contrast (Fig. 1b).

There is a small but growing body of work on computer vision-based display readers. One influential early work is Clearspeech [4], in which special markers are affixed around the borders of a display to help the system localize and read the display characters in the image. More recent work focuses on detecting and reading display text without the use of markers [6, 5]. Our current prototype uses a marker in order to provide rapid and accurate feedback to help guide the user to find the display, but we aim to eliminate the need for the marker in the future.

ASSETS '14, Oct 20-22 2014, Rochester, NY, USA
ACM 978-1-4503-2720-6/14/10.
http://dx.doi.org/10.1145/2661334.2661404

A key challenge in any camera-based system for blind or visually impaired users is that of aiming the camera properly to frame the target of interest. One such system, VizWiz [1], is a versatile tool that uses crowdsourcing to obtain descriptive information about images photographed by the user, which can include images of appliance displays. However, VizWiz has two major limitations: (a) it typically takes about 30 sec. for the user to receive feedback on an image, which is too long for some applications, and (b) it is difficult to obtain usable pictures because of frequent errors in "blur, lighting, framing, and composition" [1]. We seek to overcome both limitations in our Display Reader project.

We describe formative studies on a Display Reader prototype, which is tested with blind and low vision participants. The current phase of research focuses on a user interface (UI) that helps the user acquire one or more usable images of the desired display. The challenge of aiming the camera properly towards the display is compounded by an additional complication of LED and LCD displays: poor contrast and the presence of bright specularities (i.e., reflections from light sources). These complications make many images of displays difficult or impossible to read, even when the display is well framed in the image (e.g., as in Fig. 1b). The final stage of actually reading the display contents will be addressed in future research.

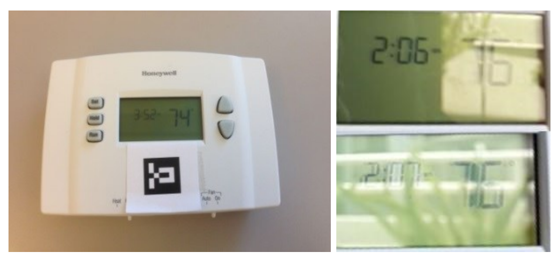

Figure 1. (a) Appliance with LCD display, with marker affixed below the display. (b) Examples of specularities (reflected glare) that severely impair visibility of display.

2. APPROACH
Our prototype consists of an Android smartphone, which is used to acquire images of the display of interest, running in conjunction with a laptop. (In the future we will port the software to the smartphone so that it runs as a self-contained app.) We affix a black-and-white fiducial paper marker [2] to each appliance, placed close to the display of interest (Fig. 1a). The marker includes a simple tactile marking to help blind users find the display; we note that it is already common practice for many people who are blind to affix tactile markings to appliances they own to help them locate and distinguish appliance buttons. Using a marker also simplifies and speeds up the process of detecting and localizing the display in the images acquired by the camera.

Rather than attempt to find and read the display "from scratch", which requires the ability to detect, localize and read arbitrary text with non-standard fonts and poor visibility, we exploit the prior knowledge of the display encoded in a *display template* that we have constructed for each appliance. The display template is a set of one or more high-quality, “clean” images of the display with annotations indicating the precise locations of display text fields. Other information, such as the locations and possible values and appearances of special symbol fields, can also be included. Such display templates may be generated by a sighted friend or a crowdsourcing process in which sighted volunteers acquire images of appliances and manually annotate them, to be shared freely on the web.

The prototype finds the marker (when visible) in each video frame, and computes a transformation between the marker in the current frame and the template to rectify (i.e., unwarp) the display, yielding a *rectified display image* (see Fig. 1b for examples).

We analyze each frame to assess its quality in two respects: the amount of blur and the severity of specularities (glare reflections) visible on the display region. Images with too much blur are rejected from further analysis. In our formative studies we found that different lighting conditions may require different thresholds for glare; we will investigate more robust methods for estimating glare in the future. Currently, our approach seeks a *single* high-quality view of the entire display, but in the future we will consider techniques for synthesizing such a view from multiple vantage points [3].

We devised a simple UI for the prototype (similar to that in [7]), using audio feedback to help guide the user until a satisfactory image is acquired. If the marker is detected and the *pose* is good (i.e., the marker and display are well centered in the camera’s field of view and are at an appropriate viewing distance), and if the blur is acceptably low, then a pleasant audio tone is issued. Verbal directions (“closer,” “farther”) indicate that the camera is too far or too close to the display; directions (“up”, “down”, “left”, “right”) help the user center the display in the image.

Our software is a Free and Open Source (FOSS) project, available at: http://www.ski.org/Rehab/Coughlan_lab/DisplayReader/

3. FORMATIVE STUDIES

We conducted formative studies of our prototype with five volunteer participants, four of whom are blind and one of whom has low vision. The feedback from each participant was used to make improvements to the UI and computer vision algorithms. Users were usually able to acquire good images of the tested appliances, demonstrating the feasibility of our approach.

Two main themes emerged from these studies. The first is that, while most participants found it fairly straightforward to frame the display properly in the camera’s field of view (with the help of a training session and the UI’s real-time feedback), they found it challenging to explore a wide range of viewing angles, which may be necessary to find a glare-free view of the display. This is because exploration of viewing angles requires the user to move the smartphone to multiple locations, while rotating the camera line of sight to keep the display in the field of view. An additional complication is that the prototype has no way of knowing in advance which viewing angle is optimal, and can offer no feedback about which viewing direction the user should try next.

The other theme that emerged is the trade-off between incorporating additional feedback in the UI, which offers more information to the user, and the need for a simple UI. Towards the end of the formative study we modulated the “pleasant” audio tone in three possible variations to communicate the severity of estimated glare (low, medium and high), which appeared to improve the search process without unduly complicating the UI.

Finally, the participants offered useful feedback about how they wanted a mature Display Reader system to operate. Three participants said they would be willing to aim the camera at the display for 30 sec. to a minute to get a reading, but no longer. Two participants expressed privacy concerns about a Display Reader system that would send images to a sighted assistant in the cloud. One enthusiastic participant recommended that training videos be posted online; this recommendation is consistent with our observation that the training sessions we conducted were essential for enabling the participants to operate the prototype.

4. CONCLUSION

We have described formative studies with a prototype Display Reader, which is a smartphone app that guides the user to take a usable image of a display. These studies demonstrate the feasibility of our approach, and suggest that future research should focus on improving the UI and training procedures to help users explore a wider range of viewing angles. Other avenues of future research include: algorithms for glare detection and estimating the image contrast); automatic reading of the display contents; visual display enhancement for low vision users; and using visual features extracted from the image template to eliminate the need for the marker. We will test Display Reader on an ongoing basis with blind or visually impaired volunteers to maximize the system’s effectiveness and ease of use during its development.

5. ACKNOWLEDGMENTS

Ladner was partially supported by the Intel Science and Technology Center on Pervasive Computing and NSF grant IIS-1116051; the other authors acknowledge support from NIH, grant 2 R01 EY018890-04, and the Dept. of Education, NIDRR grant H133E110004. David Vásquez provided extensive assistance with participant recruitment and experimental procedures.

6. REFERENCES

[1] Brady, E., Morris, M.R., Zhong, Y., White, S.C. & Bigham, J.P. “Visual Challenges in the Everyday Lives of Blind People.” Proc. ACM SIGCHI Conference on Human Factors in Computing Systems (CHI 2013), Paris, France, 2013.

[2] Garrido-Jurado, S., Muñoz-Salinas, R., Madrid-Cuevas, F. J., & Marín-Jiménez, M. J. “Automatic generation and detection of highly reliable fiducial markers under occlusion.” Pattern Recognition, 47(6), pp.2280-2292.

[3] Lee, P.S. & Ladner, R.E. “Homography-based Reflection Removal Specialized for Object Recognition on Mobile Platform.” Demo in CVPR 2013.

[4] Morris, T., Blenkhorn, P., Crossey, L., Ngo, Q., Ross, M., Werner, D., Wong, C. “Clearspeech: a display reader for the visually handicapped.” IEEE Trans Neural Syst. Rehab. Eng. Dec 2006, 14(4), pp.492-500.

[5] Rasines, I., Iriondo, P. and Díez, I. “Real-Time Display Recognition System for Visually Impaired.” Int’l Conf. on Computers Helping People with Special Needs (ICCHP '12). Linz, Austria, July 2012.

[6] Tekin, E., Coughlan, J. & Shen, H. “Real-Time Detection and Reading of LED/LCD Displays for Visually Impaired Persons.” IEEE Workshop on Applications of Computer Vision (WACV 2011). Kona, HI, Jan. 2011.

[7] Vázquez, M. & Steinfeld, A. “Helping visually impaired users properly aim a camera.” Proceedings of the 14th Intl. ACM SIGACCESS Conf. on Computers and Accessibility, (ASSETS '12), Boulder, CO, 2012, pp.95-102.