

# Eye of the Beholder: Phone-Based Text-Recognition for the Visually-Impaired

Tudor Dumitraş, Matthew Lee, Pablo Quinones, Asim Smailagic, Dan Siewiorek and Priya Narasimhan  
Carnegie Mellon University  
{tdumitra, mlee, pag, asim, ds1p, priyan}@andrew.cmu.edu

## Abstract

*Blind and visually-impaired people cannot access essential information in the form of written text in our environment (e.g., on restaurant menus, street signs, door labels, product names and instructions, expiration dates). In this paper, we present and evaluate a mobile text-recognition system capable of extracting written information from a wide variety of sources and communicating it on-demand to the user. The user needs no additional hardware except an ordinary, Internet-enabled mobile camera-phone – a device that many visually-impaired individuals already own. This approach fills a gap in assistive technologies for the visually-impaired because it makes users aware of textual information not available to them through any other means.*

## 1. Introduction

Written information pervades our environment; everyday, we read text for our convenience (e.g., street names, detours), for our needs (e.g., restaurant menus, door labels, product instructions) and even for our safety (e.g., expiration dates, warnings). Most of this written information is not easily accessible to the blind and visually-impaired because it is not available in any other form; even the information provided in Braille can be used by less than 10% of the legally-blind persons in the United States [1]. This severely limits the independence of the visually-impaired by forcing them to rely on others for help. In this paper, we present the design of a mobile text-recognition system that makes written information available to visually-impaired people using an off-the-shelf, mobile-Internet enabled camera-phone.

Consider the scenario of Jane, a visually-impaired business-woman who values her independence. Before going to work in the morning, Jane needs to take her allergy medications. Not remembering when she had renewed her prescription, she reaches for her camera-phone and takes a picture of the bottle to check the expiration date. The phone reads out the expiration date, which happens to be last month. Jane decides to throw the bottle away and to make a detour to the pharmacy. At work, Jane schedules a lunch meeting with a client. After she walks to the restaurant, she snaps a picture of the street sign at the corner of the block to verify that she is at the correct intersection. As she enters the restaurant, Jane realizes that a small object is blocking her path. She takes a snapshot of the object; her phone tells her that the text

on the object is “Wet Floor.” Jane walks carefully around the sign and meets her client. At the table, she uses her phone to “read” the menu. Jane has a successful lunch meeting. Our phone-based solution gives Jane access to textual information that cannot be conveyed through any other means, keeping her safe and giving her more independence.

Previous work, including Haritaoglu [3] and Nakajima et al. [6], have developed a mobile system for dynamic translations of text in the environment with a PDA, high-speed cellular network, and a client-server architecture. Our approach shares a similar architecture, but we use an off-the-shelf camera-phone instead of a PDA to better address the needs of visually-impaired users. Instead of requiring the user to specify where text is in the image (which is nearly impossible for our users), we simplify the interaction by removing this extra step. Several other research systems [2],[5] and commercial products [7] use onboard cameras and the processing power of PDAs for text extraction.

In this paper, we make two main contributions:

- We present the challenges for designing and implementing a mobile phone-based text-recognition system (Section 2);
- We describe a client-server architecture for this application that blind users can access from their existing mobile phones (Section 3);

## 2. Challenges of Phone-Based Text Recognition

The challenges for implementing a phone-based text-recognition system for visually-impaired users stem from limitations of mobile phones, poor image quality (especially since most OCR engines were designed for scanning documents), and interaction constraints experienced by visually-impaired users:

**2.1. The image contains *out-of-focus text*.** For example, if the subject is too close, the image is too blurred for any text to be recognizable.

**2.2. The image contains *very small text*.** This occurs when the text itself is very small (such as the ingredients list on grocery items) or when standing further away from the text to fix the out-of-focus problem described above.

**2.3. The image contains *angled text*.** It is difficult even for the sighted users to align themselves perfectly perpendicular to the text.

**2.4. The image has low contrast.** This can result if the text and background have similar colors.

**2.5. The environment has poor lighting conditions.** Poorly-lit photos are common because non-sighted people have no awareness of ambient light levels.

**2.6. Camera phones lack the memory and processing power needed for OCR.** OCR engines require a large number of processing cycles and rely heavily on floating-point operations.

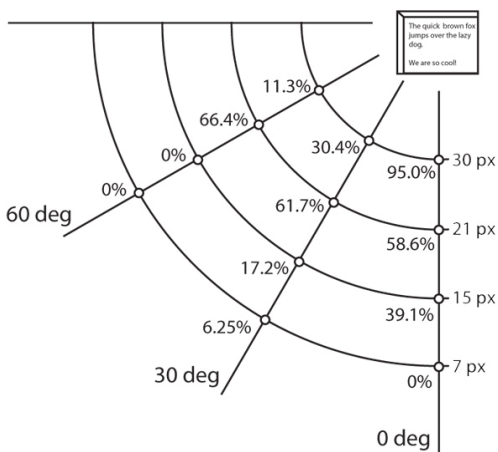
**2.7. Leveraging streaming video is impractical.** Video enables more sophisticated recognition techniques, but it is currently difficult to stream video through cellular data networks due to latency and bandwidth limitations.

**2.8. The visually-impaired user needs to locate text in the environment and point the camera-phone in the relevant direction.** To address this problem, users can leverage familiar conventions to find common locations for text (e.g. placards by doors and street signs at intersections). Users can also point the camera-phone directly at objects they hold in their hands (e.g., restaurant menus and product packaging). The vast majority of the visually-impaired can rely on their existing, albeit low, vision capabilities to locate text. Alternatively, our prototype can be used in conjunction with other systems that help blind people detect objects around them [4].

### 3. Architecture and implementation

Because of the current limitations of mobile phones, we use a client-server architecture: a custom application on a Nokia 6620 allows the user to snap a photo of an object and automatically send it to a server using an HTTP request over the GPRS network; a server-side script invokes the OCR engine, which extracts the text from the image. The server sends the extracted text back to the phone, where it is displayed and enunciated using a speech-synthesis engine. We selected a commercial OCR engine that can handle images with the characteristics described in Section 2.

**Performance Evaluation.** Figure 1 shows the recognition



**Figure 1. Recognition accuracy at different angles and distances.**

results for multiple images of an outdoor sign taken from various distances and angles. The recognition rate (computed as the number of correctly recognized characters out of the visible characters in the image) degrades significantly if the character size (height relative to the image) is less than approximately 21 pixels, or if the text angle is 30% or greater. However, in most cases, our system still provides useful information by correctly recognizing several complete words at the image center. The end-to-end latency is 33 s on average: 20.33 s for performing the HTTP request over the GPRS network, 5.5 s for compressing the image in a standard format at the client-side and 3 s for text recognition at the server-side.<sup>1</sup>

### 4. Conclusion

This paper presents a mobile text-recognition system that allows visually-impaired people to become aware of textual information normally inaccessible to them. Our approach requires the user to carry only an ordinary camera-phone and demonstrates that special-purpose hardware is unnecessary. Some of the challenges described in this paper may disappear in the future as the smart phones will be equipped with better lenses and cameras and as good text-recognition software becomes available for the phone platforms. The proof of concept outlined in this paper shows that this important problem can be addressed effectively using existing technologies.

### Acknowledgments

The authors would like to thank Dan Rossi, Wen Wu, and Anind Dey for their invaluable feedback.

### References

- [1] American Foundation for the Blind. "Estimated number of adult Braille readers in the United States," *Journal of Visual Impairment and Blindness*, 90(3), 1996, pp. 287.
- [2] V. Gaudissart et al., "SYPOLE: mobile reading assistant for blind people," In *International Conference on Speech and Computer (SPECOM)*, 2004, pp. 538-544.
- [3] I. Haritaoglu. "InfoScope: Link from Real World to Digital Information Space," In *International Conference on Ubiquitous Computing*, 2001, pp. 247-255.
- [4] P. Meijer, "An Experimental System for Auditory Image Representations," *IEEE Transactions on Biomedical Engineering*, 39(2), 1992, pp. 112-121.
- [5] M. Pilu, S. Pollard, "A light-weight text image processing method for handheld embedded cameras," In *British Machine Vision Conference*, Sept 2002.
- [6] H. Nakajima et al., "Portable Translator Capable of Recognizing Characters on Signboard and Menu Captured by Built-in Camera," In *ACL Interactive Poster and Demonstration Sessions*, 2005, pp. 61-64.
- [7] <http://www.knfbreader.com/>

<sup>1</sup> The 640x480 pixel JPEG files transferred to the server are 30 kB on average, while the text returned to the phone has less than 1 kB.