

# Drashti- An Android Reading Aid

Vikram Shirol<sup>1</sup>, Abhijit M<sup>2</sup>, Savitri A<sup>3</sup>, Pooja R<sup>4</sup>, Amrutha K<sup>5</sup>

<sup>1</sup>Associate Professor, Dept. of CSE, SKSVMACET Laxmeshwar, Karnataka

<sup>2</sup>Bachelor of Engineering, Dept of CSE, SKSVMACET Laxmeshwar, Karnataka

**Abstract-** The challenges faced by blind people in their everyday life are not well understood. They find difficulty in reading or understanding the places. It is difficult for them to know which shop board is in front of them and they have to depend on others to know about these things. To overcome the difficulties and dependencies of people we introduced an application "DRASHTT". The application that attempts to identify and interpret the text or characters displayed on the screen/board and this is been re-presented to the user in the form of voice/speech. The basic aim is to provide an easier life for blind people. The Image is converted to text by using OCR technique. Edge detection and Segmentation Algorithms are used to detect the text regions from image. The algorithms are used to an image set with different text size, font styles and text language. And TTS method is used to convert the Text to Voice, which is made Audible.

**Index Terms-** Edge detection, Image Processing, OCR technique, Segmentation Algorithms, Text Extraction, Text-To-Speech.

## I. INTRODUCTION

The next generation open operating systems are not on desktops or mainframes but on the small mobile devices that the people shall carry every day. The openness of environments leads to new applications and markets and enables greater integration. As the demand grows each time for mobile phone applications, research in OCR, a technology well developed for scanning/capturing documents/images of the documents, is shifting full focus to recognition of the text embedded in digital photographs. Optical character recognition (OCR) [1] is a powerful tool for bringing information from our analog lives into the increasingly digital world. This technology has seen long usage in building/creating digital libraries, recognizing text from images, understanding hand-written office documents / forms etc.

The proposed application would have image processing suite as well as OCR engine both installed on the smart phone implying no server communication overhead and quick and far better processing thus would not just only tackle the limited processing power and limited memory challenge but also provide improved performance. The Main Challenge is developing an OCR for Mobile/Smart Phones with an improved OCR Algorithm that must be capable enough to handle the tradeoffs between Processing time and Quality of results with proper segmentation issues and noise handling.

## II. OBJECTIVE

The main objective is to provide a platform for blind people to lead an independent life. This Product promotes greater independence by enabling blind peoples to perform

tasks that they were formerly unable to accomplish such as reading any test documents like books, newspapers, articles, notices etc. This application is also aimed toward helping normal people.

- The main objective is to help blind and visually impaired people by hearing them what they cannot see.
- To make their life independent where they depend on others to know even about a single word.
- To motivate blind people who are lagging behind and not able to read text or in terms of education.
- To help people in reading, who can understand the language but cannot read.

## III. MODULAR DESIGN

Modular design is the approach that subdivides a system into smaller parts called modules that can be independently create and then can be used in different systems. A modular design can be characterized by functional partitioning into discrete scalable, reusable modules, and rigorous use of well-defined modular interfaces, that make use of the industry standards and designed for User interfaces. Reduction in cost, and flexibility, modularity also offers other benefits such as augmentation, exclusion etc.

### Module 1:

Capturing the Image, The user Click an image that contains the text content in it. The camera through which the user clicks the image will be of high resolution. Thus the captured image will be high quality, thus the image captured will avoid the distortion to higher extent. The image will be easily transferred for the further processing to the OCR[1] process. The image is further processed to the next module that contains the main part of the application, with convert the image to text format.

### Module 2:

The thus captured image contains less distortion due to the high quality camera used in the application. The received image from the device is been uploaded to the OCR Process, that contains Tesseract engines for recognizing text [1,6]. The image uploaded to the OCR is been processed, for the processing from the image to text. The OCR recognizes the boundaries of the Character. And normalizes the character images to the M\*N and then the template matching takes place between the extracted text and the character template. The image is fully been converted into text now.

### Module 3:

The recognized text is to be converted to the voice, hence the text is been further processed by the TTS i.e. TEXT-TO-SPEECH. A text-to-speech (TTS) system converts normal language text into speech. A text-to-speech system is com-posed of two parts a front-end and backend.

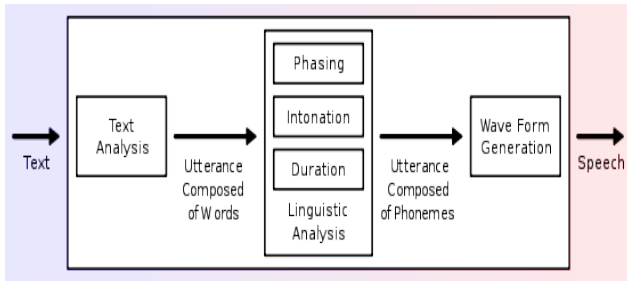


Figure 1: Overview of a TTS System

The frontend has two important tasks. The process converts raw text containing symbols like numbers and abbreviations into equivalent of written-out words. This process can generally called as text normalization, pre-processing. The front-layer or the end then assigns phonetic transcriptions to each and every word, and divides and also marks all accepted the text into prosodic units. The process of assigning phonetic transcriptions to words is called text to phoneme or from grapheme to Phoneme conversions. Mainly the two information i.e Phonetic transcriptions and also Prosody information together make up the symbolic linguistic representation that is output by the frontend. The backend—often referred as synthesizer—then it converts the symbolic linguistic representation into sound using Wave form generator.

Initially the text is been transferred to the Text analysis where the words are been recognized from each of the alphabets, then it is sent for the processing to the Linguistic Analysis, in which the phasing of the text takes place and duration of the utterance of the each word take place. Thus the text utterance composed of the phonemes is been sent to the Wave form generator for the voice transformation.

IV. ARCHITECTURE DESIGN

The Architecture Diagram depicts the overall structure of the software application or the model that is to be created or already created architectural diagram. The Architecture diagram of the application starts with the capturing the image and the uploaded to the OCR engine. Thus uploaded image is been retrieved by the OCR engine.

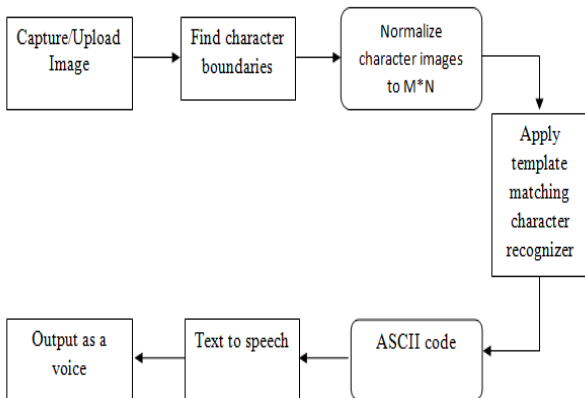


Figure 3: Architecture of Proposed Application

It starts with the finding the boundaries with the image that was uploaded to the OCR engine. Then the process continues with the normalizing the characters

within the image to the  $M * N$ . Then the extracted text is been matched with the character recognizer, will match the text contents with the pattern. The generated ASCII code is transferred to the TTS i.e. Text-to-Speech, that is been converted into the Voice from the Text.

V. FLOW OF DATA IN PROPOSED APPLICATION

Data flow diagrams are graphical tools used to describe and analyze the movement of data through a system. The DFD's also known as bubble chart or data flow graphs or context diagrams. The data flow diagram may be used to present a system or software at any level of abstraction. A fundamental system model or context model represents the entire software elements as a single bubble with input and output indicated by incoming and outgoing arrows respectively.

The following DFD shows the processing of proposed application. We can see the single sequential flow of operation, since there is no involvement of the client and server interaction. The Data Flow diagram shows the simple flow of the process that is going to occur in the specific order.

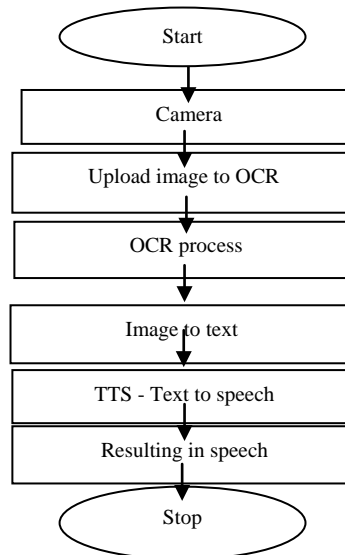


Figure 2: Flow of Data in Proposed Application

One process completes and enters into other process in the sequential flow. The process starts with the capturing image, the camera function is been direct called through the power button or the screen lock button of the respective phone. And hence the camera pops up and the user click the random image and the thus captured image will proceed with the uploading the same image to OCR process.

The OCR receives and accepts the image. Optical Character Recognition is the automated process of translating an input document image into a symbolic text file. The OCR text recognizer will detect the text region in the image through the segmentation algorithm and the Edge detection algorithm will detect the starting of the Text area or the text region. Segmentation extracts lines, words and then finally into characters from the text document images. These methods are classified into dissection, recognition and holistic. The dissection method makes use of the

properties like height, width, spacing etc. The Edge detection is a well developed field on its own within image processing.

Edge detection is basically image segmentation technique, divides spatial domain, on which the image is defined, into meaningful parts or regions. Edges characterize boundaries and are therefore a problem of fundamental importance in image processing. Edges typically occur on the boundary between two different regions in an image. Edge detection allows user to observe those features of an image where there is a more or less abrupt change in gray level or texture indicating the end of one region in the image and the beginning of another. The text thus extracted will be stored in the buffer, and then it is forwarded to the TTS process that converts into voice from the text that was recognized from the image.

Text-to-speech (TTS) technology refers to the combination of text appearing on the display together speaking text aloud with a digitized or synthesized voice. Digitized speech is a recorded (or digitized) human voice speaking, and synthesized voice is a computer-generated voice speaking the text. We can focus on TTS software tools that use synthesized speech to read any text.

#### VI. APPLICATION AND BENEFITS

- **Independent Life:** Blind people can lead an independent life. They don't need to seek help of other for their work.
- **Tries to equalise with normal people:** To Certain extent we can help Blind people to cope up of equalise with the Normal people.
- **Simple User Interface:** The user interface is created easier so that the blind people can interact with the Application.
- **Portable device:** The android device is Portable, hence it is easy to use the Application, since it is not a Desktop application. It is also portable in terms of the platform that is been coded or language.
- **Expense of Application:** It will be a Freeware, so that the Application can be downloaded and used. This Application can be used by the normal people who cannot read but understand the language. We can also learn the Pronunciation of the each word in the written languages.

#### IV. RESULTS

The Results for Blind People's Application are as follows



Figure 4: Splash Screen

The first interface of the application or splash screen that opens up / pop's up when application is open. Blind People press the power button to open the application. This screen is an intimate that an application is been opened with a music or beep sound.

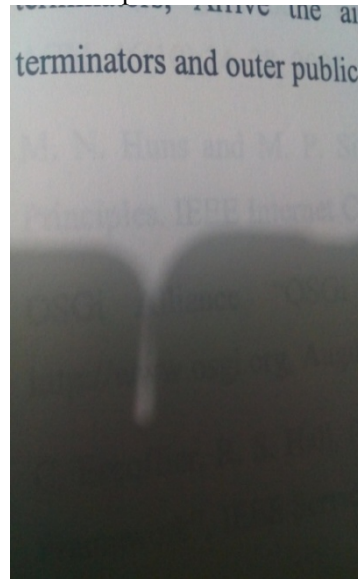


Figure 5: Camera View

Camera pops up as a second interface automatically from Application. Then a user can capture the image and the Application process the captured image to output as voice.

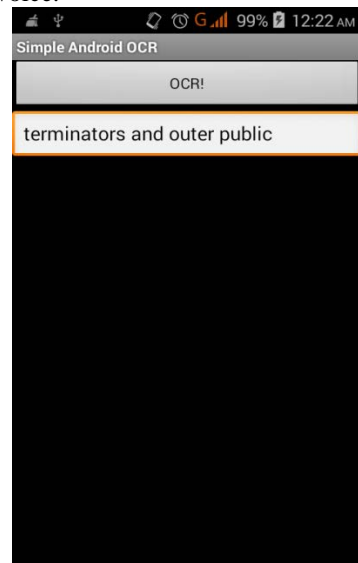
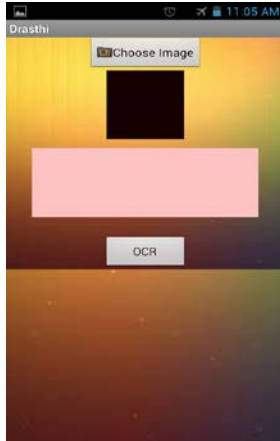


Figure 7: Output Screen

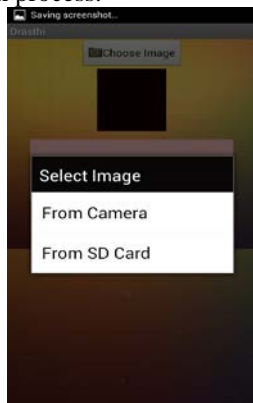
The Application accepts the image from camera and uploads to the Tesseract code and processes the captured images. The text will be displayed under the text field in the 3rd interface of the application.

Results for Normal people application, who doesn't know, how to read but understand the language. For such people we have introduced a new option to the Application of uploading from the gallery. The people who don't have time to read the complete set of pages or documents, they can capture the image and store it in the SD card for future use, later they can use the image from the gallery and read out the images.



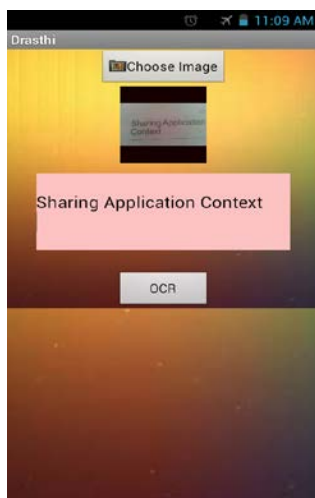
**Figure 8: Interface to Choose Image**

The Application interface for a normal people is as shown above. The choose button gives the option to capture the image and upload option, text field to display the text extracted from an image. The OCR button process the image to the extraction process.



**Figure 9: Options to Choose Image**

This interface opens when the choose button is pressed in the 2nd interface of the Application. This will help to capture an image and produce the voice. And other option is Upload an image from the existing gallery and get output image read.



**Figure 6.11: Output Screen**

When an image is been read from the camera and uploaded image from a gallery is been selected SD card. Press the OCR button to get the voice.

## V. CONCLUSION AND FUTURE WORK

In this project, we reported on blind users' interactions with "Drashti", a smart phone application that enables blind and normal people to capture photographic images using an android phone. The application installed extracts words from image using OCR (Optical Character Reorganization) and converts those words to speech using TTS (Text-to-Speech). This application is mainly focused on blind people who can get the maximum benefit of it. Along with blind people, normal people can also use this application to read stored images. This enhanced application not only consists of capturing images from camera but also provides an option for uploading image from stored images. People who understand language but cannot read can make use of this application.

In future, the application can be enhanced to capture images different languages and to output the voice in one or more languages. The performance can be enhanced using commercial tesseract.

Thus we can conclude that this application is an attempt to help blind and normal people in reading. With this application blindness problem cannot be overcome completely, but can help them at least to some extent.

## REFERENCES

- [1] Rakesh Kumar Mandal, N R Manna, "Hand Written English Character Recognition using Row- wise Segmentation Technique", International Symposium on Devices MEMS, Intelligent Systems & Communication, pp. 5-9, 2011.
- [2] K. Jain and B. Yu. "Automatic Text Location in Images". In Proc. of International Conference of Pattern Recognition (ICPR), Brisbane, pp. 1497-1499, 1998.
- [3] Naveen Sankaran and C.V Jawahar, "Recognition of Printed Devanagari Text Using BLSTM Neural Network", IEEE, 2012.
- [4] N. Senthilkumaran, R. Rajesh, "Edge Detection Techniques for Image Segmentation and A Survey of Soft computing Approaches", International Journal of Recent Trends in Engineering, Vol. 1, No. 2, PP.250-254, May 2009.
- [5] "Text to Speech Conversion System using OCR" Jisha Gopinath, Aravind S, Pooja Chandran, Saranya S S IJETAT, Volume 5, Issue 1, January 2015.
- [6] "Text Recognition from Images: A Review" Pratik Madhukar Manwatkar, Dr. Kavita R. Singh, IJARCSSE, Volume 4, Issue 11, November 2014.
- [7] Low Vision Aids for Reading - AllAboutVision.com.
- [8] Technical Devices And Special Equipment for the Blind- Future Reflections Winter 1986, Vol. 5 No. 1.
- [9] Tesseract OCR Engine Ray Smith, Google Inc OSCON 2007.
- [10] TrainingTesseract3 tesseract-ocr- How to use the tools provided to train Tesseract3 for a new language.

## AUTHORS

**Vikram Shirol:** He has received MTech in Computer science & Engineering from Basaveshwar Engineering College. He is currently working as Associate Professor in SKSVMACET Laxmeshwar Karnataka India. His Major Field of interest is Digital Image Processing, Software Engineering, Computer Graphics, and Advanced Computer Architecture. He has published many of Research Papers in Journals.

**Abhijit Murgod:** He has Received B. E in Computer Science & Engineering from SKSVMACET Laxmeshwar Karnataka India. His Field of interest is Cloud computing, Big Data, Mobile Computing, Operating System, Digital Image Processing, Web Designing. He has 6 months work experience as Web Designer/Developer in IT Software services. Laxmeshwar. Karnataka, India. Her Field of interest is Digital Image Processing, Cryptography, Cloud Computing.

**Amrutha Kanthimath:** She has Received B.E in Computer Science & Engineering from SKSVMACET