

# Bangla Optical Character Recognition and Text-to-Speech Conversion using Raspberry Pi

Aditya Rajbongshi<sup>1</sup>, Md. Ibadul Islam<sup>2</sup>, Md. Mahbubur Rahman<sup>4</sup>, Anup Majumder<sup>5</sup>, Dr. Md. Ezharul Islam<sup>6</sup>  
Department of Computer Science and Engineering  
Jahangirnagar University, Savar, Dhaka, Bangladesh

Al Amin Biswas<sup>3</sup>  
Department of Computer Science and Engineering  
Daffodil International University  
Dhaka, Bangladesh

**Abstract**—Optical Character Recognition (OCR) technology is very helpful for visually impaired or illiterate persons who are unable to read text documents but need to reach the content of the text documents. In this paper, a camera-based assistive device is used that can be applied for visually impaired or illiterate people to understand Bangla text documents by listening to the contents of the Bangla text images. This work mainly involves the extraction of the Bangla text from the Bangla text image and converts the extracted text to speech. This work has been fulfilled with Raspberry Pi and a camera module by applying the concepts of the Tesseract OCR engine, the Open Source Computer Vision, and the Google Speech Application Program Interface. This work can help people speaking Bangla language who are unable to read or have a significant loss of visual sight.

**Keywords**—Optical character recognition; Bangla text; speech conversion; Raspberry Pi; camera module

## I. INTRODUCTION

Text and speech is the primary medium for communication among human beings. For accessing the text information, an individual needs visual sight. Besides the visual sights, an individual can also know the information using their listening capability. According to the World Health Organization, the amount of visually impaired people is 285 million and the amount of blind people is 39 million in the whole world [1]. More than 90 percent of the visually impaired people exist in developing countries [2] and on the other hand, according to UNESCO [3], 27.11% of adults are illiterate. All these facts have raised the importance of research to develop systems that can help visually impaired persons to overcome their limitations.

Raspberry Pi [4] is a single-board computer. For exploring computing and learning various programs such as python, scratch, etc. this device helps people a lot. It has the capability of calculating, playing music, gaming, and other functions that are done by a computer. The main advantages of this device are portability and low cost. For experimental and innovation activities this board is designed. Its two types of model differentiate each other based on the USB port. For the above-mentioned features, Raspberry Pi has become an essential and ideal tool for IoT and automation research.

In the recent era, OCR [5] has been used for converting images to text. It helps millions of people to know the information from scripts such as airline tickets, medical documents, mail, etc. in their perspective file. In the recent advancement of OCR technology and algorithms such as the Tesseract OCR engine [6], it can recognize a huge number of characters in various languages. The application of OCR touches every technological organization in the world. It also included the recognition of characters from handwriting scripts in various languages.

Bengali or Bangla is an Indo-Aryan language. As a primary or secondary language, around 210 million people speak in Bengali, among them around 100 million, and 85 million speakers are from Bangladesh and India, respectively [7]. Bangla OCR is different from other languages because of the basic structure of the Bengali script. Bengali letters as in Fig. 1 have different transformers and edges. Besides, a large number of characters are contained in the Bengali script. There are 57 characters in Bengali scripts among them 21 characters are a vowel and 36 characters are consonants. Because of curves in the character, sliding, and stroke characters researchers face various challenges.

In this paper, we propose a Bengali OCR based system. The image which is captured by the user using the camera included in the proposed device is analyzed in various phases of Tesseract (OCR Engine) methods. The text in the input image is extracted using the method of Open Source Computer Vision (OpenCV). For converting the text into a speech, GTTs (Google Text to Speech) library is used and it works offline. In Raspberry Pi board, a slot is used for connecting the headphone and the user needs to be connected with the headset. Thus, the user can hear the speech of the text through the headphone.

This work is ordered as follows: Section II outlines the existing relevant work. Then, Section III explains the working methodology with the system architecture that includes system hardware and software implementation. System evaluation is demonstrated in Section IV. Subsequently, the conclusion of the paper by mentioning some future works is appended in Section V.



b) Pi camera: The Pi camera module has the excellency of taking pictures and recording videos in Raspberry Pi. In Raspberry Pi, there is a port to connect the Pi Camera. There are various versions of the Pi camera. Most of them can deliver a clear image. In this device, Raspberry Pi Camera V2 is used for capturing images. It has 5MP Resolution and supports 1080p30 video recording.

c) Memory (Storage): There is no hard disk or solid-state drive in the device. Instead of this, a micro SD card is used for booting the Linux kernel-based OS (operating system). In Raspberry pi, there is a slot for inserting a micro SD card. This card is also used for the data storage of capturing images. When the image is captured, it is stored on the card. The analyzing image is also stored in the memory card. The specification of the used hardware components is provided in Table I.

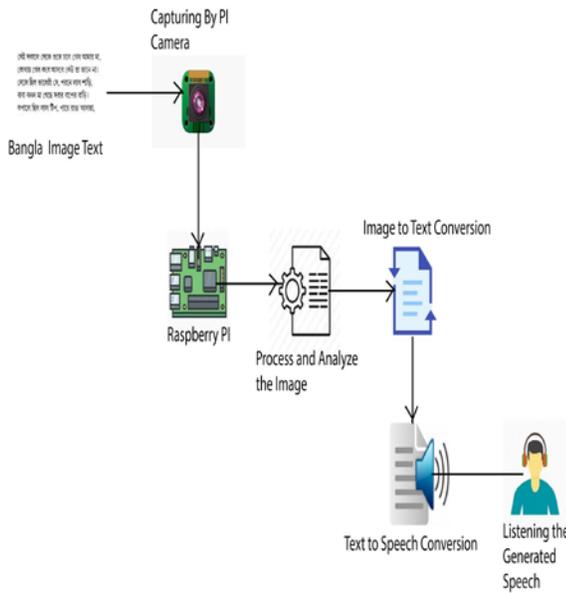


Fig. 2. Proposed System Architecture.

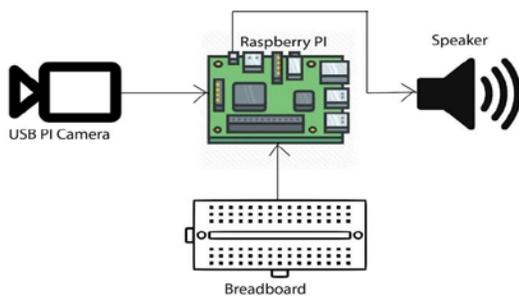


Fig. 3. System Hardware Organization.

TABLE I. SPECIFICATION OF HARDWARE COMPONENT

Component Name	Version/Storage
Raspberry Pi	3, B model
Pi Camera	V2 model
Breadboard	PMSA003
Memory	Samsung 16GB micro SD

d) Camera enabling: In Raspberry Pi, there are three ways to enable the Pi camera. One is to set manually and another one is to use the command line and also in python code. It's better to enable through the command line. Without enabling a Camera, it doesn't work. After enabling the camera, its configuration automatically adjusts to the Raspberry Pi.

e) Push-button setup in breadboard: A breadboard is a board of rectangular plastic having multiple tiny holes. In this project, a push-button is used to capture the image which is integrated with the breadboard. When the push button is hit, the image is captured. To set up the push button in the breadboard, the circuit diagram of Fig. 4 is used. Fig. 5 shows the system setup after connecting all of the hardware components.

B. System Software Implementation

In a system software implementation, we have installed the Raspbian Operating System [20] in a memory card and inserted it in the raspberry pi. After installing Raspbian OS, we have also installed various tools, packages, and libraries. Table II shows the details of some software components.

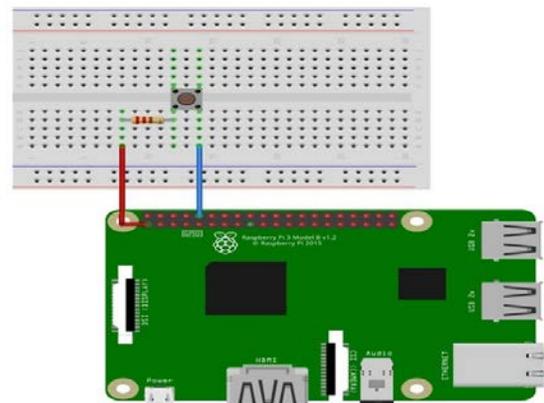


Fig. 4. Circuit Diagram for Push-Button Setup.



Fig. 5. System Setup after Connecting Hardware Components.

TABLE II. SPECIFICATION OF SOFTWARE COMPONENT

Software Component Name	Version
Raspbian OS	4.19
Python	3.6.0
Tesseract	3.05.00
OpenCV	CV2

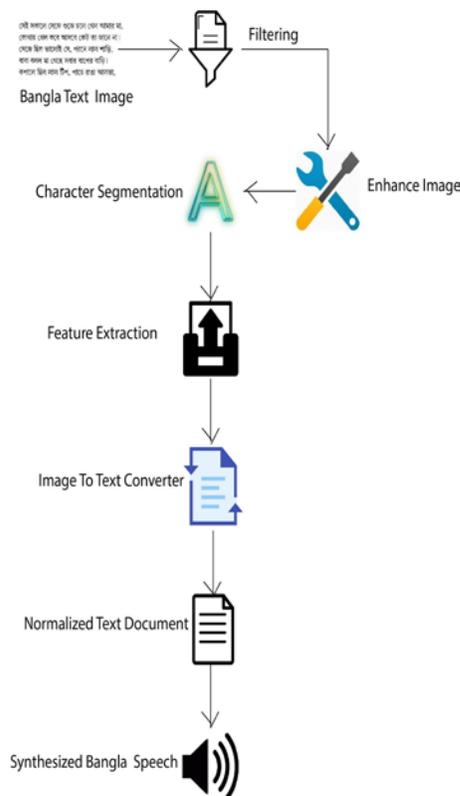


Fig. 6. System Operational Architecture.

The captured image from the Raspberry Pi camera is stored in a file as a JPEG format. The stored image is analyzed in python script through the following steps as sketched in Fig. 6.

The details of the above steps are described below:

a) Acquisition of image: In this step, a Bangla text image is captured using the Pi camera. Captured Bangla text image is sent to the preprocessing step where various unwanted noise is reduced.

b) Preprocessing of image: By applying relevant morphological transformation like dilation, back hat transformation, threshold, producing the necessary contours, discrete cosine transformations, and forming bounding box, the unwanted noise in the image is banished in the preprocessing of image. First of all, the captured image is rescaled to the relevant size and then converted into a grayscale image for further processing. The grayscale image is then compressed using the discrete cosine transformation. The compressed image is very helpful for further processing. In the compressed image, there exist various unwanted high-frequency components. Those components are omitted by setting the vertical and horizontal ratio. For decompression of the image, the inverse discrete cosine transform is applied. There are two operations like back top-hat transformation and dilation which are used in the image. Then the operation named black top-hot transformation is employed in the image. This operation helps to extract the object or elements that are smaller than the defined. After this operation, the dilation

operation is applied for adding the pixel to the edges of the object of the image. The number of pixels depends on the shape and size of the present object. Now the thresholding algorithm is applied in the present image. Among all the thresholding algorithms, here adaptive thresholding is chosen. Then using specific functions of OpenCV, the contours of the image are generated. There are many bounding boxes of the objects or elements in the present image. For drawing those, the generated contours are used. For extracting every character of the present image, the drawn bounding boxes are used. Finally, by applying the OCR (Optical Character Recognition) engine, the full text of the present image is detected.

c) Extraction of text: From the input image, the recognized text is extracted in this step. This extraction is performed using the Tesseract OCR engine.

d) Text to speech converter: Applying the GTTs engine, the extracted text is converted to speech in this step. With the help of some predefined libraries of this engine, we performed the text to speech conversions. In the GTTs engine, there are online and offline systems. In our project, we have used an offline system for user portability.

e) Desired output speech: When speech is generated, the user can easily hear it through the headphones. As it is based on a Bangla text image, the user can easily hear the speech in Bangla language.

#### IV. SYSTEM EVALUATION

For the system analysis, there needs to be some Bangla scripts paper. A monitor is included for proper monitoring of recognition of text from the captured image in the system. The Fig. 7 shows the final view of our system.

For the result, an image is captured using the Raspberry Pi camera. It is considered as the input image. Image preprocessing, extraction of texts are handled by the various defined methods of the tesseract (OCR engine). The accuracy of the extraction of text in captured images is not 100% because of the mid resolution pi camera. After extraction of text, a user can easily hear the speech through a headphone that is connected to the earphone slot in Raspberry Pi. For the precise image, the accuracy of text extraction is satisfactory. Sometimes the accuracy of text extraction is 100% for some precise images. Fig. 8, 9, 10, and 11 show the input image, the Bangla text extraction from the input image, the texts displayed on the screen, and a user of hearing speech, respectively.



Fig. 7. The System Connecting with Monitor.

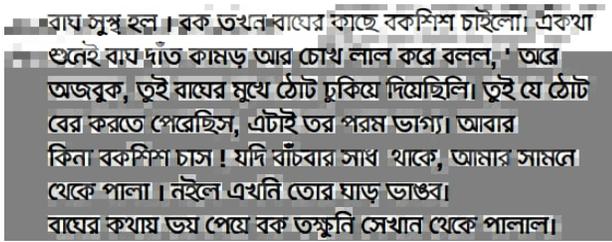


Fig. 8. Input Image.

বাঘ সুস্থ হল। বক তখন বাঘের কাছে বকশিশ চাইলো। একথা শুনেই বাঘ দাঁত কামড় আর চোখ লাল করে বলল, 'অরে অজবুক, তুই বাঘের মুখে চোঁট ঢুকিয়ে দিয়েছিলি। তুই যে চোঁট বের করতে পেরেছিলিস, এটাই তর পরম ভাগ্য। আবার কিনা বকশিশ চাস! যদি বাঁচবার সাধ থাকে, আমার সামনে থেকে পাল। নইলে এখনি তোর ঘাড় ভাঙব। বাঘের কথায় ভয় পেয়ে বক তক্ষুনি সেখান থেকে পালাল।

Fig. 9. Extracted Text from the Input Image.

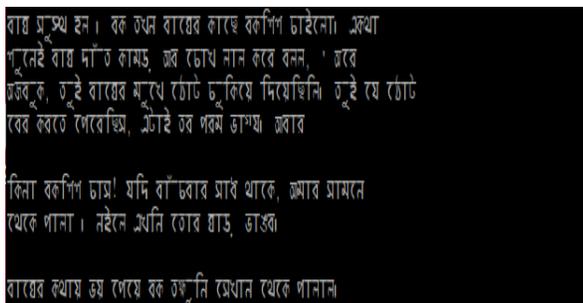


Fig. 10. Extracted Text Displayed on the Monitor.



Fig. 11. Outcome Observation via Headphone.

## V. CONCLUSION AND FUTURE WORKS

This research uses raspberry pi, pi camera, Tesseract OCR engine, etc. to help the people listening to the content of the Bangla text image who are visually impaired or illiterate. It can also be used by any person who wants to listen to the content of the image instead of reading the content of the image. We have achieved 97.4% accuracy for precise Bangla text images. For the middle range of a Pi camera, the quality of the captured image is not so good in low light. During the night the quality of the captured image is obscure. As a result, sometimes the accuracy of the extraction of text is not up to mark.

In the future, we would like to enhance this system by appending the higher resolution Pi camera to increase accuracy for text extraction and by eliminating noise from speech using advanced algorithms. Furthermore, we would like to improve the portability of the system by compacting the hardware design through design improvement and hardware upgrade. We

would also like to extend our research to extract text on Bangla handwritten script.

## REFERENCES

- [1] "Globaldata on Visual Impairment," Available Online: <https://www.who.int/blindness/publications/globaldata/en/> [Last Access: 20 May 2020].
- [2] "WHO |Sight savers International," Available Online: [https://www.who.int/workforcealliance/members\\_partners/member\\_list/si/en/](https://www.who.int/workforcealliance/members_partners/member_list/si/en/) [Last Access: 20 May 2020].
- [3] "Bangladesh Literacy Rate," Available Online: <https://countryeconomy.com/demography/literacy-rate/bangladesh> [Last Access: 20 May 2020].
- [4] J. Arthur, Raspberry Pi: The complete guide to Raspberry Pi for beginners, including projects, tips, tricks, and programming. CreateSpace Independent Publishing Platform, 2017.
- [5] S. Mori, H. Nishida, and H. Yamada, Optical character recognition, John Wiley & Sons, Inc., 1999.
- [6] C. Patel, A. Patel, and D. Patel, "Optical character recognition by open source OCR tool tesseract: A case study," International Journal of Computer Applications, vol. 55, no. 10, pp. 50-56, 2012.
- [7] "Bengali Language," Available Online: <https://www.britannica.com/topic/Bengali-language> [Last Access: 20 May 2020].
- [8] R. Naveen, S. A. Sivakumar, M. Cholavendan, S. S. Kumar, A. S. K. Singh, S. Thilak, and C. Arunkumar, "Face and text recognition for the visually impaired persons based on Raspberry Pi," International Journal of Advanced Engineering and Research Development, vol. 5, no. 03, pp. 90-94, 2018.
- [9] L. Nagaraja, R. S. Nagarjun, M. Nishanth, D. Nithin, and V. S. Murthy, "Vision based text recognition using Raspberry Pi," National Conference on Power Systems & Industrial Automation (NCPISA 2015), 1-3, 2015.
- [10] V. A. Devi and S. S. Baboo, "Embedded optical character recognition on Tamil text image using Raspberry Pi," International Journal of Computer Science Trends and Technology (IJCTST), vol. 2, no. 4, 2014.
- [11] G. K. Sagar, and T. Shreekanth, "Real time implementation of optical character recognition based TTS system using Raspberry Pi," International Journal of Advanced Research in Computer Science and Software Engineering, vol. 7, no. 7, pp. 149-156, 2017.
- [12] S. A. James, S. Sanjana, M. Monisha, "OCR based automatic book reader for the visually impaired using Raspberry Pi," International Journal of Innovative Research in Computer and Communication Engineering, vol. 4, no. 7, pp. 1111-1118, 2016.
- [13] C. A. Todd, A. Rounaq, U. K. Nur, and F. Boufarhat, "An audio haptic tool for visually impaired web users," Journal of Emerging Trends in Computing and Information Sciences, vol. 3, no. 8, pp. 1136-1145, 2012.
- [14] D. Velmurugan, Srilakshmi, S. Umamaheswari, S. Parthasarathy, and K. R. Arun, "Hardware implementation of the smart reader for visually impaired people using Raspberry Pi," International Journal of Advanced Research in Electrical Electronics and Instrumentation Engineering, vol. 5, no. 3, pp. 2055-2063, 2016.
- [15] C. S. T. Thu and T. Zin, "Implementation of text to speech conversion," International Journal of Engineering Research & Technology (IJERT), vol. 3, no. 3, pp. 911-915, 2014.
- [16] H. M. Htun, T. Zin, H. M. Tun, "Text to speech conversion using different speech synthesis," International Journal of Scientific & Technology Research, vol. 4, no. 7, pp. 104-108, 2015.
- [17] A. Goel, A. Sehrawat, A. Patil, P. Chougule, and S. Khatavkar, "Raspberry Pi based reader for blind people," International Research Journal of Engineering and Technology, vol. 5, no. 6, pp. 1639-1642, 2018.
- [18] H. Rithika and B. N. Santhoshi, "Image text to speech conversion in the desired language by translating with Raspberry Pi," 2016 IEEE International Conference on Computational Intelligence and Computing Research, IEEE, 2016.
- [19] S. Gollapudi, Learn computer vision using OpenCV, Apress, 2019.
- [20] W. Harrington, Learning raspbian, Packt Publishing Ltd, 2015.