REST Journal on Emerging trends in Modelling and Manufacturing

Vol: 9(2), June 2023

REST Publisher; ISSN: 2455-4537 (Online)

Website: https://restpublisher.com/journals/jemm/

DOI: https://doi.org/10.46632/jemm/9/2/3

Image-To-Text Conversion with Text-To-Speech Technology in Android

***Ms. S. S. Kiruthika, Neelakantam Guna, G Akhil Reddy, S. Adnan Haseeb**

SRM Institute of Science and Technology, Chennai, India.

*Corresponding Author Email: kiruthis3@srmist.edu.in

Abstract: *One to use as Automatic text detection The process of conversion is crucial to the eventual success of multimedia content retrieval. In this work, we offer an algorithm that can automatically recognise horizontally aligned text in photos and then isolate it from its surrounding context, regardless of how complicated the backdrop may be. The suggested method relies on the use of a color-reduction methodology, an edge recognition method, and the positioning of text sections via studies of projection profiles and geometrical features. As well we are using some files with the help of deep machine level-based technique. And we are detecting text from the real time video also. The text is embedded in a wide variety of papers and natural settings, and it offers information that is both crucial and beneficial. It's not easy to pull text from a natural picture. We recognise and extract the text in a manner that makes it easily readable by a human. Here, we offer a quick approach to text localization by using an edge detection and clustering approach to pinpoint all potential picture edges.*

1. INTRODUCTION

Every day, people must deal with hunks of data. These days, it's common practise to send files like scanned papers or photos instead of text. Data supplied in the form of photos often has to be edited before being delivered. As a result, we need programmes that can transform these files into more usable formats. Therefore, it is necessary to be able to recognise texts from photographs of printed or typed text. The goal of this study is to find ways to digitise photographs of printed text so that they may be used in word processors. The main goal is to identify individual characters in photographs of handwritten or typed text. These days, a wide range of methods and algorithms may be used to achieve the same results. One of the most used methods for identifying text is optical character recognition. It's a strategy for changing static pictures of printed or manually written text into dynamic, machine-meaningful person streams that might be modified for extra purposes. It is much simpler to complete the necessary steps after transforming a printed or composed text picture into a machine-intelligible configuration. You may type in a term or phrase, get relevant results, alter the text, copy and paste it into an email, post it on your website, etc. There is a wealth of literature in this field to draw on. Otsu's methodology for picture division and the Hough change for slant identification are only two examples of the methods that have been utilised to recognise characters. Character recognition makes use of pattern recognition methods. Germany awarded a patent for optical character recognition to Gaurav Tauschek in 1923. In 1933, Paul W. Handel was granted a patent in the United States after him. A patent in the United States was later awarded to Tauschek as well. Tauschek's device included a photo detector and pattern templates. In 1949, RCA engineers developed the first rudimentary OCR system. The goal was to aid the US Veterans Administration's visually challenged clients. However, their gadget did more than just translate human-written text into computer code. The letters were then spoken by it as well. Due to the prohibitive price tag, this was shelved after initial testing. There is a lot of literature on the topic of document image defects models. Sarkar et al. utilized a ten-boundary picture debasement model to construe a portion of the known material science of machine printing and imaging of text. This document image decoding (DID) model suggests it is possible to improve accuracy while decreasing the amount of time spent on tedious human labour. Even in the worst cases of visual deterioration, this is assured. After extensive testing with a wide variety of photos, they found that even severely deteriorated photographs still yielded 99% accuracy. This has allowed DID to be utilized on inferior quality photographs without the requirement for manual division, thus reducing the amount of human labour required. Enhanced document images are required to raise the reliability of

the suggested approach. Canon et al. developed an apparatus to fix and improve photographs. The system's in-house label was QARC. The visuals were of actual typewritten manuscripts. Using quality metrics and picture restoration methods, they automated the process of enhancing document scans for OCR processing. They employed metrics like the We've got the "font size factor," the "small speckle factor," the "touching character factor," the "white speckle factor," and the "broken character factor."Using the document's picture as input, the algorithm was able to characterise it based on the quality metrics. Then, a linear classifier that had already been trained was used to create a filter for restoring images. It was stated that the character mistake rate would be reduced by 38% what's more, the word blunder rate by 24% thanks to this suggested approach. In a later piece of software, Image Refiner, Summers et al. used a very similar approach. They made enhancements to the system by taking into account non-Latin scripts and non-fixed fonts, while QUARC only dealt with Latin scripts. To QUARC's already extensive set of quality metrics, we added measurements of length, breadth, depth, and aspect ratio. Image Refiner, like QUARC, uses document image quality measurements to choose the best approach to restoration, but it uses an error propagation neural network as its classifier instead of support vector machines. It is crucial that text lines be correctly recognised while working with text recognition. Many text line finding strategies have been developed. Breuel presented an algorithm that would first determine the correct page rotation and then pick out the lines of text. He employed the branch and bound method of optimum line finding. It can be implemented quickly and just requires five inputs. A technique was presented to identify the desired text line inside an image acquired by a portable pen scanner. Experiments were conducted on a testing information base comprising of 117 record pictures to validate and test the suggested technique. The outcomes met with my approval. In this research, we offer an algorithm for efficiently transforming scanned text pictures into a more suitable format for further editing. In order to effectively identify and recognise characters, the suggested algorithm makes use of modified versions of methods like Otsu's approach and the Hough transform. The suggested algorithm used very basic methods, but after some tweaking, it performed well even when subjected to rotation and scaling. English letters (a-z, A-Z) and numbers (0-9) are processed by the suggested algorithm.

2. LITERATURE SURVEY

Extracting useful textual information[1] from a set of photos is the focus of the field known as "text extraction from image." Content extraction from pictures has gotten a great deal of consideration in late examination in the space of picture handling. Objects, colour, texture, shape, and even their connections to one another may all serve as content. picture indexing and categorization, as well as content-based picture retrieval, may benefit from the semantic information they give. The difficulty[2] of identifying the relevant text area is made more difficult by the fact that text data might be contained in a picture in a variety of font styles, sizes, orientations, colours, and against a complicated backdrop. There is no way to extract text from a complicated or textured backdrop using current Optical Character Recognition (OCR) methods. By "text detection," we mean the process of identifying individual photographs within a sequence in which text appears. Finding where text appears in a picture and creating bounding boxes around it is called text localization. To accelerate the limitation cycle and keep the text in the same relative location between frames, text tracking is employed. Bounding boxes may show where text is in a picture, but it still has to be separated from the backdrop before it can be recognised. Therefore, before to being given into an OCR engine, the retrieved text image must be transformed into a binary image and improved. In the text extraction process, the textual elements are isolated from the surrounding context. Since the text area often has poor goal and is defenseless to commotion, text improvement of the gathered text components is necessary. After text has been retrieved from photos, OCR technology[4] may be used to convert the images into text. Images of text recorded by a camera include curved text lines because of page twist and the camera's perspective. Therefore, it is essential that the text be aligned correctly and straight so that it may be easily scanned. However, de-wrapping systems have a hard time with text lines segmentation for curved text. In this examination, we offer a methodology to character segmentation and extraction from curved text lines in document pictures that is based on image processing methods. By using an x-line and a base line, the algorithm[5] is able to segment curved text. Bounding boxes are then drawn around each word in the document picture that has been determined to contain text. Words are broken apart using the qualities of their related parts. We are introduced to a novel technique for extracting information from unconstrained handwritten manuscripts via statistical shallow parsing. Hidden Markov Models treat a line of text as a unit independent of its parts. By focusing on individual lines of text, shallow parsing may rapidly extract the data that matters from any document while ignoring the rest. There are two types of knowledge that are necessary for this.

1. HMM (Hidden Markov Models using the Baum-Welch Algorithm) character models.
2. use probabilities to smoothly move between them.

The RecogniseP[6] Step Text lines are deciphered in terms of the model during the recognition stage. According to this paper, in the current context, text extraction from filtered reports and text pictures is a crucial undertaking. For this purpose, we use a technology called optical character recognition (OCR). The scanned page serves as the

input for the proposed technique[7], which uses Otsu's calculation for division and the Hough change strategy for slant recognition to extract the text from the picture. Only the English letters (A-Z, a-z) and numbers (0-9) were recognised by the system. Character recognition through OCR technology has been deployed. Images of scanned documents and screenshots of written words from online sources were used in the validation testing. The suggested technique[8] has been tested and shown to be effective in recognising letter sets written in Verdana textual style, size 14, and has shown phenomenal outcomes with pivoted photographs. With an average accuracy of 93%, the suggested system is able to correctly recognise characters from text pictures. It also performed well when tested against rotation and scaling, and it successfully reduced picture noise. It was determined that 90% was the average rotation accuracy for properly rectifying skew from photos. In this method[9], Separate the areas of interest from the background using a technique that uses the inclination dim level to part the 2D histogram locale and afterward plays out the traditional Otsu's thresholding approach two times on two projection histograms. The experimental findings verify that the approach is computationally efficient and resistant against noise. Thresholding[10] is a technique used to separate areas of interest in a picture from the background depending on the grayscale of the original image. Several thresholding techniques are thoroughly detailed and ranked using several error metric types. By maximizing the 'between-class variance,' as one popular thresholding approach does, Otsu's method finds the best possible threshold. However, since a single-dimensional histogram of an image does not capture the spatial information between picture pixels, achieving adequate results may be challenging when images include noise. To find the best threshold vector[11], Lui et al. extended the use of a 1D-histogram to a 2D-histogram, which considers the dim level dispersion of pixels as well as the normal dark level conveyance of their neighborhood. Although it takes more time to run, this approach outperforms 1D Otsu's in terms of thresholding accuracy. According[12] to the proposed technique, object pixels, background pixels, edge pixels, and noise pixels each have their own distinct qualities. The original grey levels of the object's inner pixels are rather near to those of the object's average grey level, and the same is true of the interior pixels of the background. However, the original grey levels of the pixels used to depict borders and sounds are significantly different from the average grey levels. To show the differences between the shades of grey, they used a gradient grayscale. The gradient magnitude[13] values in the 2D histogram are 0 at the diagonal and grow at large distances from the diagonal, as shown by the coordinates (i, j). Similar to how tiny gradient magnitude values tend to show on pixels portraying the article and foundation, large angle size values tend to depict edges and noise. As a result, we may roughly assign a chance of 1 to classes that are close to the diagonal and a probability of 0 to classes that are distant from it. It is evident that the classic 2D Otsu method's region separation does not adequately divide the regions. By taking use of these features, we may segment the 2D histogram into areas that identify the object, background, edges, and noise. Two hundred genuine photos from the Segmentation evaluation database were used to put the approaches to the test; the ground truth for each image was available. The experimental findings demonstrate the method's effectiveness in producing desirable outcomes, its resistance to noise, and its need for less computing time than competing approaches producing equivalent outcomes. Researchers in this paper, have the difficult task of finding an effective way to filter out unwanted background noise in photos before using them in further analyses. When an image is being captured or transmitted, noise may significantly affect the quality of the picture. An picture's noise is first and foremost dealt with before any other image processing techniques are used. The many kinds of noise in a picture need different kinds of noise reduction methods. The best outcomes are achieved when the testing image model conforms to the presumptions. In this study, we shed some light on certain crucial forms of noise and conduct a comparative examination of methods for suppressing noise. Different forms of noise are applied to a picture model, and the impacts of different sound decrease approaches are explored in this work. The degree[14] to which a picture is corrupted by noise depends on the nature of the disturbance. Both linear and non-linear techniques may be used to remove noise. While linear approaches are quick, they sacrifice picture quality in favour of speed, non-linear methods maintain image quality. Mean filters, medial filters, order statistics filters, and adaptive filters are all distinct varieties. The most effective filter for picture de-noising is identified using all of the available filters. Based on their experiments, they found that BM3D and median filters gave good results/ Character recognition is performed by feeding the probable target text line into an OCR engine, which is based on a score derived from geometric features. The acceptance of the target text line depends on the accuracy of the recognition result. If none of the target text lines can be found, the leftover message lines are shipped off the OCR motor to check whether any of them can be found. Probes a testing data set of 117 report pictures filtered with C-Pen and ScanEye pens verify the efficacy of the aforementioned strategy. They ran a benchmark test to see how well the strategy worked for extracting target lines. [15]In early benchmark testing, its efficacy was established. It is pointed out that their method involves adjusting a number of control settings. Using a consistent set of values for these regulating factors yields the same outcomes throughout. Space constraints prevent them from providing further detail on how they settle on certain values for these knobs. The responsiveness of the control settings on the presentation of line acknowledgment and rubbish picture dismissal may also be further studied with additional data. In this paper, we present a mechanized archive section framework that can distinguish what kind of document a picture is and then extract textual information like titles and authors from those photos.

The system stores documents in a database, makes it simple to get the files you need for your regular workflow, does layout analyses using document-specific models, and takes for granted that you already know what each document type is. In this study, we explore a system for determining a document's classification.

3. EXISTING METHOD

Text is one of the most important ways information is transported in today's digital world. To pass on cultural norms and values from one generation to the next, text may be seen as a system of symbols. One of the most consequential innovations in human history, texts are fundamental to the survival of our species. Every aspect of our lives is surrounded by text. Therefore, it is essential to get information from literature that is present in everyday life circumstances. There are two main tasks involved in processing the text that occurs in a natural scenario: text identification what's more, text affirmation. Text acknowledgment alludes to the course of classifying the characters from a scene's picture that form expressive words, while Text detection refers to the approach of extracting text bodies from natural scenes. Astute examination, blind route, modern mechanization, car help, robot route, quick interpretation, independent driving, item search, and so on are just some of the numerous sectors that may benefit from scene text recognition. That's why, in recent years, research into street view text detection and identification has exploded. As a preliminary start, this research explores techniques for detecting text in photographs of natural scenes.

4. PROPOSED METHOD

1. It's app-based The phrase "image pre-processing" is used to describe the most fundamental picture manipulations.
2. Using entropy as a metric, we see that these manipulations have no positive effect on the information content of the picture.
3. Applications
4. A system for automatically recognising signs
5. Vision-based computer processing

There is a visual representation of the suggested algorithm. Pre-processing is a multi-step procedure applied to each of the input photos. First, we take the picture from the source and convert it to grayscale using the global OTSU technique, so the image is black and white. The results of the first process might be contaminated by extraneous data or distortion in the second. As a result, we use image de- noising methods to get rid of this kind of disturbance. Step three involves creating a histogram for the noise-free picture, which aids with text extraction. Once the picture has been segmented along its borders, the OCR software Tesseract can determine whether or not a given line contains text by comparing it to a database of known text strings. Every snippet of text is taken directly from the picture. After that, an algorithm is used to put the text into a structure that humans can comprehend.

1. The basic component of the proposed system for text extraction from photos to text using text-to-speech in Android is a set of OCR libraries, and the use of Machine Learning (ML) methods is recommended to increase the accuracy and efficiency of these libraries.
2. To enhance the precision of word extraction from photos, we will train ML algorithms on a huge dataset of images and their accompanying text, using methods like element extraction and profound learning approaches like Convolutional Brain Organizations (CNNs).
3. The system will include a user interface that allows for the quick and simple submission of photos for text extraction, as well as language selection and customization choices, and voice output settings.

5. SYSTEM ARCHITECTURE

To begin using the Android app, the user inputs an image which can either be selected from the device's gallery or taken as a new picture using the device's camera. Next, the image goes through a pre-processing stage, which involves resizing it to a standard size, converting it to grayscale, and enhancing the contrast to improve OCR accuracy. Following this, any unwanted noise is removed from the image by eliminating stray marks and lines, sharpening the edges of characters, and smoothing out any imperfections. OCR technology is used in the next step to extract text from the pre-processed image, which is then stored as a digital file. Feature files are created from the extracted text and image, which are analyzed in the following step to identify relevant keywords and features. Natural language processing (NLP) techniques are employed to identify the different parts of speech like nouns, verbs, adjectives, and more, which are then used to generate speech output.

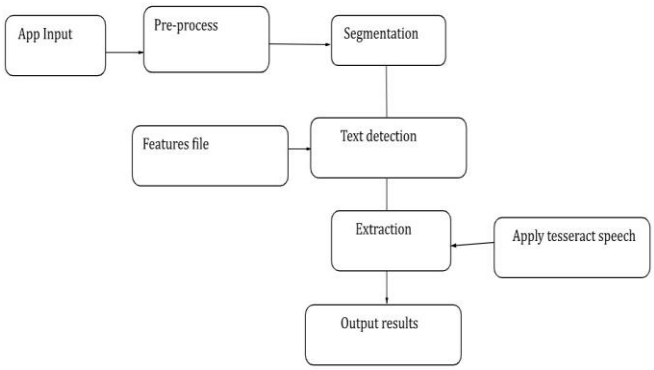

FIGURE 1. Architecture of Conversion of Image

1) App input: The user starts by inputting an image into the Android app. This can be done by selecting an image from the device's gallery or by using the device's camera to take a new picture.
2) Preprocess: Once the input image is received, it needs to be preprocessed to prepare it for feature and text extraction. This step can involve resizing the image to a standard size, converting it to grayscale, and enhancing the contrast to improve the accuracy of the OCR process.
3) Remove unwanted noise: The next step is to remove any unwanted noise from the image. This can include removing any stray marks or lines, enhancing the edges of characters, and smoothing out any imperfections.
4) Text extraction: Once the image has been preprocessed, OCR technology is used to extract text from the image. OCR is a process that uses computer algorithms to recognize and extract text from images. The text extracted from the image is then stored as a digital file.
5) Feature files: In this step, feature files are created from the extracted text and image. These feature files can be used for subsequent analysis and processing, such as identifying relevant keywords and features for generating speech output.
6) Extraction: The extracted text is analyzed to identify relevant keywords and features. This can involve using natural language processing (NLP) techniques to identify nouns, verbs, adjectives, and other parts of speech. These features can then be used to generate speech output that accurately reflects the content of the original image.
7) Apply text to speech: The relevant keywords and features identified in the previous step are used to generate speech output using text-to-speech technology. This involves converting the digital text file into speech that can be played back through the device's speakers.
8) Output result: The final output is presented to the user in the form of speech output or displayed text. The user can then choose to listen to the generated speech or read the extracted text on the device's screen.

Overall, this system architecture involves a series of complex steps that work together to extract features and text from an image and generate speech output that accurately reflects the content of the image. By processing images in this way, the system can provide a valuable tool for users who need to extract information from images and generate speech output on their Android devices.

6. ALGORITHM USED

Future Works: The study identified the limitations of existing OCR and machine learning algorithms and suggests optimizing them to improve accuracy and efficiency. The study evaluated the effectiveness of a specific text-to-speech technology on Android devices. The precision of the retrieved text and the quality of the produced speech may both be enhanced with the help of natural language processing methods.

REFERENCES

[1]. Shejwal, M. A., & Bharkad, S. D. (2017), “Segmentation and extraction of text from curved text lines using image processing approach”. 2017 International Conference on Information, Communication, Instrumentation and Control.

[2]. Thomas, S., Chatelain, C., Heutte, L., & Paquet, T. (2010), “An Information Extraction Model for Unconstrained Handwritten Documents”. 2010 20th International Conference on Pattern Recognition

[3]. Kochi, T., & Saitoh, T. (1999), “User-defined template for identifying document type and extracting information from document”s. Proceedings of the Fifth International Conference on Document Analysis and Recognition. ICDAR ’99.
[4]. Agrawal, N., & Kaur, A. (2018), “An Algorithmic Approach for Text Recognition from Printed/Typed Text Images”. 2018 8th International Conference on Cloud Computing, Data Science & Engineering (Confluence
[5]. S. P. Chowdhury, S. Mandal, A. K. Das,“Automated Segmentation of Math-Zones from Document Images”
[6]. Zhen-Long BAI and Qiang HUO, “An Approach to Extracting the Target Text Line from a Document Image Captured by a Pen Scanner.”
[7]. K.N. Natei, J. Viradiya, S. Sasikumar, “Extracting Text from Image Document and Displaying Its Related Information.”
[8]. Chandan Singha,Nitin Bhatiab,Amandeep Kaurc, “Hough transform based fast skew detection and accurate skew correction methods.”
[9]. Chinnasamy, Sathiyaraj, M. Ramachandran, and Vidhya Prasanth. "Recent Advances in Selection Techniques for Image Processing." Electrical and Automation Engineering 1, no. 2 (2022): 98-105.
[10]. Puthipong Sthitpattanapongsa and Thitiwan Srinark, “A Two-stage Otsu’s Thresholding Based Method on a 2D Histogram.”
[11]. U. Bhattacharya, S. K. Parui and S. Mondal, “Devanagari and Bangla Text Extraction from Natural Scene Image”.
[12]. "Real-Time Image Captioning and Text-to-Speech Conversion for Visually Impaired People," by A. Hussain, A. M. Khan, and A. Khan. (2021)
[13]. "An Image-to-Text Conversion Approach Using Deep Learning for Visually Impaired Individuals," by R. Agrawal, S. K. Jain, and R. Kumar. (2020)
[14]. "Real-Time Image Captioning and Speech Synthesis for Mobile Applications," by C. Chi, C. Lee, and J. T. Tsai. (2019)
[15]. "An Android Application for Image Captioning and Text-to-Speech Conversion for the Blind," by A. A. Khanna and K. K. Sharma. (2019)
[16]. "Image Captioning and Text-to-Speech Conversion for Visually Impaired Individuals Using Deep Learning Techniques," by P. B. Jadhav, M. V. Shinde, and A. S. Khobragade. (2019)

