Appliance Displays: Accessibility Challenges and Proposed Solutions

Giovanni Fusco[1], Ender Tekin[1], Nicholas A. Giudice[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 Maine
Dept. Spatial Information Science & Engineering
Orono, ME 04469 USA
nicholas.giudice@maine.edu

ABSTRACT

People who are blind or visually impaired face difficulties using a growing array of everyday appliances because they are equipped with inaccessible electronic displays. We report developments on our "Display Reader" smartphone app, which uses computer vision to help a user acquire a usable image of a display and have the contents read aloud, to address this problem. Drawing on feedback from past and new studies with visually impaired volunteer participants, as well as from blind accessibility experts, we have improved and simplified our user interface and have also added the ability to read seven-segment digit displays. Our system works fully automatically and in real time, and we compare it with general-purpose assistive apps such as Be My Eyes, which recruit remote sighted assistants (RSAs) to answer questions about video captured by the user. Our discussions and preliminary experiment highlight the advantages and disadvantages of fully automatic approaches compared with RSAs, and suggest possible hybrid approaches to investigate in the future.

Categories and Subject Descriptors

I.5.5 [**Pattern Recognition**]: Applications: Computer Vision

Keywords

Access; blindness; low vision

1. INTRODUCTION

People who are blind or visually impaired face severe difficulties using a growing array of everyday appliances because they are 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. Current 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.

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 [5], and [3] focuses on the use of computer vision and crowdsourcing to help visually impaired persons find and actuate appliance controls such as buttons. Our current "Display Reader" (DR) system builds on [2] and uses 3-4 markers around the display border in order to provide rapid and accurate feedback to help guide the user to find the display, but we aim to minimize or eliminate the need for markers in the future.

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 [6]. We have designed the user interface (UI) for our system to provide explicit guidance to the user to help them capture the display of interest in the camera's field of view, and from a vantage point with minimal glare.

Systems that use remote sighted assistants (RSAs) to offer visual assistance, such as VizWiz [1], are becoming increasingly popular. However, a major drawback of using VizWiz for the purpose of interacting with appliances is that it typically takes about 30 sec. for the user to obtain feedback from a sighted operator on an image they have taken, and a frequent problem is that the image does not frame the object of interest properly. This latency is too slow to help the user obtain a usable picture of the appliance, which is typically an iterative process. Modern successors to VizWiz such as Be My Eyes (BME, see bemyeyes.org) provide a live video feed to a sighted operator, who can speak with the user and see whatever is captured by their camera. This approach offers real-time interaction between the user and operator and makes it possible for the operator to help the user acquire a usable picture of any desired target.

We describe recent developments to the DR system, including an improved UI and the ability of the system to automatically read seven-segment digits (which were chosen because they appear in many appliance displays and are non-standard fonts that are unreadable by many standard OCR systems, such as the KNFB Reader for iOS). We also report the start of a formative study that assesses the usability of DR and compares it with the use of BME to read displays, highlighting the advantages and disadvantages of fully automatic approaches compared with RSAs, and suggesting possible hybrid approaches to investigate in the future.

2. APPROACH

The current DR system 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 3-4 black-and-white fiducial paper markers (4 is an optimal number but space limitations may require the use of fewer markers) to each appliance, bordering the display of interest, to simplify and speed up our computer vision algorithms. We exploit the prior knowledge of the display encoded in a *display template*

ASSETS '15, October 26-28, 2015, Lisbon, Portugal
ACM 978-1-4503-3400-6/15/10.
http://dx.doi.org/10.1145/2700648.2811392

that we have constructed for each appliance [2]. 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, as in [3].

Figure 1. Sample LED display contains a string of digits, each of which is a standard seven-segment digit. For each digit, our algorithm estimates its bounding box (shown in white) and then reads it (results in yellow).

Based on past and new studies with blind and visually impaired volunteer participants and discussions with two blind accessibility experts (one of whom is a co-author on this paper), we improved upon our original UI [2]. Upon launch, the system sounds an ambient clock tone every few seconds to signal that the system is active but that no display is visible. The user is instructed to begin searching for the display using a strategy we call the "back-away" strategy: starting by holding the smartphone body flush with the surface of the display and then backing the smartphone away while keeping the body parallel to the display surface until audio feedback is heard. This strategy has proven effective with commercial apps such as the Digit-Eyes barcode scanner.[1]

The user's next goal is to move the smartphone camera to the *zone* (set of appropriate camera poses) that enables the display to be read, such that the display is roughly centered in the camera's field of view and the camera is far enough from the display to capture it in its entirety but still close enough to be well resolved. Any time the app detects one or more markers but determines that the camera is not in the designated zone, appropriate speech feedback ("up", "down", "left", "right", "closer", "back") is issued to help the user move the camera to enter the zone. For each video frame in which the camera is in the zone, the system estimates the amount of glare visible in the display region, and if the glare is acceptable then it interprets the display contents (see Fig. 1) and reads them aloud using text-to-speech. If the glare is above acceptable limits, then the system sounds another ambient tone every second to signal that the camera is in the zone but that a different vantage point is needed to find a view with less glare.

3. FORMATIVE STUDY

We conducted a formative study of the new version of DR with one volunteer participant who is completely blind. After a brief training session he was able to use the system to read three appliance displays (an LED clock, an LCD thermostat and an LED microwave). He read each display three times (for a total of 9 trials), with the correct reading obtained in all trials except for one erroneous reading caused by motion blur. The time it took to obtain a reading ranged from 7-48 sec. across all trials, with a median of 11 sec.

Next we asked the participant, who had previously installed BME on his own iPhone 6 but hadn't yet gotten it to work properly, to use his BME app to read the LED clock display with our guidance, which was successful. We then asked him to read the three appliance displays on his own using BME, but we found that he was unable to establish a connection more than once (despite a stable Wi-Fi connection). His first attempt to use BME on his own to read a display was partially successful; it took 64 sec. to get connected with a helper, who announced the reading 29 sec. later after providing verbal guidance to help him aim the camera. However, the helper made an interesting error, which was to read the display despite its left-most digit lying just outside of the camera's field of view. While BME is an extremely valuable app, even our limited experience with it in this study highlights the need for greater reliability and for some training on the part of the helpers in an RSA framework.

4. CONCLUSION

We have described recent developments on our DR project, which includes improvements to the UI, the addition of automatic digit recognition algorithms and the start of a formative study that assesses the usability of DR and compares it with the use of BME to read displays. While our new study is very preliminary, it highlights the relative advantages and disadvantages of fully automatic systems compared with RSAs, and reinforces the need for a hybrid approach (as in [3]) that combines elements of both technologies. In particular, the computer vision algorithms used in DR are effective for helping the user find a usable view of a display in much less time than might be required with RSAs such as BME, which requires a substantial amount of time to connect with a helper and frame a usable picture. However, it might be more effective to use an RSA to read the display contents rather relying on a computer vision algorithm, which could falter when confronted with unfamiliar display fonts and symbols.

5. ACKNOWLEDGMENTS

Giudice was supported by NSF grant CHS 1425337; the other authors acknowledge support from NIH, grant 2 R01 EY018890-05 and the National Institute on Disability, Independent Living and Rehabilitation Research, grant 90RE5008-01-00.

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 Conf. on Human Factors in Computing Systems (CHI 2013), Paris, France, 2013.

[2] Fusco, G., Tekin, E., Ladner, R. E., & Coughlan, J. M. (2014, October). "Using computer vision to access appliance displays." In Proc. 16th Int'l Conf. on Computers & Accessibility (ASSETS '14), pp. 281-282.

[3] Guo, A., Chen, X. A., & Bigham, J. P. (2015, April). "ApplianceReader: A Wearable, Crowdsourced, Vision-based System to Make Appliances Accessible." Proc. 33rd Annual ACM Conf. on Human Factors in Computing Systems (pp. 2043-2048). ACM.

[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] 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

---

[1] http://www.digit-eyes.com/cgi-bin/digiteyes.fcgi?action=scanningTips