

An Optical Character Recognition Framework based Newspaper Reader Application for Blind

Sundar C¹, Abinaya O², Ajithkumar V³
{csr.mca@psgtech.ac.in¹, 18mx202@psgtech.ac.in², 19mx251@psgtech.ac.in¹}

Assistant professor (Sr. Grade), Computer Applications, PSG College of Technology
Coimbatore, India¹, Master of Computer Applications, Computer Applications, PSG College of
Technology, Coimbatore, India², Master of Computer Applications, Computer Applications, PSG
College of Technology, Coimbatore, India¹

Abstract. One of humanity's greatest disadvantages is having a visual handicap and being illiterate or having a learning disability. This paper addresses this shortcoming and gives promise of making newspapers available on daily basis to the blind people using mobile application. There is a demand for an acceptable and easily accessible newspaper reader for the visually impaired society. In this work, newspaper images are converted into audio using in character recognition, artificial intelligence and computer vision. The recognition process is carried-out using python's Optical Character Recognition (OCR) and OPENCV libraries. This system is divided into three modules: image processing for object count, picture cropping, and image reading. A collection of frameworks for OCR and Text to Speech Synthesis (TTS) have been combined, allowing a user to capture a photo with a smart phone and hear the text in the picture.

Keywords: OCR, blind, newspaper, mobile app, reader, TTS, illiterate, python, flask, android.

1 Introduction

Visual tasks are impossible for blind individuals and illiterate to complete. Text reading, for example, necessitates the use of a braille reader or a digital voice synthesiser. The bulk of printed books do not have braille or audio adaptations, and digital versions are still uncommon. Blind individuals, on the other hand, are unable to read newspapers. As a result, the creation of a mobile application that can convert images to voice on a newspaper is essential.

Over the last decade, technology has advanced significantly, with many systems now capable of producing synthetic speech that is very accurate. Speech synthesis research has increased as a result of its growing relevance in a variety of new applications. A mobile application is useful because it allows a blind person to utilise the device camera to read text from a newspaper recorded in the image. The system makes use of current OCR and TTS frameworks, integrating them in such a way that they may deliver the necessary results when used together.

2 Literature Review

All India Radio and News channels are the types of media used in by the blind to hear the news. But currently there is no application which converts the newspaper content into voice so that blind can hear the news.

There is a need to develop mobile application for blind and illiterate people in all segment irrespective of language barrier. This proposed work addresses this with help of python's frameworks for Optical Character Recognition (OCR) and Text to Speech Synthesis (TTS).

It allows a user to take a photo of the newspaper with their mobile and hear the text that is included in the image. The workflow process is shown in figure 1.

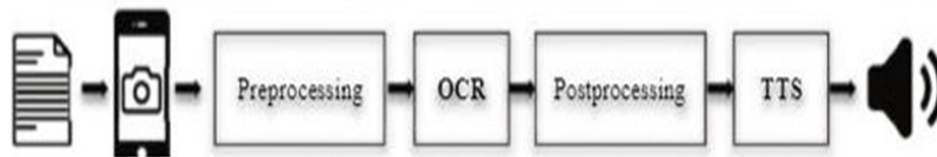


Fig 1. Workflow

3 Technology

This work is carried out with python libraries and Android studio framework. This section exemplifies the tools and technologies

3.1 Python

Python is a portable programming language that may operate on any system. Python supports a wide number of libraries and gives developers access to hundreds of features that help them build successful applications. Python libraries are essential for creating applications in machine learning, data science, data visualization, image and data processing, and more.

3.2 Optical Character Recognition

Extracting text as string values from images is called optical character recognition (OCR) or simply text recognition. Optical character recognition is a method of reading or extracting text from printed or scanned photos, handwritten images, and converting them into a digital format that can be edited and searched. Pytesseract is a wrapper for Tesseract-OCR Engine. Tesseract is a Google-managed FOSS OCR engine.. The following steps are used to implement OCR.

- i. Convert image to Grayscale
- ii. Thresholding is used to convert grayscale images into binary images
- iii. invert the image to using bitwise_not operation
- iv. Applying various noise reduction techniques like eroding, dilating, etc.
- v. Applying the text extraction method to the preprocessed image

Figure 2 shows the conversion of image into text. After that, the final text is transformed into the chosen document format by the user.



Fig 2. OCR conversion

3.4 Flask

It is a web development framework using python as a server side scripting. Web applications created by flask is responsive in nature. Flask supports the features of Unicode, WSGI compliance, URL routing, secure cookies, sessions and request dispatching.

4 Design And Implementation

The interface design depicts how the programme communicates with itself and with the systems with which it interacts.

4.1 Use case Diagram

The purpose Use case diagram illustrates the user's interaction with the system.

In this work, the blind / illiterate is the end user interact with the android application. Figure 3 shows the use case diagram of the application.

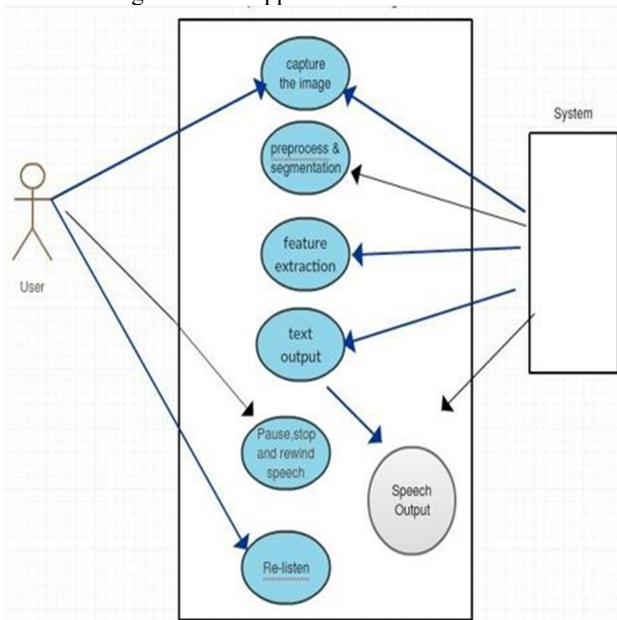


Fig 3. Use case diagram

4.2 Image to Text Conversion

Image to Text Conversion module is broken into two sub-modules that are critical for conversion.

- Capture an image
- Upload an image

Text extraction from a taken image necessitates natural image processing. It starts by finding the text-containing portions of the picture. Following that, Optical Character Recognition (OCR) is used to combine the identified characters into words and phrases. Finally, a text file is generated holding the extracted text.

To improve efficiency, all noise is clipped prior to processing the image. Following that, OCR is applied to this image. The extracted text is subsequently saved in a file in the working directory for further processing. It is important to remember that the algorithms used to analyse natural images are designed in such a way that the quality of the image does not interfere with the precision with which the text is retrieved. The OCR works well enough for photos with text orientations that aren't always straight and poor quality. In an image or a PDF, text can be in any font style and size. The algorithms have been designed to recognise text in a variety of styles and sizes. When it is translated to text format, however, these features are lost since the algorithms are only prepared to read the text and not its font size and font style. The complete logic has been designed to minimise latency and maximise efficiency.

4.3 Text to Speech Conversion

Text retrieved from any type of input is saved in the working directory as a text file. Microsoft supplies a Speech API that is utilised to supply different voices, and an algorithm is created to translate this text into speech. Figure 4 shows the process of conversion.

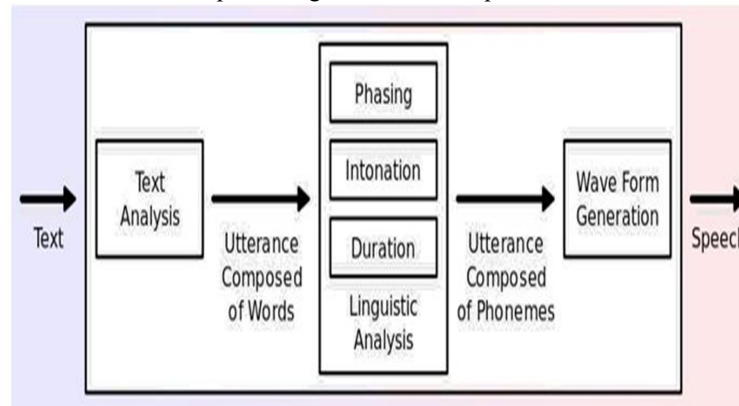


Fig 4. Text to Speech

4.4 User Interface Design

Intro Slider is the first screen in the mobile app that is used to display the highlights of application shown in figure 5.



Fig 5 Intro Screen

Figure 6 shows the camera module part of the application which is used to capture images and shows the generated text output.



Fig 6. Captured Image and text output

5 Testing

A series of testing are performed for this application.

5.1 UI Testing

This testing is carried out for some devices and different OS in emulator. The conventions where as expected. This way of approach ensure that the elements of UI like toolbar, menu , image, list in the application work consistently across all the target platform and the layout , resolution remains as expected. On a whole it provides better usability experience. Table 1 depicts the test case taken for User Interface Testing.

ID	Conditions	Expected results	Actual results	Status
1	Check for the initial display of the page	Page should displayed with background, images, text	Page displays all the images, text and background properly	Pass
2	Check the camera	Application used to take a photo and video	Photo and video captured properly	Pass
3	Text view	To display the converted text in appropriate place	Text displayed properly on the user interface	Pass
4	Audio play	Text to played properly in audio format	Converted text is played properly using speech API	Pass

Table 1. UI test cases

5.2 Performance Testing

The performance of the app depends on the device performance, Server/API performance and network. The app start-up takes some seconds and even though the notifications are send the memory consumption is decreased. The retrieval of the data from the server is handled efficiently where it does not take much in loading the app. In case if the app does not respond, restart the app for effective functionality.

5.3 System Testing

This system generally checks whether the hardware, software, and other third party installed plug-ins work well to the device on which app is installed. The below table 2 illustrates the testing carried out on the application and its result.

Description	Yes /No
Can the app be installed on the device	Yes
Does the app behave as desired if there is an incoming call or SMS	Yes
Does the app behave as desired if there is a charger connected or disconnected	Yes
Does the app behave as desired if the device goes to sleep mode or resume from lock screen?	Yes
Does the app behave as desired if the device is tilted or shaken	Yes

Does the app behave as desired if there is a local message coming from other app?	Yes
Is the functionality of all the buttons or keys on this device defined for this app?	Yes
Does the app navigate through the "back" button available on device	Yes
Do all the touch screen positions (button) working correctly	Yes
Does the application work properly on device rotation	Yes
Does the application shows notifications properly	Yes

Table 2. System testing

6 Conclusion

The technique allows the visually impaired to not feel at a disadvantage while reading literature that is not printed in braille. The picture pre-processing section extracts the needed text region from the complicated backdrop and provides high-quality input to the OCR. The OCR output text is delivered to the TTS engine, which generates the voice output. A battery may be used to power the system to enable for mobility of the gadget. Future development might include creating devices that identify objects and extract text from films rather than static photos.

Acknowledgment

We are thankful to Dr. K. Prakasan, Principal In-Charge, PSG College of Technology, Coimbatore for permitting us to take up this work. We extend our earnest gratitude to Dr A. Chitra, Professor and Head, Department of Computer Applications, PSG College of Technology, Coimbatore, for her support, encouragement and above all her ardent motivation.

References

- [1] Hagargund, Asha G., et al. "Image to speech conversion for visually impaired." *International Journal of Latest Research in Engineering and Technology* 3.06 (2017): 09-15.
- [2] James, S. Aaron, S. Sanjana, and M. Monisha. "OCR based automatic book reader for the visually impaired using Raspberry PI." *International Journal of Innovative Research in Computer and Communication Engineering* 4.7 (2016): 1111-1118.
- [3] Liambas, Christos, and Miltiadis Saratzidis. "Autonomous OCR dictating system for blind people." *2016 IEEE Global Humanitarian Technology Conference (GHTC)*. IEEE, 2016.
- [4] https://en.wikipedia.org/wiki/Visual_impairment