ISSN : 2347 - 2243

# Indo - American Journal of Life Sciences and Biotechnology

www.iajlb.com
Email : editor@iajlb.com or iajlb.editor@gamil.com

ISSN 2347-2243 www.iajlb.com
Vol.19, Issue.4, Oct 2022

Vol.19, Issue.3, July 2022

Indo-Am. J. of LifeSc & Bt.,2022

Indo-American Journal of Life Sciences and Biotechnology

# Text to Audio Conversion System with Efficient Portable Camera for Visually Impaired People

**Dr.AR.Sivakumaran**,

**Abstract**-- It's a lot of work to accurately recognize text from a picture and turn it into audio. Our suggested system analyzes and compares the many methods utilized for text extraction from color photos, such as text detection and identification. Multiple subtasks, such as text detection, text localization, text classification, text segmentation, and text recognition, make up this overall assignment. Transcribing the data included in these photos into English will make using the data more efficient and convenient. Text extraction is the technique of removing text from a picture. Information retrieval, keyword searching, editing, documenting, archiving, and reporting all use text extraction in one way or another. However, poor picture contrast and complex backgrounds, as well as differences in text size, demanding subject to solve. Character properties of fonts used in texts and picture quality provide further difficulties. Because of these obstacles, computers can't read the characters perfectly and identify them. Using Python 2.7.15, Optical Character Recognition (OCR) technology, and an audio converter, we have created a system that can extract text from photos and play it back..

**Index Terms**— Optical Character Recognition, Text Detection, Localization, Classification, and Segmentation (OCR).

## I. INTRODUCTION

TODAY, there are 285 million persons in the globe who are visually impaired. Rapid growth is being seen in this area as a result of a rise in the incidence of inherited genetic disorders. Those individuals rely only on a guiding cane in order to read any kind of text, including those found in ads, nature scenes, handwritten materials, etc. With this in mind, our suggested approach is geared at assisting users in hearing the image's accompanying text as an audio recording. Someone does not need to be around all the time to help a blind person. The suggested system is employed by

Professor, Department of Information Technology, Malla Reddy Engineering College for Women, Secunderabad, Telangana,India

.

**The Proposed System: An Overview (Part II)**

The challenges of guiding the blind and visually handicapped are constant and significant. Here, we want to record a picture of some text. Text extraction from a picture requires a series of steps to be conducted once the image has been acquired. At first, the picture is filtered, and any unwanted noise is eliminated, in what are called "pre-processing" processes. After filtering, a thresholding operation is conducted to transform the picture to grayscale. The aforementioned procedures are performed to improve and extract data from photos that may be used in further processing. Next, the text is processed, and the result is transformed into an audio file that can be played via headphones.

In our suggested system, we show how OCR may be used in Python. Heuristic-based techniques, including leveraging gradient information or grouping the text into paragraphs which will appear on a straight line, may be used to detect text in controlled, confined contexts. Tesseract is an OCR engine that can read text from noisy photos on a number of different operating systems. An audio converter is used to turn the read text into an audible format.

Architecture of Systems III

From a scanned document, a photo of the document, a scene-photo (for instance, the text on signs and billboards in a landscape photo), or subtitle text superimposed on an image, optical character recognition (OCR) is the mechanical or electronic conversion of images of typed, handwritten, or printed text into machine-encoded text (for example from a television broadcast).

The resulting text is then recast as a spoken word.

Figure 1 depicts the proposed system architecture, and Figure 2 is a schematic of the proposed system.

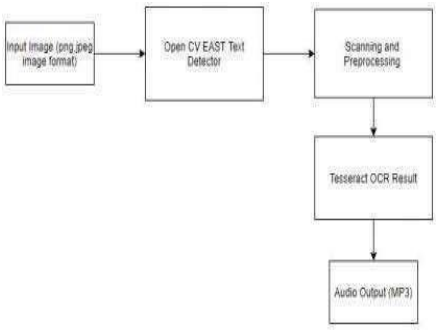



**Figure 1: Proposed System Architecture Image to Text Recognition**

To successfully extract text properties, one must have prior knowledge of certain text qualities. Thus, a first step is to investigate textual

characteristics. The geometry, size, color, motion, edge, and compression of text are all features that may be studied separately. During the text recognition process, raw text is broken down into individual words and characters. Since the word is the most fundamental entity that people employ for visual recognition, it is crucial that pictures of text be converted into words. Character recognition and word recognition are two distinct methods of recognition. Some techniques of character identification work by chopping up a picture of text into many individual characters. An key component of these techniques is a separator for neighboring characters. Word recognition is the process of extracting words from a text or picture using character recognition outputs and language models or lexicons. In the event that characters are downgraded, this may be utilized. Optical character recognition (OCR) is a method of reading text by analyzing a picture of the text. The Python-based program called "python-tesseract" is an OCR (optical character recognition) application. The text in photos will be "read" by the system.

**Figure 2: Schematic Diagram of the Proposed Systems**

Changing Text into Sound

The output of the text areas is informative words thanks to OCR's text recognition. These contrasted text sections are sent as audio signals via the audio jack, amplified, and shown to the consumers. The acquired picture then undergoes a text to string transformation using the Tesseract speech engine. Next, the converted text is converted into an audio file. It works for texts that can be read clearly even when seen via a camera.

**II. HOW THE SYSTEM IS INTENDED TO WORK**

Object area recognition, text localisation using pre-processing methods, text extraction, and text-to-speech conversion are the four components that make up the suggested OCR approach.Detecting the Physical Boundaries of an Object We use a camera with a rather wide angle in our prototype to guarantee that the item in hand is visible to the camera. The Logitech C270 webcam is more than enough for your purposes. However, this might have the unintended consequence of bringing into the camera's field of vision additional, although perhaps textual, items. such as when a user is grocery shopping. Extraction is challenging for texts of any size. It's likely that the quality of the text might suffer if it were enlarged to a larger size. Improved pixel quality from a camera may be exploited for mass manufacturing. Due of its ability to record concentrated text that yields clear results.

Object-based text localization

We have employed many methods, such as morphological procedures and image processing, to localize text on an object. Once the foreground is blurred, obtaining the Region of Interest is simple. Thresholding is applied to this fuzzy picture, resulting in a binary representation in which the foreground is made up of all black pixels and the background of light ones. The first order derivative of a picture may be calculated from its gradient. It will keep track of any lateral movement. Other thresholding methods exist, such as adaptive thresholding and Otsu's thresholding, in addition to the conventional kind. Otsu's thresholding requires a uniform distribution of foreground and background pixel values. Otsu's thresholding was disregarded since its findings were imprecise. In OpenCV, you may boost the efficiency by increasing the number of repetitions of these morphological procedures to get more accurate results.

Taken from the original text

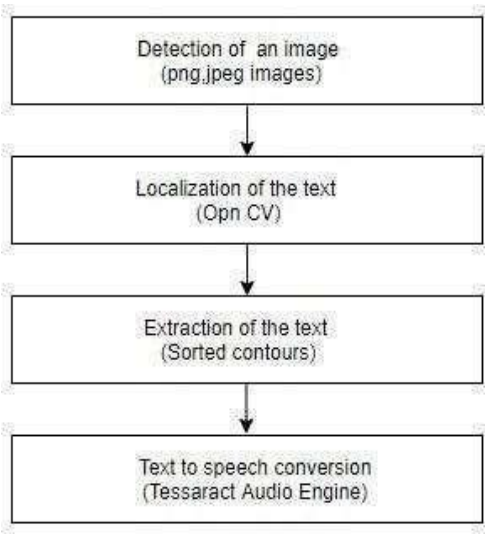
By using the Sorted Contour Method, a square bounding box has been generated. An image's text coordinates are retrieved during a pre-processing procedure, and after the width and height of the contour have been determined, a bounding box may be calculated using these values in addition to the image's dimensions.

Transforming Text

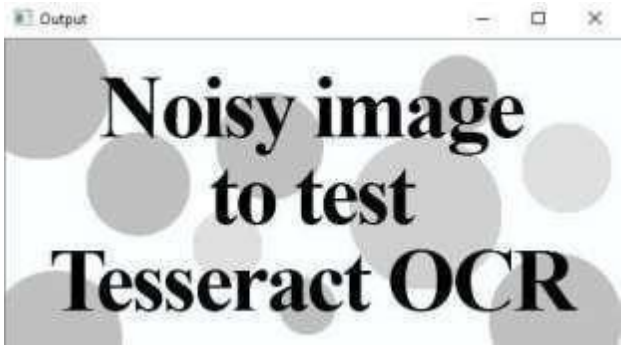
The acquired picture then undergoes a text to string conversion using the Tesseract audio engine. Numerous tools convert written material into spoken language. The pyttsx3 library is used to convert text to speech. It has been noted that the aforementioned approach only yields results within specified bounds. Texts legible via a camera, such as those set in Arial, qualify for this rule. The device also stops working if lightning strikes nearby.

**Automatic Detection of Text and Playback of Audio**

The product's text is detected and rendered into audio form after extraction from the visual backdrop. Here, a little camera attached to a pair of eyeglasses will be used to snap a picture of a product or any other image with writing on it. Text extraction from a picture requires a series of steps to be conducted once the image has been acquired. In the first stage of processing, the picture is filtered and noise is removed. We next do thresholding operations on the grayscale picture that we've created after filtering. At this point, segmentation is used to split the next frame (i.e., divide the characters). Character analysis using OCR (optical character recognition) is used to compare written text to a preexisting database. Now that they've been identified, these symbols may be turned into audible output.




making it simpler for them to go about their regular routine. It's also utilized for the purpose of signature-reading technology




in cheques. Text recognition is also used in automated document scanning.

Figure 3: Test image for Optical Character Recognition

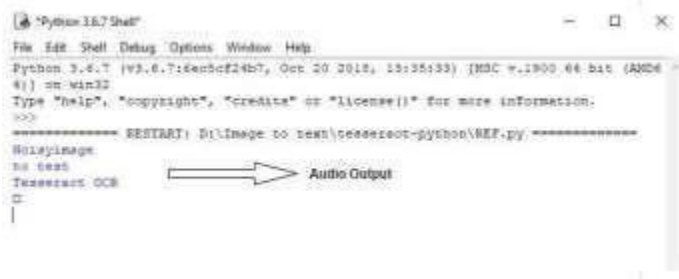



**Figure 4: Final Result**

## II. RESULTS AND DISCUSSION

It is recommended to transform the original colored picture into a gray image so that morphological procedures may be performed with less difficulty. The inverted binary conversion results are shown in Figure 3. Because of this, the image's backdrop is removed, leaving us with clean text for processing.

The OCR method's test picture is shown in Figure 3. The camera's output is the picture shown in the illustration. In Figure 5, we see the highlighted end result of text recognition.

Here, a system test picture is recorded, compared to a database of template photos, and the end result is the product name, as seen highlighted in Figure 4.

## III. APPLICATIONS

In the business world, our suggested technology may be put to use for text detection and identification tasks like scanning barcodes and product labels. It may be used to find particular text on the web, such as video subtitles or article abstracts. It is also used for reading street signs in the event of unmanned vehicles and automated license plate readers at toll booths. When it comes to helping the visually handicapped read, text detection and identification has a crucial use case.

A Final Thought and Suggestions for the Future

Our suggested system, a portable camera-based assistive text reader, is an aid that takes object data from its surroundings, pre-processes the picture, and then identifies what it sees, allowing the visually handicapped to hear what others can see. To help the targeted users read text labels and product packaging from photos of hand-held products in their everyday lives, we present a camera-based assistive text reading framework that uses the extracted text. All texts with font sizes of 1 inch or higher on simple backgrounds provide positive results when processed with OCR.

Our system currently struggles to recognize texts that are small in size, have complicated backgrounds, or are written in a language other than English, but we want to apply more advanced algorithms in the future to remedy this. Finding text from ancient scripts and turning it to audio is another area where our

suggested approach has room for improvement. We can also do future work using e-handwritten and elaborate typeface content.

## III. REFERENCES


[1] Text Detection and Recognition in Imagery: A Survey, Qixiang Ye and David Doermann, IEEE Transactions on Pattern Analysis and Machine Intelligence, 2014.

[2] Using NLP, K. Elagouni, C. Garcia, and P. Sbillot wrote "A Comprehensive Neural-Based Approach for Text Recognition in Videos," which was published in the proceedings of the 2011 ACM conference on multimedia retrieval.

[3] This is supported by the following reference: [3] "Detecting Texts of Arbitrary Orientations in Natural Images," C. Yao, X. Zhang, X. Bai, W. Liu, Y. Ma, and Z. Tu, in Proc. IEEE Int'l Conf. Computer Vision and Pattern Recognition, pp.1083-1090, 2012.

[4] In 2010, at the European Conference on Computer Vision, K. Wang and S. Belongie published "Word Spotting in the Wild" on pages 591 and 604.

[5] On Combining Multiple Segmentations in Scene Text Recognition, L. Neumann and J. Matas, in Proceedings of the 2013 IEEE International Conference on Document Analysis and Recognition, pages 523–528, 2013.

[6] Pattern Recognition, volume 31, issue 12, pages 2055-2076, 1998; A.K. Jain and B. Yu, "Automatic Text Location in Images and Video Frames."

[7] "A Cascade Detector for Text Detection in Natural Scene Images," by S.M. Hanif, L. Prevost, and P.A. Negri; published in Proceedings of the IEEE International Conference on Pattern Recognition; pages 1–4, 2008.

[8] Scene text detection via connected component clustering and nontext filtering. IEEE Transactions on Image Processing, volume 22, issue 6, pages 2296-2305, 2013. [8] H. Koo and D.H. Kim.

[9] A Laplacian Approach to Multi-Oriented Text Detection in Video, P. Shivakumara, T.Q. Phan, and C.L. Tan, IEEE Transactions on Pattern Analysis and Machine Intelligence, volume 33, issue 2, pages 414–419.

[10] According to "AdaBoost for Text Detection in Natural Scene," written by J. Lee, P. Lee, S. Lee, A. Yuille, and C. Koch, and published in Proc. IEEE Int'l Conf. Document Analysis and Recognition, pages 429-434, 2011, you may get a more thorough explanation of this technique.

[11] To cite this paper: A. Coates, B. Carpenter, C. Case, S. Satheesh, B. Suresh.

[12] Text Detection and Character Recognition in Scene Images with Unsupervised Feature Learning, T. Wang, D. J. Wu, and Andrew Y. Ng, Proceedings of the 2011 IEEE International Conference on Document Analysis and Recognition, pages 440-445.

[13] Localizing and Segmenting Text in Images and Videos, by R. Lienhart and A. Wernicke, IEEE Trans. Circuits System on Video Technology, 12(4):257–268 (2002).

[14] An Adaptive Text Detection Approach in Images and Video Frames," by M. Li and C. Wang in Proceedings of the 2008 International Joint Conference on the Neural Network, pages 72–77.

[15] IEEE Transactions on Image Processing, Volume 9, Issue 2, Pages 147–156, 2000, Authors: H. Li, D. Doermann, and O. Kia, "Automatic Text Detection and Tracking in Digital Video."

[16] Image text detection using a bandlet-based edge detector and stroke width transform. A. Mosleh, N. Bouguila, and A. Ben Hamza. 2012. Proc. British Machine Vision Conference, pp. 1-2.

[17] "A New Approach for Overlay Text Detection and Extraction from Complex Video Scene," by W. Kim and C. Kim, IEEE Transactions on Image Processing, volume 18, issue 2, pages 401-411, 2009.

[18] To read more about this topic, check out the following article: [17] "Toward Integrated Scene Text Reading," by J.J. Weinman, Z. Butler, D. Knoll, and J.J. Feild, published in the IEEE Transactions on Pattern Analysis and Machine Intelligence, volume 3, issue 2, pages 375–387, 2014.

[19] To wit: "A Gradient Vector Flow-Based Method for Video Character Segmentation," by T. Phan, P. Shivakumara, B. Su, and C.L. Tan, published in Proceedings of the 2011 IEEE International Conference on Document Analysis and Recognition, pages 1024-1028.

[20] According to "A Novel Adaptive Morphological Approach for" by S. Nomura, K. Yamanak, O. Katai, H. Kawakami, and T. Shiose.

[21] Pattern Recognition, volume 38, issue 11, pages 1961-1975, 2005, "Degraded Character Image Segmentation."

[22] According to "Text Detection and Recognition in Images and Video Frames," by D. Chen, J.M. Odobez, and H. Bourlard, published in Pattern Recognition, volume 37, issue 3, pages596-608.