IJIIT

International Journal of Infinite Innovations in Technology|ISSN:2278-9057
IJIIT|Volume-IV|Issue-II|2015-2016 October|Paper-01
Reg. No.:20150901|DOI:V4I2P01

FINGER READER: A WEARABLE DEVICE FOR THE VISUALLY IMPAIRED

Pooja Deole
Computer Science Department
Vivekanand Education Society's Institute of Technology.
Mumbai, India.
pooja.deole@ves.ac.in

Shruti Kulkarni
Computer Science Department
Vivekanand Education Society's Institute of Technology,
Mumbai, India.
shruti.akulkarni@ves.ac.in

Abstract: **The current day scenario of reading for blind people is with the help of braille. Braille is a code- a system of dots that represent letters of an alphabet. All books are not written in Braille, thus the library of a visually impaired person is limited to countable number of books. The technology currently used in the market is having problem like focusing, accuracy, mobility and efficiency. Hence here we want to propose a device that will solve all the problems. A camera will be mounted on the device which will fit on the finger of the reader**
.

INTRODUCTION:

According to the estimates from World Health Organization (WHO) about 285 million people are visually impaired worldwide: 39 million are blind and 246 million have low vision (severe or moderate visual impairment).
So, basically what is a finger reader? Finger Reader is a device that assists visually impaired users with reading texts or words. It's basically a ring the user wears on their index finger that houses a small camera and some tactual actuators for feedback. When a visually impaired person wants to read some text, for example a newspaper, a paper book, any document or for that matter even an electronic book, they point their finger at the text that they wish to read and the device will read the words out loud. They can go faster, slower, go back, etc. that is the wearer can move over the text at whatever pace he wants to and the device will read it aloud.

Although, the visually impaired can read with the help of Braille, this type of device would be more beneficial for them as they can interpret almost any form of text. For example, the restaurant's menu card is not made taking into consideration the visually impaired. Having such a wearable device at their dispense would lead to a sense of independence in them.

Concept:

The concept of optical character recognition is used in this device. Optical Character Recognition (OCR) is mechanical or electronical conversion of typed, handwritten or printed text into machine-encoded text. It widely accepts data from any sort of document .It is a common method of digitizing printed texts so that it can be electronically edited, searched, stored more compactly, displayed on-line, and used in machine processes such as text to speech, machine translation, key data and text mining. OCR is a field of research in artificial intelligence, pattern recognition and in computer vision.

-The device reads printed text out loud with a synthesized voice, with the help of heavily modified open source software. One of the important concerns can be the weight of the device as it should be easily wearable and comfortable for the user. But fortunately, the weight of the device is nearly same as that of any regular ring.

Hardware details:

Multimodal feedback mechanism via vibration motors and a high-resolution mini video
Camera can be used. Vibration motors embedded on the top and bottom of the ring can be used to provide feedback. The dual material design can be used to improve flexibility.

Software details:

To accompany the hardware, a software stack that includes a text extraction algorithm, hardware control driver, integration layer can be developed. We start with image binarization and selective contour extraction.

Thereafter we look for text lines by fitting lines to triplets of pruned contours; we then prune for lines with feasible slopes. We look for supporting contours to the candidate lines based on distance from the line and then eliminate duplicates using a 2D histogram of slope and intercept.

Words with high confidence are retained and tracked as the user scans the line. For tracking we use template matching, utilizing image patches of the words, which we accumulate with each frame. We record the motion of the user to predict where the word patches might appear next in order to use a smaller search region.

When the user veers from the scan line, we trigger a tactile and auditory feedback. When the system cannot find more word blocks along the line we trigger an event to let users know they reached the end of the printed line. New high-confidence words incur an event and invoke the TTS engine to utter the word aloud. When skimming, users hear one or two words that are currently under their finger and can decide whether to keep reading or move to another area.

Advantages:

We focused on runtime efficiency, and typical frame processing time of our machine is within 20ms, which is suitable for real time processing. Low running time is important to support randomly skimming text as well as for feedback.

Drawback:

The voice is clipped but work is going on in order to improve the quality of sound. It doesn't work with text as small as, say, on a medicine bottle, but it can read 12-point printed text. Certain issues are observed associated with text alignment, inaccurate word recognition , slow speed of OCR software, and obscurity of photographs.

Difficulties were observed associated with reading minute texts such as a menu, text on a screen, or a business card.

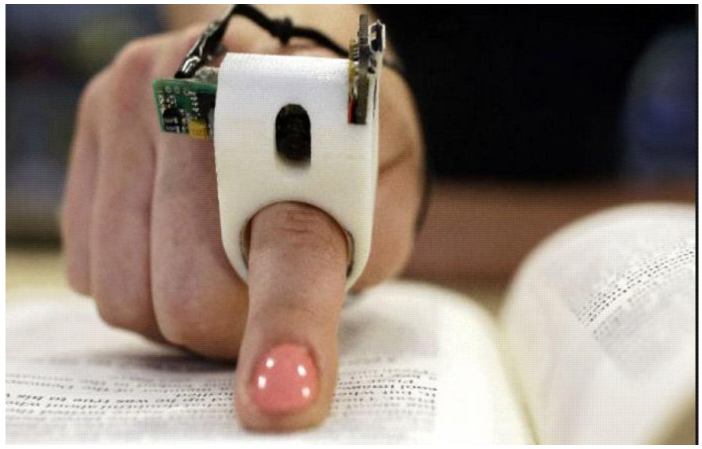
Solution to the existing problems:

In order to combat the existing problems, a novel hardware and software can be used that includes quick response, video-processing algorithms and different output mechanisms.

The ring prototype adjusts the camera at a fixed distance and utilizes the sense of touch when scanning the surface. The device can be made to contain few buttons and a simple user interface thus making it compact and user-friendly.

Improvement:

Yet more research has to be done on this device and a lot of improvisations are to be made. Moreover this device has not been brought to the market yet due to the cost associated with it. Regarding the future plans one of my suggestions is that the device should be able to accept many more languages as input and generate output is any language as desired by user. This will make the device more useful universally and will definitely increase the utility of the device.



What is the future of wearable tech?

Mobile phones and laptops are very fragile and at times become very complex to use. There is a growing need of more and more user-friendly devices which are sturdy and innovative. We expect in the coming years we will see a lot more wearable devices such as glasses, bracelets and watches enabling us to glance at some relevant information without running to places.

Evaluation:

The Finger Reader was evaluated in a two-step process: an evaluation of Finger Reader's text-extraction accuracy and a user feedback session for the actual Finger Reader prototype from four VI users. The accuracy of the text extraction algorithm in optimal conditions at 93.9% ($\sigma = 0.037$), in terms of character misrecognition, on a dataset of test videos with known ground truth was measured, which tells us that part of the system is working properly.

User Feedback:

Qualitative evaluation of Finger Reader with 4 congenitally blind users was conducted. The goals were (1) to explore potential usability issues with the design and (2) to gain insight on the various feedback modes (audio, haptic, or both). The two types of haptic feedbacks were: fade, which indicated deviation from the line by gradually increasing the vibration strength, and regular, which vibrated in the direction of the line (up or down) if a certain threshold was passed. Participants were introduced to Finger Reader and given a tablet with text displayed to test the different feedback conditions. Each single-user session lasted 1 hour on average and we used semi-structured interviews and observation as data gathering methods. Each participant was asked to trace through three lines of text using the feedbacks as guidance, and report their preference and impressions of the device. The results showed that all participants preferred a haptic fade compared to other cues and appreciated that the fade could also provide information on the level of deviation from the text line. Additionally, a haptic response provided the advantage of continuous feedback, whereas audio was fragmented. One user reported that "when [the audio] stops talking, you don't know if it's actually the correct spot because there are no continuous updates, so the vibration guides me much better." Overall, the users reported that they could envision the Finger Reader helping them with everyday tasks, explore and collect more information about their surroundings, and interact with their environment in a novel way.

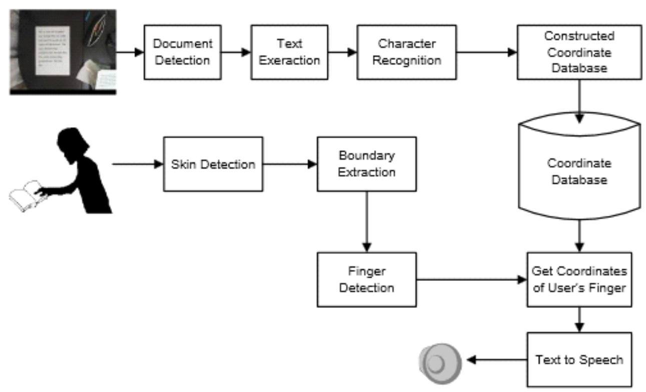


Figure 1. Block diagram for proposed blind reading system

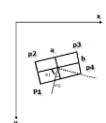


Figure 2. A rectangle

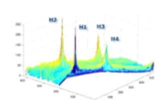
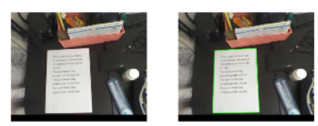
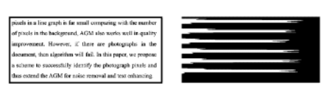


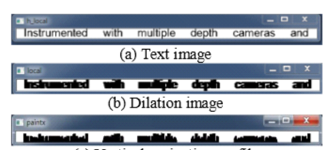
Figure 3. Hough space transform



(a) Text image (b) Detection result
Figure 4. Document Detection



(a) English document *(b) Horizontal projection profile*
Figure 5. Extraction of Textlines



(a) Text image
(b) Dilation image
(c) Vertical projection profile
Figure 6. Extraction of words

OVERVIEW OF THE PROPOSED SYSTEM

System Overview

Line Extraction: Within the focus region, we start with local adaptive image binarization (using a shifting window and the mean intensity value) and selective contour extraction based on contour area, with thresholds for typical character size to remove outliers. We pick the bottom point of each contour as the baseline point, allowing some letters, such as ‘y’,’g’ or ‘j’ whose bottom point is below the baseline, to create artifacts that will later be pruned out. Thereafter we look for candidate lines by fitting line equations to triplets of baseline points; we then keep lines with feasible slopes and discard those that do not make sense. We further prune by looking for supporting baseline points to the candidate lines based on distance from the line. Then we eliminate duplicate candidates using a 2D histogram of slope and intercept that converges similar lines together. Lastly, we recount the corroborating baseline points, refine the line equations based on their supporting points and pick the highest scoring line as the detected text line. When ranking the resulting lines, additionally, we consider their distance from the centre of the focus region to help cope with small line spacing, when more than one line is in the focus region.

Word Extraction: Word extraction is performed by the Tesseract OCR engine on image blocks from the detected text line. Since we focus on small and centric image blocks, the effects of homography between the image and the paper planes, and lens distortion (which is prominent in the outskirts of the image) are negligent. However, we do compensate for the rotational component caused by users twisting their finger with respect to the line, which is modeled by the equation of the detected line. The OCR engine is instructed to only extract a single word, and it returns: the word, the bounding rectangle, and the detection confidence. Words with high confidence are retained, uttered out loud to the user, and further tracked using their bounding rectangle as described in the next section. See Fig. 6 for an illustration.

Word Tracking and Signaling: Whenever a new word is recognized it is added to a pool of words to track along with its initial bounding rectangle. For tracking we use template matching, utilizing image patches of the words and an L2 norm matching score. Every successful tracking, marked by a low matching score and a feasible tracking velocity (i.e. it corresponds with the predicted finger velocity for that frame), contributes to the bank of patches for that word as well as to the prediction of finger velocity for the next tracking cycle. To maintain an efficient tracking, we do not search the entire frame but constrain the search region around the last position of the word while considering the predicted movement speed. We also look out for blurry patches, caused by rapid movement and the camera’s rolling shutter, by binarizing the patch and counting the number of black vs. white pixels. A ratio of less than 25% black is considered a bad patch to be discarded. If a word was not tracked properly for a set number of frames we deem as “lost”, and remove it from the pool. See Fig. 7 for an illustration.

DISCUSSIONS:

Efficiency over independence: All participants mentioned that they want to read print fast (e.g. “to not let others wait, e.g. at a restaurant for them to make a choice”, P3) and even “when that means to ask their friends or a waiter around” (P1). Though, they consider the Finger Reader as a potential candidate to help them towards independence, since they want to explore on their own and do not want others suggest things and thus subjectively filter for them (e.g. suggesting things to eat what they think they might like). From our observations, we conclude that the Finger Reader is an effective tool for exploration of printed text, yet it might not be the best choice for “fast reading” as the speed of the text synthesis is limited by how fast a user actually flows across the characters.

Exploration impacts efficiency: The former point underlines the potential of Finger Reader-like devices for exploration of print, where efficiency is less of a requirement but getting access to it is. In other words, print exploration is only acceptable for documents where (1) efficiency does not matter, i.e. users have time to explore or (2) exploration leads to efficient text reading. The latter was the case with the business cards, as the content is very small and it is only required to pick up a few things, e.g. a particular number or a name. P2, for instance, read his employment card with the Finger-Reader after finishing the business cards task in session 1. He was excited, as he stated *“I never knew what was on there, now I know”*.

Visual layouts are disruptive: The visual layout of the restaurant menu was considered a barrier and disruption to the navigation by P2 and P3, but not by P1. All of the three participants called the process of interacting with the Finger Reader “exploration” and clearly distinguished between the notion of *exploration* (seeing if text is there and picking up words) and *navigation* (i.e. reading a text continuously). Hence, navigation in the restaurant menu was considered a very tedious task by P2 and P3. Future approaches might leverage on this experience by implementing meta-recognition algorithms that provide users with layout information. A simple approach could be to shortly lift the finger above the document, allowing the finger-worn device to capture the document layout and provide meta-cues as the user navigates the document (e.g. audio cues like “left column” or “second column”).

Feedback methods depend on user preference: We found that each participant had his own preference for feedback modalities and how they should be implemented. For instance P1 liked the current implementation and would use it as-is, while P2 would like a unified audio feedback for finger rotation and straying off the line to make it easily distinguishable and last, P3 preferred tactile feedback. Thus, future Finger Reader-like designs need to take individual user preferences carefully into account as we hypothesize they drastically impact user experience.

LIMITATIONS:

The current design of the Finger Reader has a number of technical limitations, albeit with ready solutions. The camera does not auto-focus, making it hard to adjust to different finger lengths. In addition, the current implementation requires the Finger Reader to be tethered to a companion computation device, e.g. a small tablet computer.

The studies presented earlier exposed a number of matters to solve in the software. Continuous feedback is needed, even when there is nothing to report, as this strengthens the connection of finger movement to the "visual" mental model. Conversely, false realtime-feedback from an overloaded queue of words to utter caused an inverse effect on the mental model, rendering "ghost text". The speech engine itself was also reported to be less comprehensible compared to other TTSs featured in available products and the audio cues were also marked as problematic. These problems can be remedied by using a more pleasing sound and offering the user the possibility to customize the feedback modalities.

CONCLUSION:

We contributed Finger Reader, a novel concept for text reading for the blind, utilizing a local-sequential scan that enables continuous feedback and non-linear text skimming. Motivated by focus group sessions with blind participants, our method proposes a solution to a limitation of most existing technologies: reading blocks of text at a time. Our system includes a text tracking algorithm that extracts words from a close-up camera view, integrated with a finger-wearable device. A technical accuracy analysis showed that the local-sequential scan algorithm works reliably. Two qualitative studies with blind participants revealed important insights for the emerging field of finger-worn reading aids.

First, our observations suggest that a local-sequential approach is beneficial for document exploration–but not as

much for longer reading sessions, due to troublesome navigation in complex layouts and fatigue. Access to small bits of text, as found on business cards, pamphlets and even newspaper articles, was considered viable. Second, we observed a rich set of interaction strategies that shed light onto potential real-world usage of finger-worn reading aids. A particularly important insight is the direct correlation between the finger movement and the output of the synthesized speech: navigating within the text is closely coupled to navigating in the produced audio stream. Our findings suggest that a direct mapping could greatly improve interaction (e.g. easy "re-reading"), as well as scaffold the mental model of a text document effectively, avoiding "ghost text". Last, although our focus sessions on the feedback modalities concluded with an agreement for cross-modality, the thorough observation in the follow-up study showed that user preferences were highly diverse. Thus, we hypothesize that a *universal* finger-worn reading device that works uniformly across all users may not exist and that personalized feedback mechanisms are key to address needs of different blind users. In conclusion, we hope the lessons learned from our 18month-long work on the Finger Reader will help peers in the field to inform future designs of finger-worn reading aids for the blind. The next steps in validating the Finger Reader are to perform longer-term studies with specific user groups (depending on their impairment, e.g. congenitally blind, late-blind, low-vision), investigate how they appropriate the Finger Reader and derive situated meanings from their usage of it. We also look to go beyond usage for persons with a visual impairment, and speculate the Finger Reader may be useful to scaffold dyslexic readers, support early language learning for preschool children and reading non-textual languages.

References:

http://www.robotica-up.org/PDF/Wearable4Blind.pdf
http://fluid.media.mit.edu/sites/default/files/paper317.pdf
http://www.techtimes.com/articles/9949/20140708/device-helps-blind-read-print-fingerreader-audio-reading-gadget-index.htm
http://fluid.media.mit.edu/sites/default/files/FingerReaderFAQ%20(4).pdf
http://techcrunch.com/2014/04/17/mits-fingerreader-helps-the-blind-read-with-a-swipe-of-a-digit/
http://www.humanity.org.uk/articles/blindhttp://mashable.com/2014/07/13/blind-fingerreader/
ness-visual-impairment/learning-to-read-braille
http://www.ijetae.com/files/Volume5Issue1/IJETAE_0115_62.pdf

[1] www.electronic-circuits-diagrams.com

[2] www.circuitstoday.com
[3] www.circuitlake.com

A. Software Books & Websites

[1] Gary Cornell & Jonathan Marrison, Programming VB.Net: A Guide for experienced programmers, Second Edition 2002, ASPToday Publication, ISBN (pbk): 1-893115-99-2, 424 Pages.