Image To Speech Converter Android App for partially blind and Tourist

Mrs.Asma Parveen I. Siddavatam, Asst. Prof., Dept of Information Technology, VESIT.

Aniket Nighot,Uddesh Karda, Vaibhav Gaikwad Dept of Information Technology, VESIT.

Abstract - This paper presents an android application which allows visually impaired people to access information from newspapers, magazines etc by transforming textual information into speech. The project aims to increase computation speed of mobile based optical character recognition systems by having server based processing systems and also aims at achieving high accuracy rate.

Keywords - OCR, Image to Text, Text to Speech, Image Processing.

1. Introduction

Real world contains too many significant messages and useful information which cannot be ignored or left unread. Sometimes a signboard or any other notice could carry an important message or even danger notice that could be missed by partially blind and illiterate people. This application is mainly beneficial to access printed text which may carry significant messages.

A. Optical character recognition

Optical Character Recognition, or OCR, is a technology that enables the conversion of different types of documents, such as scanned paper documents, PDF files or images captured by a digital camera into editable text.
The comparison in the table below explains the differences between “on device” processing with Mobile OCR Engine or a custom developed back end server.

Parameter	**Mobile OCR**	**Server Based OCR**
Application Type	Thick” Apps with OCR technology included, OS native	Thin” Client approach Image capture, data communication with the back-end, More generic Apps for multiple devices
CPU’s	ARM, x86	Any Architecture
Device Memory	Limited	Unlimited
Online connectivity	Not Required	Required

Technology	Compact & tuned to make it work on mobile devices	Broad set of technologies can be used for image preprocessing, document analysis, character recognition, data extraction etc.
Processing speed	less	More

Table Ⅰ. Comparison between Mobile OCR and server based OCR
Source:(Sathiapriya Ramiah, 2015)[1]

B. OCR Technologies

There are numerous open source as well as commercial OCR engines available in the market with their own strengths and weaknesses for character recognition. Many open source communities offer engines such as GOCR, Cuneiform, OCRAD, Tesseract and OCROPUS. And also commercial OCR engines are available such as ABBYY Finereader, OmniPage, and Microsoft Office Document Imaging.
Based on the literature survey conducted on various open source engines like Cunieform**,** GOCR, OCRAD and Tesseract, it was found that the Tesseract OCR engine outperformed other open source engines. Despite noisy data, Tesseract proved to be the best free and open source OCR engine in terms of accuracy and processing time as shown in below table .1.
Tesseract works well on all computer operating systems as well as Android and IOS mobile platforms. Due to Tesseract being an open source engine, it is very popular and there are a lot of academic experiments and OCR software developments conducted successfully.
In the proposed system Tesseract OCR engine is being used for character recognition.

Parameter	**Cunieform**	**gocr**	**ocrad**	**tesseract**
license	BSD	GPL 2	GPL3	Apache 2.0
courier/ black	61%	67%	21%	81%
courier/ gray	×	67%	21%	81%
times/black	94%	76%	82%	92%
times/gray	×	76%	82%	92%
verdana/	95%	97%	97%	95%

black				
justy/black	3%	31%	1%	15%
justy/gray	×	31%	15%	15%

Table Ⅱ. Comparison between various open source OCR engines.
Source: (Andreas Gohr, 2010), Published in [1].

2. The proposed system

The android application has been developed for converting captured text as an image to an audio as an output. The system has been developed using open libraries, engines and translator. The system uses OpenCV image processing library for image preprocessing techniques such as Image Filtering and Noise Reduction, Thresholding etc. For character recognition the system uses Tesseract OCR engine and Yandex translator for translation. For the purpose of giving an audio (i.e speech) as an output, android inbuilt TTS (Text to speech) engine is used.

A. Flowchart of the Proposed System

The flowchart of the proposed system is as shown below:

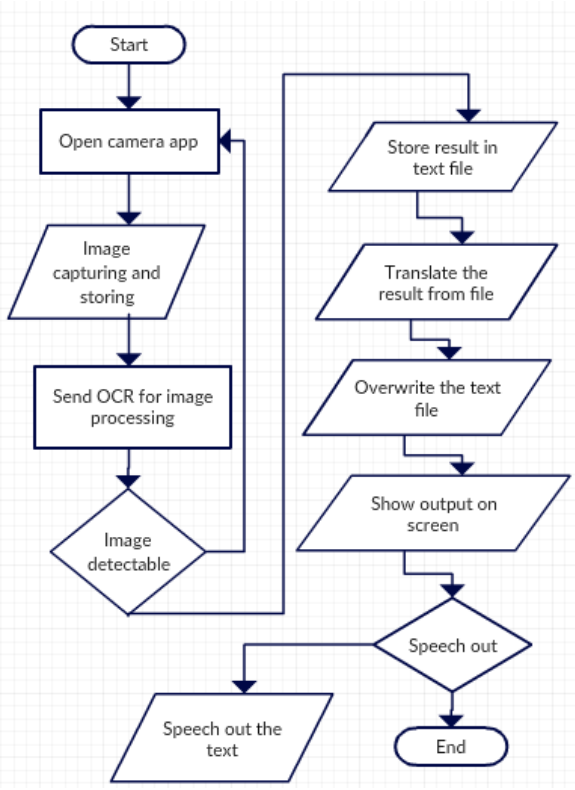


Figure 1. Proposed system Flowchart

B. Image Capturing and Preprocessing

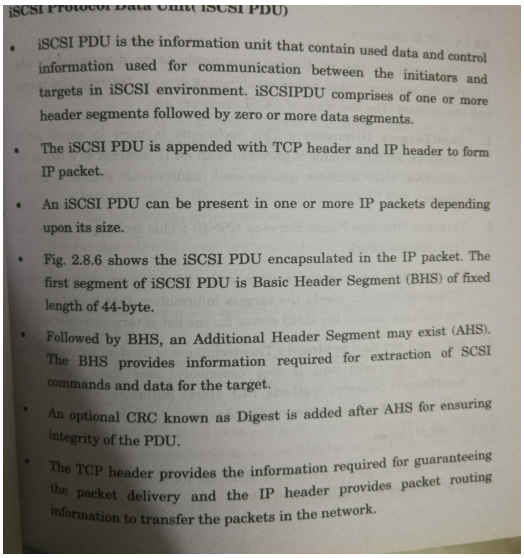
Firstly, images are captured or loaded from mobile device.This images may contain noise which affects output of the OCR. Therefore, there are some techniques such as Image filtering for noise reduction, binarization, text segmentation to be done in the preprocessing phase to improve performance and accuracy of the character recognition system.

1.Image Filtering and Noise Reduction

In image processing, filters are mainly used to suppress either the high frequency noise or the low frequency noise which can be done in both the frequency domain and spatial domain[4].
Frequency domain uses Fourier transform techniques that requires substantial computations therefore, only spatial filtering techniques are used for OCR. Spatial filtering uses a small convolution mask, such as 3x3 matrix[4]. Convolving this mask over an image is much easier and faster than performing Fourier transforms and multiplication[4].
Hence, in the proposed application, two spatial filtering techniques for noise reduction are used such as Gaussian filter and Median filter[4].
OpenCV (Open Source Computer Vision Library) is an open source computer vision and machine learning software library[4]. OpenCV was built to provide a common infrastructure for computer vision applications used to do image preprocessing for OCR. It has C++, Python and Java interfaces and supports Windows, Linux, Mac OS, iOS and Android[4].

2.Binarization or Thresholding

Processing of coloured images needs high computation power as compared to binary (grayscale) images.Hence most of the OCR's need binary image as an input instead of coloured image[4]. The technique used for converting a coloured image to a binary image is called binarization or thresholding[4].
In simple thresholding, If a pixel value is greater than a threshold value, it is assigned one value (white), else it is assigned another value (may be black)[4].
But images generally have different lighting conditions in different areas[4]. Hence instead of global thresholding (fix threshold value for all pixels in an image) adaptive thresholding is used in this application[4].
In adaptive threshold unlike global threshold, the threshold value at each pixel location depends on the neighboring pixel intensities[4]. Threshold value is calculated for smaller regions and therefore, there will be different threshold values for different regions[4].


iSCSI Protocol Data Unit(iSCSI PDU)

- iSCSI PDU is the information unit that contain used data and control information used for communication between the initiators and targets in iSCSI environment. iSCSIPDU comprises of one or more header segments followed by zero or more data segments.
- The iSCSI PDU is appended with TCP header and IP header to form IP packet.
- An iSCSI PDU can be present in one or more IP packets depending upon its size.
- Fig. 2.8.6 shows the iSCSI PDU encapsulated in the IP packet. The first segment of iSCSI PDU is Basic Header Segment (BHS) of fixed length of 44-byte.
- Followed by BHS, an Additional Header Segment may exist (AHS). The BHS provides information required for extraction of SCSI commands and data for the target.
- An optional CRC known as Digest is added after AHS for ensuring integrity of the PDU.
- The TCP header provides the information required for guaranteeing the packet delivery and the IP header provides packet routing information to transfer the packets in the network.

Figure 2. Input Image

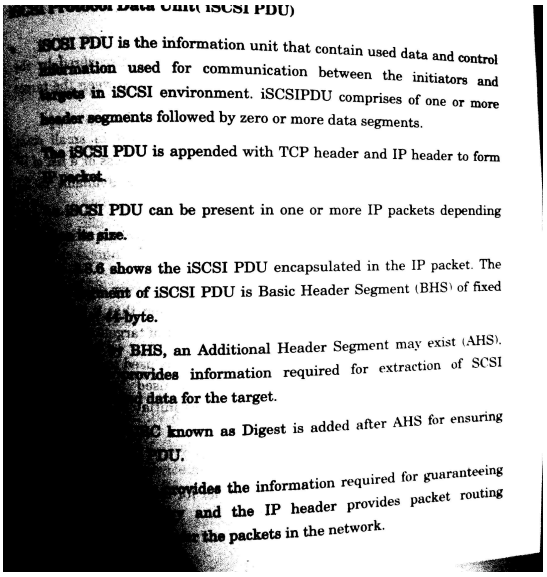


Figure 3. After Fixed Thresholding

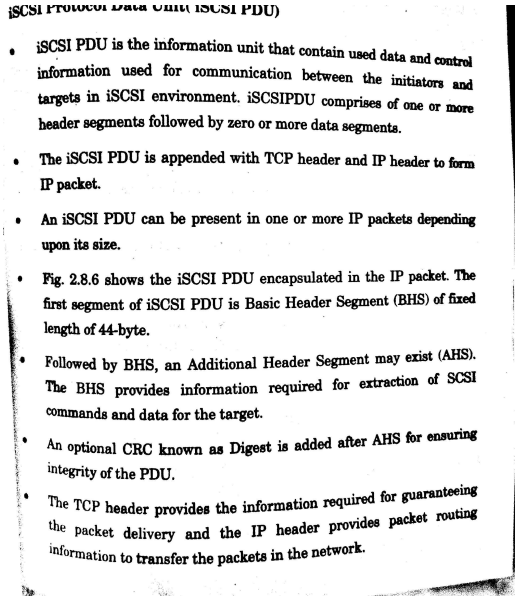


Figure 4. After Local Adaptive Thresholding

Fig.4 shows the final preprocessed image which is passed to OCR engine for text extraction.

C. Translator

There are various translators available in the market such as Google Translator, Bing Translator, etc. Google Translator and Bing Translators are paid API's and provide access to only one user per access key but Yandex translator provides better features and multiple client access using one key.
It also provides synchronized translation for 95 languages, predictive typing, dictionary with transcription, pronunciation and usage examples and many other features.

Figure 5. Captured Image

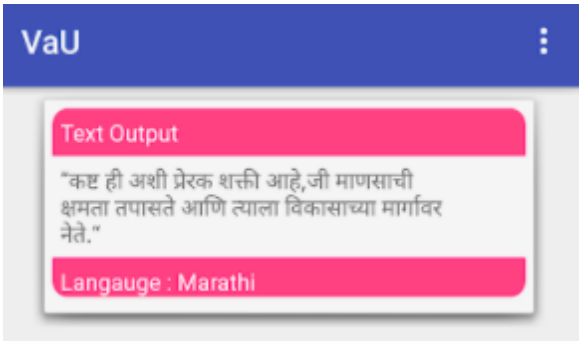


Figure 6. Text Recognized

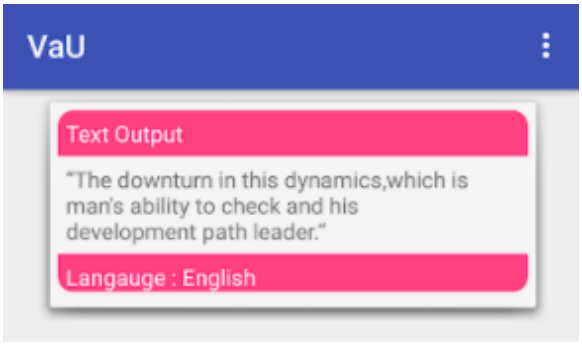


Figure 7. Translated output

Hence, in the proposed system, Yandex translator API has been used for translation from one language to another.

D. Text To Speech

For the purpose of providing audio (i.e speech), as output, the proposed system uses built-in feature i.e the text-to-speech (TTS) engine as the best option provided by google itself, which performs the speech out service of the recognized characters. Android libraries such as android.text and android.speech could also be used. Other available options were espeak, live-text-view and AndroidMary-TTS.

Conclusion

The implemented android application is able to recognize and extract the text in images with a good accuracy Moreover with the use of the android accessibility features such as talkback, the application proves to be useful in helping the blind or visually impaired people in listening to text that they were previously not able to read.
Moreover the use of translator and text to speech feature makes it really easy for the tourists to not only understand but also communicate in different languages with native people.

References

1. Sathiapriya Ramiah,Tan Yu Liong,Manoj Jayabalan 'Detecting Text Based Image With Optical Character Recognition for English Translation and Speech using Android' -2015 IEEE Student Conference on Research and Development (SCOReD)
2. Teddy Mantoro,Abdul Muis Sobri,Wendi Usino 'Optical Character Recognition (OCR)Performance in Server-based Mobile Environment'-2013 international conference on advanced computer science applications.
3. Derek Ma,Qiuhau Lin,Tong Zhang 'Mobile Camera Based Text Detection and Translation'
4. OpenCV Documentation-https://docs.opencv.org/2.4/doc/tutorials/tutorials.html