International Journal of Engineering and Advanced Technology (IJEAT)
ISSN: 2249-8958 (Online), Volume-8 Issue-5, June 2019

Handwritten English Character Recognition and Speech Synthesis to aid Text-to-Speech Applications

Akshay S, Chandana J R, Namita G

Abstract: ***Handwritten character recognition is principal area for research. This domain is widely challenging. It has various applications like helping the visually impaired and in government offices where there is a need to recognize handwritten documents. Through this paper we propose a system which recognizes characters and converts them into voice output using text-to-speech system. Datasets of uppercase letters, lowercase letters and digits have been used as templates for matching with the input image. every input image of the character that we have considered is of the standard size 42x24 pixels. Our proposed system manages to produce good recognition rates and successful voice output.***

Index Terms*: *Handwritten Character Recognition, image processing, OCR, text-to-speech.

I. INTRODUCTION

Character recognition is a technique which recognizes written or printed letters or numbers and converts them to digital text. Our project is based on OCR. The function of OCR is to read the characters from the scanned document or image, process it and extract the feature and recognizing the pattern. Before using OCR first we should do the preprocessing on the image. Preprocessing includes noise removal, compression and thresholding to make the image easy to operate on. Segmentation and feature extraction are the crucial operations in character recognition. In Segmentation, the image is divided into many segments. In our project, segmentation is done using regionprops and boundingbox functions of MATLAB. This helps to segment each character in the image. Templates have been created for different characters to enhance the recognition of characters using OCR. The recognized character will be converted to a .txt format file. Abbreviated as TTS, Text-to-Speech conversion means converting the text into a spoken voice output. We have implemented TTS using Microsoft .NET framework and assembly, which includes the namespace System.Speech.Synthesis. When a new speech synthesizer object is created, it uses the default system voice. Our project is useful in assisting blind and visually impaired people. It helps in data entry tasks where handwritten documents can be computerized. We have the following sections: 2. Papers Reviewed, 3. Proposed Method, 4. Experimental Results, 5. Conclusion and 6. Future work.

Manuscript published on 30 June 2019.
***** Correspondence Author (s)
Akshay S, Department of Computer Science, Amrita Vishwa Vidyapeetham, Amrita School of Arts and Sciences, Mysuru, Karnataka, India.
Chandana J R, Department of Computer Science, Amrita Vishwa Vidyapeetham, Amrita School of Arts and Sciences, Mysuru, Karnataka, India.
Namita G, Department of Computer Science, Amrita Vishwa Vidyapeetham, Amrita School of Arts and Sciences, Mysuru, Karnataka, India.

II. LITERATURE SURVEY

[1] Raspberry Pi 3 motherboard with camera and audio jack is used in this project. They have used OCR system. Google Speech API and Microsoft Translator translates to desired user language. [2] Pattern recognition, its scope and its applications have been emphasized in this paper. They have also discussed about NLP and its architecture consisting of six levels. [3] Line detection and character detection algorithms for the segmentation are used. Matrix feature extraction method for segmentation of the text in image is done. The Kohonen neural network is used for the classification of the characters. [4] They have developed a system which converts image's text to speech that helps visually impaired people. Raspberry Pi is used and Raspberry Pi camera captures the image. Tools like OCR (Optical Character Recognition) and TTS(Text-To-Speech) engines helps in recognition and speech synthesis. [5] This paper compares and contrasts between offline and online character recognition methods. The online input approach is represented as function of time and order of stroke of a specialized pen on a digital surface. Whereas in offline approach the image is scanned and loaded in the frame buffer. The conclusion says online recognition systems are better than offline recognition systems. [6] They have combined the concept of OCR and TTS synthesizer in MatlabR2011b. MATLAB is used for Preprocessing. (Preprocessing includes Gray scaling, contrast adjustment, thresholding). The preprocessed image is then converted to text document using OCR which includes line finding, base line fitting, and word recognition. Microsoft SAPI is used for text-to-speech conversion. [7] This paper talks about identifying objects from images and converting them to text and a speech synthesis. The location of the object is found out using LG graphs. They have used similar methodology like image analysis and understanding which includes segmentation, feature extraction, recognition and classification. [8] The authors propose a work of image to text to speech conversion system. Edge detection is the tool for segmenting the image into regions of discontinuity. In this project they have concentrated mainly on image segmentation as they consider image segmentation as the main work in the image processing. After conversion of image to text; next is text to speech.

[9] An already existing method which uses OCR for the conversion of image to text is used. So they mainly focus on TTS or speech synthesis. A TTS system has: front end and back end. The front end converts text into written-out words and then it allots phonetic transcriptions to each word. The back end is the synthesizer which converts the transcrpited words into sound. [10] Their system has: number plate area extraction and character recognition. Morphological operations are done. Character segmentation is done after plate area extraction. Histogram-based character segmentation is a simple and efficient technique used for segmentation. To extract the characters template matching is used. For character segmentation, histogram-based method is used. [11] Preprocessing, binarization of the image and small region removal is done in the first phase of their project. As the second phase of their project, segmentation is done based on thresholding using Otsu's Method. In the third phase for recognition they have used Tesseract OCR. It is then followed by token matching. [12] We observed that in this paper, their method consists of three stages- preprocessing, feature extraction, recognition using MATLAB. Preprocessing-removal of noise. Feature extraction-defining classes for particular features. Recognition-assigning each character pattern to a predefined class set. Their recognition rate is 98%. [13] They have done a survey on all the available recognition techniques using OCR and their accuracy percentage. According to their survey the technique with 98.86% accuracy is Support Vector Machine (SVM) with dataset 74K. Initially in this paper they have discussed about various generations of OCR. [14] Various terminologies and techniques involved in handwritten character recognition are discussed. They have discussed all the steps used by researchers for various techniques in detail. According to the authors this problem still requires increase in recognition activities. They have suggested some optimal systems for better recognition accuracy. [15] In this paper we have observed that they have discussed about the various techniques of feature extraction. They have also discussed about the usefulness of feature extraction and also types of features in image processing system. Using this paper, we can decide which feature extraction technique can be used for which type of image and also based on the complexity of image. [16] They have used MATLAB to do preprocessing and segmentation of character using bounding box. They have stated that they have achieved promising results using these techniques although there are few shortcomings. If a character image is not completely connected it is divided as two different objects.

III. PROPOSED WORK

Captured images of handwritten characters are processes in three phases. In the first phase, the image is preprocessed to make it operable for further phases. The second phase includes segmentation, feature extraction and recognition using OCR. The output we get is the recognized characters in the form of a .txt file. The last phase is about text-to-speech conversion.

A. PREPROCESSING

In this phase we have converted rgb image to grayscale using function rgb2gray. This function helps in elimination of the hue and saturation information and retaining the luminance by doing this we can reduce the complexity of the image. Next for binarizing the image we have used graythresh function which uses Otsu's method that chooses the threshold to minimize intraclass variance of black and white pixels. The graythresh function calculates a global threshold which is used to convert an intensity image to binary image. Lastly we have used bwareaopen function to remove connected components that have lesser than the specified pixel values from the binary image, giving rise to another binary image.

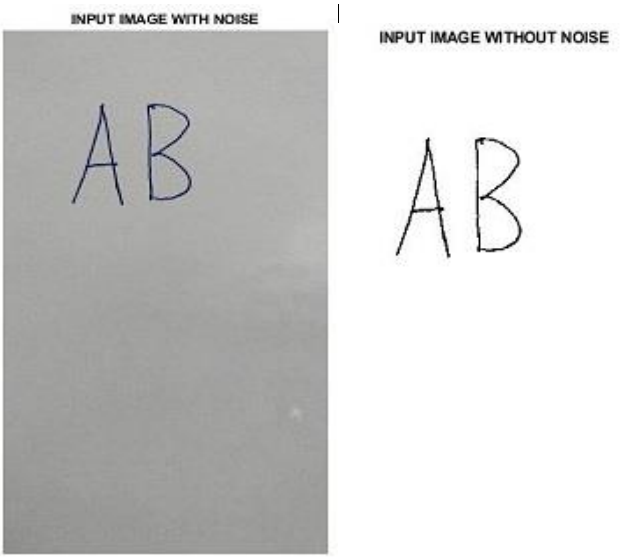

Fig.1- Noise Removal

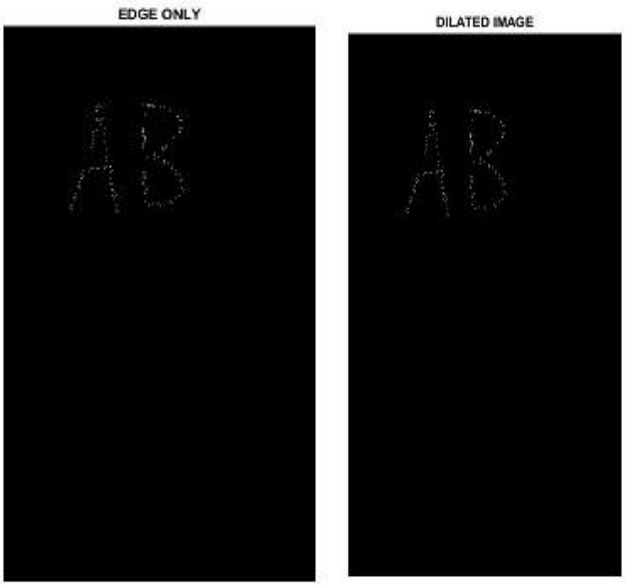

Fig.2- Morphological Operations

B. CHARACTER RECOGNITION

Character recognition has two sub-phases.

1. *Segmentation and Feature Extraction*.

Segmentation is done on the binarized image as it reduces the complexity of the process. Firstly, we have detected the edges of characters in the image.

The function edge outputs a binary image having 1s where there are edges in the image and 0s in different places.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication (BEIESP)

By default, edge uses the sobel edge detection method. Vertical direction and horizontal direction are the two types of edges in an image which are detected using Sobel operator. After edge detection, dilation and erosion are being performed using the structuring elements. By doing this we can find out the perimeter of the object. Using that perimeter value, we can get the outline of the object. we have used imdilate, imerode, BWperim functions. The function regionprops measures properties of each region which is labeled in label matrix. The property that we used to provide the shape measurements is BoundingBox, smallest rectangle which consists of the region returned as 1-by-Q*2 vector where Q is the number of image dimensions. The output we got is the segmented characters inside the rectangle.

Fig.3- Segmentation

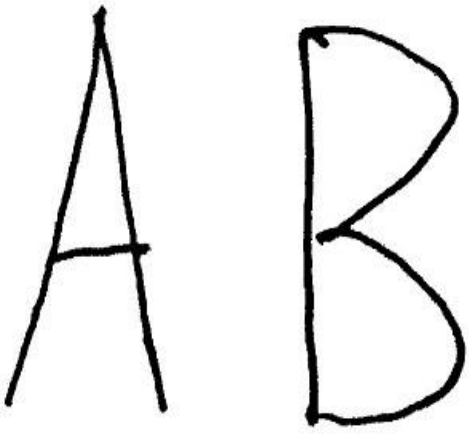

Fig.4- Segmented Characters

2. Optical Character Recognition.

Firstly, we have created templates for all the characters in different styles and we read them. We then created a function to read the letters. Correlation is computed between the template and the input image to give a string containing letters as the output. The size of that input image must be 42x24 pixels. We have used corr2 function to correlate the templates and the images. The corr2 function is used to find out the similarities between the two images. The created templates of all the letters are read and stored in a matrix. Initially, the segmented image is stored in a matrix for the ease of further operations. Then we open a '.txt' file to write the recognized characters to it. The templates are then loaded and read so that they can be matched with the input dataset. The next step is to find the number of letters and white spaces in the input image. The first letter in the line, the remaining line and the space between adjacent letters are stored together. Letters are then resized so that correlation can be performed. Then the user defined function compares the resized image and the template image stored in a folder. Then the recognized character is written to the .txt file and a new window pops up containing the recognized character.

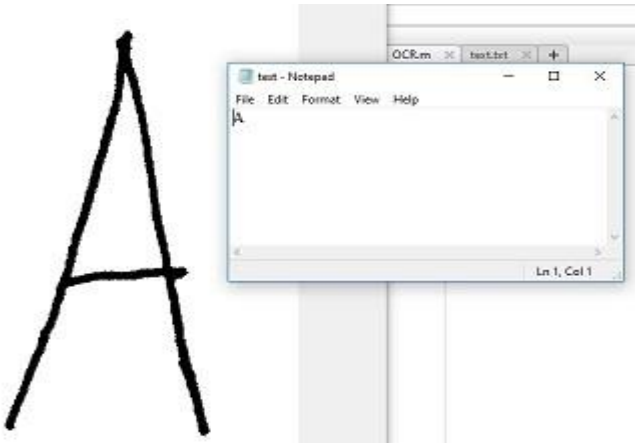

Fig.5- Recognized character written to notepad file

C. TEXT-TO-SPEECH

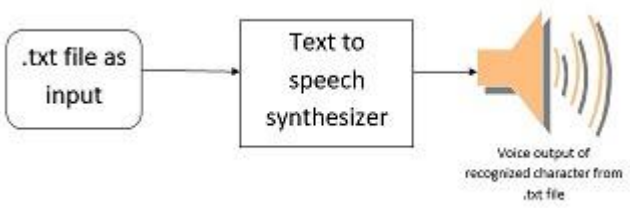

Fig.6- Text-to-speech System

This is the last phase in our project where we convert the content in the .txt file to a voice output. To achieve this, we have used Microsoft .NET framework. In our project we are creating the object for speech synthesizer. The object uses default system voice to give the voice output. The .txt file which we got as output from the second phase will be fed as an input to the speech synthesizer. The namespace System.Speech.Synthesis contains the classes for configuring a speech synthesis engine and also prompts for generating speech for modifying the voice characters.

IV. EXPERIMENTAL RESULTS

We considered dataset of uppercase letters, lowercase letters and digits. The results of these were found accurate but there was a hindrance in recognizing cursive characters.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication (BEIESP)

The text-to-speech system shows 100% accuracy for all the characters as well as the digits. The below figures Fig. 7 and Fig. 8 show the results of character recognition.

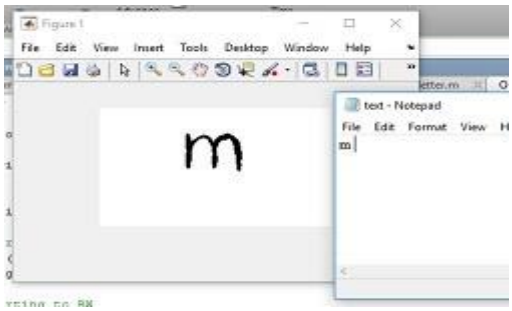

Fig.7- Recognized character ‘m’

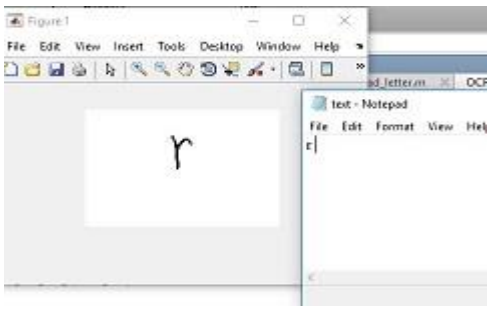

Fig.8- Recognized character ‘r’

Fig. 9 and Fig. 10 represent the lowercase character 'f' written in two different ways and both the characters are being recognized as 'f'. This is to say that we were successful in recognizing most of the characters written in variant styles. Although we faced failures in recognizing few complicated cursive characters

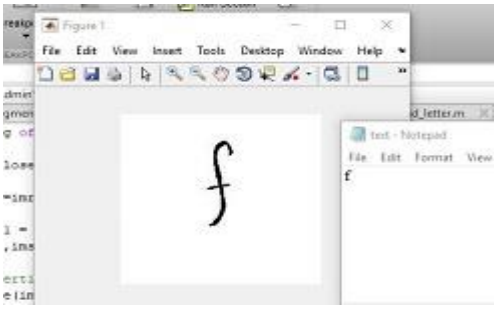

Fig.9- Lowercase character ‘f’

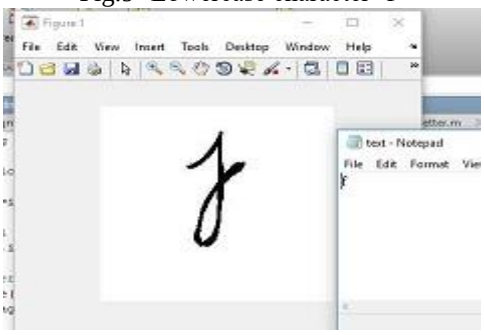

Fig.10- Lowercase character ‘f’

V. CONCLUSION

The three datasets which we have considered are lowercase characters, uppercase characters and digits. We have achieved at most results for the datasets that we considered which consisted of cursive letters and complicated structure of characters. We calculated the F-measure for each of the datasets and we obtained the results as 82% for lowercase characters, 89% for digits and 93% for uppercase characters.

Table-1
F-Measure Calculation

DATASET	F-MEASURE
Uppercase Characters	93%
Lowercase Characters	82%
Digits	88%

VI. FUTURE WORK

The images that we have considered were taken from phone camera. Only the preprocessing techniques were being performed on the input images. We can get better results if we use a compression technique to compress the images before performing segmentation. The compression reduces the image size and thus facilitates it to be better to operate on.

REFERENCES

1. Rithika.H, B. Nithya santhoshi, "Image Text To Speech Conversion In The Desired Language By Translating With Raspberry Pi", IEEE International Conference on Computational Intelligence and Computing Research, 978-1-5090-0612-0/16.
2. Mirjana Kocaleva, Done Stojanov, Igor Stojanovik, Zoran Zdravev, "Pattern Recognition and Natural Language Processing: State of the Art", TEM Journal 5(2) 236–240, Volume 5, Number 2, 237, 238, 239, 240.
3. Pratik Madhukar Manwatkar, Shashank H. Yadav, "Text Recognition from Images", IEEE Sponsored 2nd International Conference on Innovations in Information,Embedded and Communication systems (ICIIECS)2015, 978-1-4799-6818-3/15.
4. Asha G. Hagargund, Sharsha Vanria Thota , Mitadru Bera, Eram Fatima Shaik, "Image to Speech Conversion for Visually Impaired", International Journal of Latest Research in Engineering and Technology (IJLRET), ISSN: 2454-5031, Volume 03, Issue 06.
5. Anisha Priya, Surbhi Mishra, Saloni Raj, Sudarshan Mandal, Sujoy Datta, "Online and Offline Character Recognition: A Survey", International Conference on Communication and Signal Processing, April 6-8, 2016, India, 978-1-5090-0396-9/16.
6. K Kalaivani,R Praveena, V Anjalipriya, R Srimeena,"Real time Implementation of Image Recognition and Text to Speech Conversion",2014
7. Nikolas Bourbkis, "Automatic image-to-Text-to-Voice Conversion for Interactively Locating Objects in Home Environments",20th IEEE Interactive conference on Tools with Artificial Intelligence,2008,1082-3409/08.
8. Mrunmayee patil, Ramesh kagalkar, "Review on conversion of image to text as well as speech using edge detection and image segmentation", International journal of science and research(IJSR),2319-7064,2012,3.358
9. Jishna Gopinath, Arvind S, Pooja Chandran, Saranya S S, "Text to Speech conversion using OCR", International Journal of Emerging technology and advanced engineering, ISSN 2250-2459, volume 5, issue 1, January 2015.
10. Chetan Pardeshi, Priti Rege, "Morphology Based Approach for Number Plate Extraction‘‘, Proceedings of the International Conference on Data Engineering and Communication Technology, ICDECT 2016, Volume-2.
11. Apoorva P, Bindushree M, Bhamati H A, "Nutritional facts label processing using image segmentation and token matching based on OCR", International Journal of Applied Engineering Research(IJAER), ISSN:0973-4562, Volume 11, 2016.
12. M. Sivasankari, Dr. S.P. Victor,"Handwritten text Recognition", International Journal of Computer Science and Information Technologies, ISSN:0975-9646, Volume 5, 2014.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication (BEIESP)

13.V S Dhaka, Manoj Kumar, Prashant Chaudhary, "Offline Handwritten English Script Recognition: A Survey", Special Conference Issue: National Conference on Cloud Computing and Big Data.https://www.overleaf.com/project/5c8bd8c2b7a95f14c03b1f14
14.Madhuri Maheshwari, Deepesh Namdev, Saurabh Maheshwari," A systematic review of automation in handwritten character recognition", International Journal of Applied Engineering Research , ISSN: 0973-4562, Volume 13 , 2018
15.Gaurav Kumar, Pradeep Kumar Bhatia, "A detailed review of feature extraction in image processing systems",2014.
16.Amit Choudhary,Rahul Rishi, Savitha Ahlawat ," A New Approach To Detect And Extract Characters From Offline Printed Images And Text", Procedia Computer Science 17 , 434 – 440, 2013.
17.Akshay S,"Single Moving Object Detection and Tracking Using HornSchunck Optical Flow Method", International Journal of Applied Engineering Research, Vol:10, No:11, 2015.
18.Akshay S, Apoorva P,"Segmentation and classification of FMM compressed retinal images using watershed and Canny Segmentation and Support Vector Machine",2017 International Conference on Communication and Signal Processing(ICCSP), 2018.

AUTHORS PROFILE

AKSHAY S received M.S degree in Computer Science from University of Mysore in 2012 and is currently pursuing PhD. He is working as an Assistant Professor in Department of Computer Science, Amrita Vishwa Vidyapeetham, Mysore campus.

CHANDANA J R, BCA student at Amrita Vishwa Vidyapeetham, Mysore campus.

NAMITA G, BCA student at Amrita Vishwa Vidyapeetham, Mysore campus.