THESIS READER FOR VISUALLY IMPAIRED

Tejaswini Patil[1], Prof. Ajay Talele[2]
[1,2]VIT, Pune

Abstract

Typically, the research scholars with visual impairment (VIRS) is on a rise. But difficulty in assessing the quality research resource is the major challenge for the VIRS individuals. The current research proposes a thesis reader setup which preliminarily focuses on increasing the reading experience for VIRS. The proposed system includes a camera to capture the contents from the thesis, convert it to image separating the text from the images and reading it loud for the VIRS, additionally the research scholar can save the document with his/her personal comments for easy future reference reading. The OCR is illumination-invariant and angle/tilt-invariant. The current setup has typical challenges such as, not being easy to move around and difficulty with initial setup. The current system has a score of 3 on the scale of 1-5 for the readers on the comfort and ease of use, based on the survey with 15 scholars with VIRS. Also, the scholars with normal vision appreciated the current system.

Index Terms: Text reading, Optical character recognition, Text to speech, Text extraction

I. INTRODUCTION

WORLDWIDE, 285 million people are visually impaired (WHO, 2012)[1]. This population faces important challenges related to orientation and mobility. Visual impairment may result in a reduction of autonomy in daily life[2]. Very often information is presented in visual form only and visually impaired people are thus excluded from accessing this information. This concerns important domains such as administrative tasks and education[3]. Challenges are also related to mobility and orientation. This issue presents a social challenge as well as an important research area. Therefore, if assistive technology can support visually impaired in at least one of these tasks, it is going to make a very relevant social impact[4]. This dissertation turns to the development of technology for assisting visually impaired to access visual information.

Moreover, textual information is everywhere, not only in the user's living room, and can exist under different forms such as newspapers, books, or text in natural scenes (signs, screen, schedules, etc.).

The idea of this system is to increase their autonomy by using wireless devices anytime and anywhere. To realize this dedicated platform, we have adopted a user-centered design in close relationship with low-vision people[5].

Fig. 1 gives an overview of the system. Users interact with a dedicated human-machine interface, specifically created for their disabilities. The image taken by the embedded camera is sent to the text detection module. When the text zone is established, the OCR module tries to extract the useful information and sends it to the text-to-speech module. The system is also supported with ASR(Automatic Speech Recognition) so that the user has control over the device through speech commands.

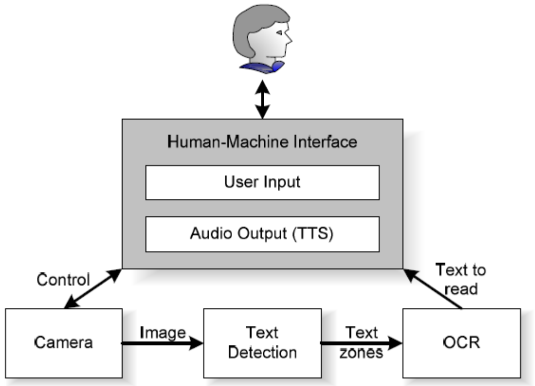


Fig.1 Overview of the system

- With design space we address the existing solutions for interactive reading for visually impaired people. We performed an exhaustive research of the literature and analyzed the corpus regarding non-visual interaction.
- Design is an iterative design process that includes users from the start to the end of the development. We worked in close collaboration with visually impaired people. However, the methodology of this design process is usually largely based on the use of the visual sense.
- We propose that the design of usable interaction may benefit from better understanding of how visually impaired people explore books. Second, we studied the possibility to include further functionality into the thesis reader prototype by making use of advanced non-visual interaction. These investigations were only of preliminary nature and open up avenues for future work in this field.

This paper is organized as followed. Section 3 describes text detection challenges and the approach we have followed. Section 4 details the choices we made for the human-machine interface. In the final section, we address perspectives in research activities and conclude the paper.

II. Related Work

Available technologies, such as smartphone applications, screen readers, flatbed scanners, e-Book readers, and embossers, are considered to have slow processing speeds, poor accuracy or cumbersome usability[2]. Hence the need to develop a prototype of thesis reader which would overcome the above drawbacks was strongly felt.

III. OCR for Text Detection

Optical character recognition (OCR) technology offers blind and visually impaired persons the capacity to scan printed text and then speak it back in synthetic speech or save it to a computer[7]. Little technology exists to interpret graphics such as line art, photographs, and graphs into a medium easily accessible to blind and visually impaired persons. It also is not yet possible to convert handwriting, whether script or block printing, into an accessible medium.

There are three essential elements to OCR technology—scanning, recognition, and reading text. Initially, a printed document is scanned by a camera[8]. OCR software then converts the images into recognized characters and words. The synthesizer in the OCR system then speaks the recognized text. Finally, the information is stored in an electronic form, either in a personal computer (PC) or the memory of the OCR system itself[9].

The recognition process takes account of the logical structure of the language[10]. An OCR system will deduce that the word "tke" at the beginning of a sentence is a mistake and should be read as the word "the." OCR's also use a lexicon and apply spell checking techniques similar to those found in many word processors[16].

Word extraction is performed by the Tesseract OCR engine on image blocks from the detected text line[11]. 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[12]. However, we do compensate for the rotational component caused by users twisting their finger with respect to the line.

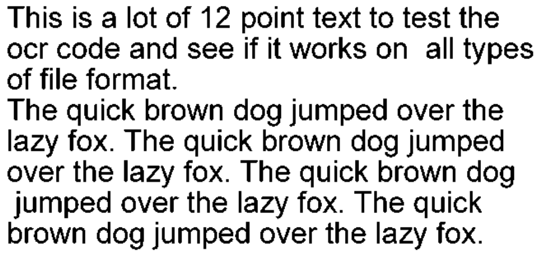
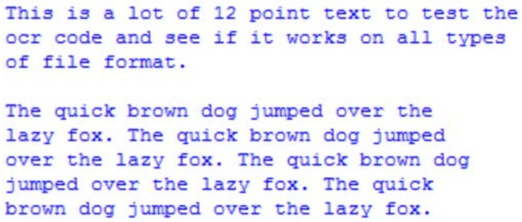

This is a lot of 12 point text to test the
ocr code and see if it works on all types
of file format.
The quick brown dog jumped over the
lazy fox. The quick brown dog jumped
over the lazy fox. The quick brown dog
jumped over the lazy fox. The quick
brown dog jumped over the lazy fox.

Fig.2 Sample Input Image

Early versions needed to be trained with images of each character, and worked on one font at a time. Advanced systems capable of producing a high degree of recognition accuracy for most fonts are now common, and with support for a variety of digital image file format inputs.



```
This is a lot of 12 point text to test the
ocr code and see if it works on all types
of file format.

The quick brown dog jumped over the
lazy fox. The quick brown dog jumped
over the lazy fox. The quick brown dog
jumped over the lazy fox. The quick
brown dog jumped over the lazy fox.
```

Fig.3 Text Detected

The OCR engine is instructed to only extract a single word, and it returns: the word, the bounding rectangle, and the detection confidence.

IV. Text to Speech

A TTS Engine converts written text to a phonemic representation, then converts the phonemic representation to waveforms that can be output as sound[17]. TTS engines with different languages, dialects and specialized vocabularies are available through APIs[18].

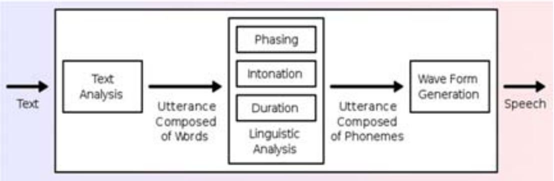


Fig.4 Overview of a TTS system

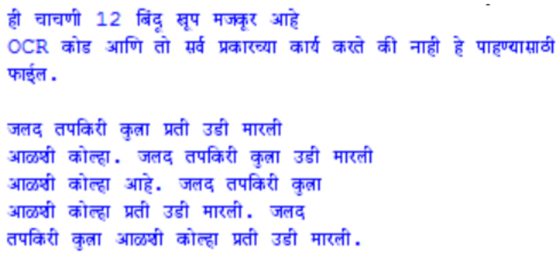
We can make the computer speak with Python. Given a text string, it will speak the written words in the English language. This process is called Text To Speech or shortly TTS.

gTTS is a module and command line utility to save spoken text to mp3 via the Google Text to Speech (TTS) API. This module supports many languages and sounds very natural. **gTTS text to speech module** has been used in the development of this prototype.

V. Translation

While interpreting—the facilitating of oral or sign-language communication between users of different languages—antedates writing, translation begins only after the appearance of written literature[19].

The Google Cloud Translation API can dynamically translate text between thousands of language pairs. The Cloud Translation API lets websites and programs integrate with the translation service programmatically[20].



```
ही चाचणी 12 बिंदू खूप मजकूर आहे
OCR कोड आणि तो सर्व प्रकारच्या कार्य करते की नाही हे पाहण्यासाठी
फाईल.

जलद तपकिरी कुत्रा प्रती उडी मारली
आळशी कोल्हा. जलद तपकिरी कुत्रा उडी मारली
आळशी कोल्हा आहे. जलद तपकिरी कुत्रा
आळशी कोल्हा प्रती उडी मारली. जलद
तपकिरी कुत्रा आळशी कोल्हा प्रती उडी मारली.
```

Fig. 5 Text translated into Marathi

The Google Translation API is part of the larger Cloud Machine Learning API family. This API is used in the prototype so as to enable the user with various output languages.

A translation that meets the criterion of fidelity (faithfulness) is said to be "faithful"; a translation that meets the criterion of transparency, "idiomatic". Depending on the given translation, the two qualities may not be mutually exclusive[21].

The criteria for judging the fidelity of a translation vary according to the subject, type and use of the text, its literary qualities, its social or historical context, etc[22].

The criteria for judging the transparency of a translation appear more straightforward: an unidiomatic translation "sounds wrong"; and, in the extreme case of word-for-word translations generated by many machine-translation systems, often results in patent nonsense.

Nevertheless, in certain contexts a translator may consciously seek to produce a literal translation. Translators of literary, religious or historic text often adhere as closely as possible to the source text, stretching the limits of the target language to produce an unidiomatic text. A translator may adopt expressions from the source language

VI. Speech Recognition

Speech recognition (SR) is the inter-disciplinary sub-field of computational linguistics that develops methodologies and technologies that enables the recognition and translation of spoken language into text by computers. It is also known as "automatic speech recognition" (ASR), "computer speech recognition", or just "speech to text" (STT). It incorporates knowledge and research in the linguistics, computer science, and electrical engineering fields.

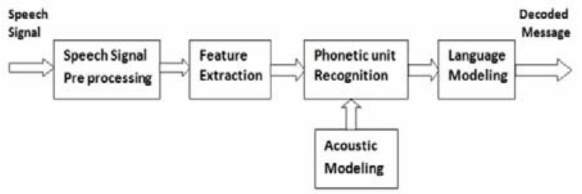


Fig. 6 Speech Recognition Block Diagram

Neural networks emerged as an attractive acoustic modeling approach in ASR. Neural networks have been used in many aspects of speech recognition such as phoneme classification, isolated word recognition,[13] and speaker adaptation[14].

In contrast to HMMs (Hidden Markov models), neural networks make no assumptions about feature statistical properties and have several qualities making them attractive recognition models for speech recognition. When used to estimate the probabilities of a speech feature segment, neural networks allow discriminative training in a natural and efficient manner[23].

Google Cloud Speech API enabled us to **convert audio to text** by applying **powerful neural network models** in an easy to use API. The API **recognizes over 80 languages and variants,** to support your global user base. We could transcribe the text of users dictating to an application's microphone, enable command-and-control through voice[24].

Speech API can **stream text results**, returning partial recognition results as they become available, with the recognized text appearing immediately while speaking. The prototype can successfully **handle noisy audio** from a variety of environments.

VII. Conclusion

This paper presented a system for camera-based online text recognition in a system designed for blind and visually impaired people. This particular design is challenging and relevant for both the document recognition and the embedded systems fields. Currently, the system performs well in detection and recognition of text with nearly uniform backgrounds. But it is not accurate enough to take into account more complex pictures like outdoor situations. New approaches using colour information are under investigation.

A first prototype device has already been implemented in order to have a quick feedback from the blind users. This methodology of work is the key aspect of our user-centered approach. Even though the system is in its infancy and needs further works, first opinions are satisfying and promising. Practical applications such as those already implemented can easily be added in order to build a complete talking assistant.

VIII. References

[1] WHO. (2012). Visual Impairment and blindness Fact Sheet N° 282. World Health Organization.

[2] Anke Brock. Interactive Maps for Visually Impaired People: Design, Usability and Spatial Cognition. Human-Computer Interaction [cs.HC]. Universite Toulouse 3 Paul Sabatier, 2013.English. <tel-00934643>

[3] WHO. (2001). International Classification of Functioning, Disability and Health (ICF). Geneva, Switzerland: World Health Organization.

[4] WHO. (2010). International Classification of Diseases (ICD-10). Geneva, Switzerland: World Health Organization.

[5] D. Doermann, "Progress in camera-based document image analysis", 7th ICDAR Conference, vol.1, pp.606 – 616, August 2003

[6] V. Wu, R. Manmatha, "Text finder: an automatic system to detect and recognize text in images", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 21, n.11, pp. 1224-1229, 1999

[7] H. Li, D. Doermann, "Automatic text detection and tracking in digital video", IEEE Transactions on Image Processing, vol. 9, n.1, pp. 147-156, 2000

[8] P. Clark, M. Mirmehdi, "Recognizing text in real scenes", International Journal on Document Analysis and Recognition, vol.4, pp. 855-877, 1998

[9] J. Kittler, J. Illingworth, "Threshold selection based on a simple image statistic", CVGIP vol. 30 (1985) 125-147

[10] W. Niblack, "An introduction to Digital Image Processing", Englewood cliffs, N.J.: Prentice Hall (1986) 115—116

[11] R.G.Casey, E. Lecolinet, "A survey of methods and strategies in character segmentation", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 18, n°7, 1996

[12] O.D.Trier, A.K.Jain and T.Taxt, "Feature extraction
methods for character recognition", a survey,
Pattern Recognition, vol. 29, n°4, pp 641-662, 1996

[13] Waibel, A.; Hanazawa, T.; Hinton, G.; Shikano, K.; Lang, K. J. (1989). "Phoneme recognition using time-delay neural networks". IEEE Transactions on Acoustics, Speech and Signal Processing. 37 (3): 328–339. doi:10.1109/29.21701.

[14] Wu, J.; Chan, C. (1993). "Isolated Word Recognition by Neural Network Models with Cross-Correlation Coefficients for Speech Dynamics". IEEE Transactions on Pattern Analysis & Machine Intelligence. 15 (11): 1174–1185. doi:10.1109/34.244678

[15.] Peters, J.-P., Thillou, C., and Ferreira, S. Embedded reading device for blind people: a user-centered design. In Proc. of ISIT, IEEE (2004), 217–222.

[16] Rissanen, M. J., Vu, S., Fernando, O. N. N., Pang, N., and Foo, S. Subtle, natural and socially acceptable interaction techniques for Ringter faces-Finger-Ring shaped user interfaces. In Distributed, Ambient, and Pervasive Interactions. Springer, 2013, 52–61.

[17] Sears, A., and Hanson, V. Representing users inaccessibility research. In Proc. of CHI, ACM (2011), 2235–2238.

[18] Shen, H., and Coughlan, J. M. Towards a real-time system for finding and reading signs for visually impaired users. In Proc. of ICCHP, Springer (2012), 41–47.

[19] Shilkrot, R., Huber, J., Liu, C., Maes, P., and Nanayakkara, S. C. Finger reader: A wearable device to support text reading on the go. In CHI EA, ACM (2014), 2359–2364.

[20] Shinohara, K., and Tenenberg, J. A blind person's interactions with technology. Commun. ACM 52,8 (Aug. 2009), 58–66.

[21] Shinohara, K., and Wobbrock, J. O. In the shadow of misperception: Assistive technology use and social interactions. In Proc. of CHI, ACM (2011), 705–714.

[22] Smith, R. An overview of the tesseract OCR engine. InProc. of ICDAR, vol. 2, IEEE (2007), 629–633.

[23] Stearns, L., Du, R., Oh, U., Wang, Y., Findlater, L., Chellappa, R., and Froehlich, J. E. The design and preliminary evaluation of a finger-mounted camera and feedback system to enable reading of printed text for the blind. In Workshop on Assistive Computer Vision and Robotics, ECCV, Springer (2014).

[24] Yi, C., and Tian, Y. Assistive text reading from complex background for blind persons. In Camera-Based Document Analysis and Recognition. Springer, 2012, 15–28.