International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-9 Issue-10, August 2020

Optical Character Reader & Text To Speech Conversion using Correlations & Speech Synthesis

Avinash Rai, Shivani Sonker

***Abstract**: In the modern era of image processing, recognizing content or information from an image is process of electronic conversion into machine encoded text. Advanced systems that are capable of producing high accuracy for multi-font recognition are now becoming commonplace, and with the support of digital consent formatting. Some programs are able to retrieve formats that are very close to the original page including images, columns, and other non-text items. Proposed system is able to recognize text from an image and convert it into editable text along with speech conversion. System uses Correlation model for OCR (Optical Character Recognition) and Speech Synthesis for TTS (Text To Speech) conversion. Correlation is a measurement of the similarities between two similar objects such as the predefined alphabets and recognizing a combination of those alphabets from an image. Speech synthesis is an artificial expression of human speech. The computer program that has been used this feature is called a speech computer as well as speech synthesizer that can be implemented on the basis of software or hardware primitives. The text-to-speech system (TTS) converts a standard language text into a speech; some programs provide figurative language presentations such as typed text in speech. System is capable enough to acquire high level of accuracy with less false recognition. It is required to built an effective text scanner that can recognize text from an image with less error rate. System has been implemented in MATLAB and various pre-processing filters have been applied for better enhancement and extraction. Hand written text can also be recognized with an effective manner.*

Keywords: OCR, TTS, Speech Synthesis, Correlation Model, Machine Encoding, Image Processing.

I. INTRODUCTION

The OCR and TTS visual system based on low vision include: image acquisition module, used to scan the object, as well as image acquisition and removal; the processing module, which includes the OCR character recognition unit, is linked to the image acquisition module and is used to obtain the image and to increase the recognition and recognition of a single character in the image, in order to find the text file associated with the image; a TTS engine unit, linked to an OCR character recognition unit and used to convert a text file into an audio file; and an output module, which is linked to the processing module and is used to output the text file and audio file respectively. The low-vision learning aid system combines OCR and TTS technology, image acquisition module scans a readable image and captures the module, processes the image, and finally modes the text and associated audio, thus providing the user with a way of learning where listening plays. The main role and gaze play a secondary role, thus achieving the benefits of using simple eye fatigue and reducing it. Compared to the application for obtaining audio for customized audio files, Speech Engine TTS has only a few million sizes, does not require a large number of audio files, can therefore save a huge amount of storage space, and can read aloud even if an unknown statement is in advance. These TTS software applications have to monitor the phone functionality, such as that some broadcasting methods can be used to read a novel or read testimonials, can also read aloud an email, some electronic dictionaries can read a word, and can be used by the Help Center and automatically play information about services etc [1].

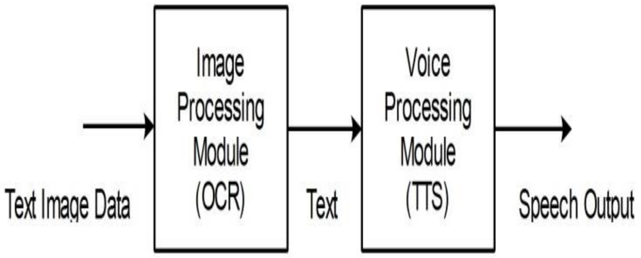


Fig. 1.Process Model

A low-vision visual based OCR and TTS-enabled design learns a program that facilitates vision and is integrated with OCR character recognition technology and TTS speech recognition technology, with image capture module scanning and image collection, by processing a module image compiled for processing. and finally by using the same display module for reading text that emits compatible audio frequency, thus being read to be professional, the user's visual aid

Revised Manuscript Received on August 30, 2020.
* Correspondence Author

Dr. Avinash Rai*, Department of Electronics and Communication, University Institute of Technology, Rajiv Gandhi Proudyogiki Vishwavidyalaya, Bhopal, Madhya Pradesh, India. E-mail: avinashrai@rgtu.net

Ms. Shivani Sonker, Department of Electronics and Communication, University Institute of Technology, Rajiv Gandhi Proudyogiki Vishwavidyalaya, Bhopal, Madhya Pradesh, India. E-mail: shivanisonker19@gmail.com

reading method reaches the listener dried, white line, with black written characters, eye-catching pattern, reduces eye strain, achieving the result that a low-level visual patient, presbyopia crowd and blind users complete the reading.

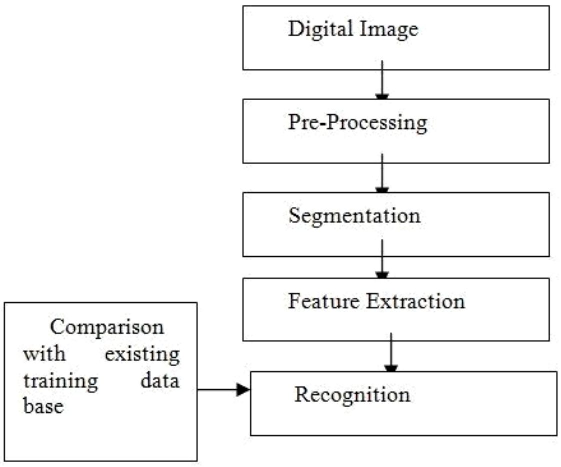


Fig. 2.OCR Components [2]

OCR word recognition unit, linked to image caption module, image acquisition, semantic image input and individual character spacing, find image caption matching image description.

II. RELATED WORKS

A. Literature Survey

Sneha.C.Madre [3] et al. proposed a system which is based on traditional OCR approach that directly segments the image. This paper shows the actual effort of the text being extracted from the image using Optical character recognition and the known character is converted to sound using the MATLAB environment. This suggested application is cost-effective, efficient and real-time. For this system, read the text in a text, newspaper, E-mail. This program can also be beneficial to people with visual impairments. And the interesting fact is that I'm trying to create a single hardware space where it's MATLAB. But the system can be installed and can be carried everywhere easily. The main purpose of this program is to meet the needs of Hand-Managed people and can interact with people who do not understand sign language. The TTS system can also be used in domain-specific programs such as train announcements. Shimona Gupta [4] et al. proposed a system which is based on Adaptive Neuro Fuzzy Inference System (ANFIS). This paper provides an automated system for finding text in a picture, dividing lines and letters and extracting individual character traits that can be used for character recognition followed by speech structure. In the pre-processing part the image is expanded, filtered and converted to a gray image to facilitate additional work, after pre-processing, the pre-processed image is activated when the text is found in the image and the lines and characters are removed. The morphological function is used to locate the text in the image. Linear segmentation was performed using the Y histogram and character removal was performed using the BoundingBox property of the linked component analysis. The feature release is the start of a debugging module. Zoning process is used for character extraction of characters; the features of the feature released are provided by the Adaptive Neuro Fuzzy Inference System for character identification that gives the character its identity and transforms it into speech. The overall accuracy achieved for this program is approximately 92.9%.

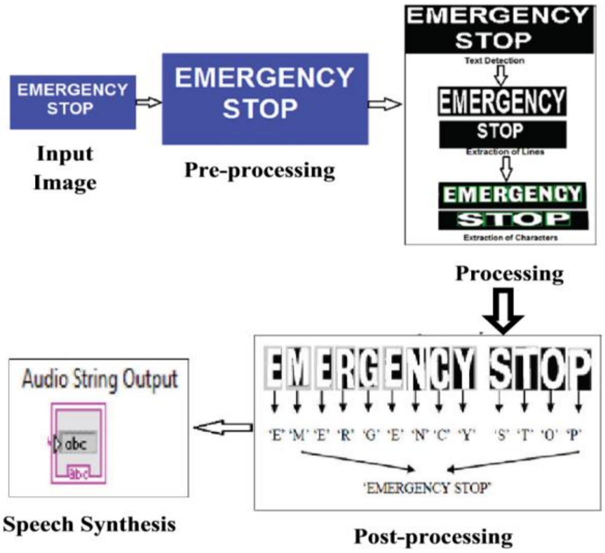


Fig. 3.Overview of the System [4]

Mullani.J.J [5] et al. proposed a system which is based on LABView approach which is Microsoft speech conversion software. This work introduces the OCR based Speech Synthesis System to produce record yields that can be used as a direct means of communication between people. The framework is available in the Lab section to review 7.1. There are two phases of the framework initially by the OCR and then the second by Speech Synthesis. In OCR printed or embedded in character databases it is scanned and the image is acquired using LabA's IMAQ view and then the characters are identified using the classification and composition techniques done in LabVIEW. The content acquired in the Insecond section is converted into speech. The ACTIVE X inferior Communication bed is used to trade information between applications. ActiveX's new design provides a general model in text conversion that different programming languages can operate at different stages. Nidhi Kalidas Sawant [6] et al. proposed a system which is based on Support Vector Machine. In this paper, the method presented provides text on the speech translation of the Devanagari printed machine using OCR. As can be seen from the results we get the Optical Character Recognition correctly using the multiclass SVM classifier. The classifier gives the result of a text stored in a notepad and is also converted into a speech output. This applies to images or documents scanned with a printed Devanagari script. Abhishek Mathur [7] et al. proposed a system which is based on AI-OCR, An AI-based learning system that uses OCR is an artificial learning system produced using smart phone cameras integrated with OCR (Optical Character Recognition). The program captures the text using the camera and scans the text and converts it to a digital text visualized by the program and displays the translated text and gives a speech effect. To understand the power of a project, a basic idea of what AI and OCR is required. This report describes all the functionality of the Language Translator, as well as a few requirements for using it.

Therefore, a visually impaired person can easily use this AI-based learning program as a simple application friendly across the globe.

III. PROPOSED WORK & IMPLEMENTATION

Proposed system is based on correlations and speech synthesis models where OCR recognizes the words effectively with high precision rate along with correct text to speech conversion. System is efficient for separating two closely spaced characters. System pre-processed the data using scaled color mapping with binary thresholding for better recognition then process the model with correlation for recognizing characters from the images and then apply speech synthesis for test to speech conversion effectively. There are two main approaches that system used i.e. Correlation and Speech Synthesizer. Where, correlation is a basic operation that we perform to extract information from images. Correlation is an image processing model which is also knows as a filtering technique for enhancing image and detecting edges from an image. It is also useful for image compression, color correction, object recognition & image segmentation. It accepts certain frequencies and rejects the other frequencies for removing background and highlighting the foreground objects. Correlation is a measurement of the similarities between two similar objects such as the predefined alphabets and recognizing a combination of those alphabets from an image.

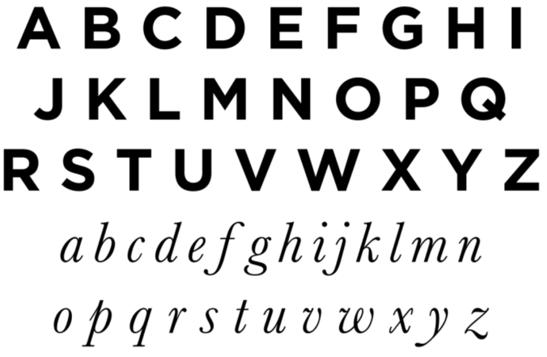

A B C D E F G H I
J K L M N O P Q
R S T U V W X Y Z
a b c d e f g h i j k l m n
o p q r s t u v w x y z

Fig. 4.Predifined Alphabets

BHOPAL

Fig. 5.Recognizing String

Fig. 4 & 5 both are from similar family but one is the predefined or universal group and another one is composite string that may be a dictionary words or may not.

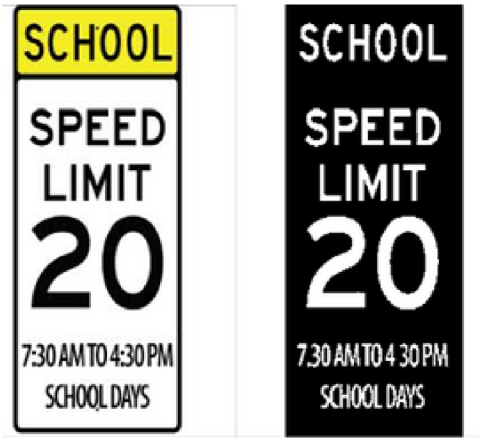
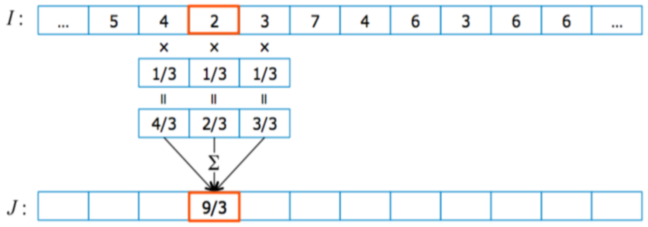


Fig. 6.Correlation Module

Effect of Correlations over an image that highlighted the active frequencies and masks the background and may extract these characters.

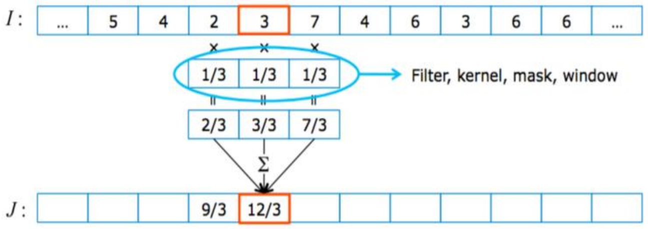
B. Correlation Model

Correlation is the process of moving a filter mask that is often referred to as a kernel over the image and calculating the number of products in each area. Integration is a function of filter migration. In other words, the first value of the connection corresponds to the total removal of the filter, the second value corresponds to the single displacement unit, and so on.



It is an averaging filter and its boundary conditions are –

- Ignore filtered values at the boundaries
- Pad with zeros
- Pad with first/last image values



$$F \circ I(x) = \sum_{i=-N}^{N} F(i)I(x+i)$$

Where o (circle) denotes correlation, F has 2N+1 element with –N to N limit and F(0) is the center element. Gaussian blur as common practice –

$$Log\,(x,y) = -\frac{1}{\pi\sigma^4}\left[1 - \frac{x^2+y^2}{2\sigma^2}\right]e^{-\frac{x^2+y^2}{2\sigma^2}}$$

The value of σ is 1 for 3x3 matrix and 2 for 5x5 matrix, (x, y) are native pixels.

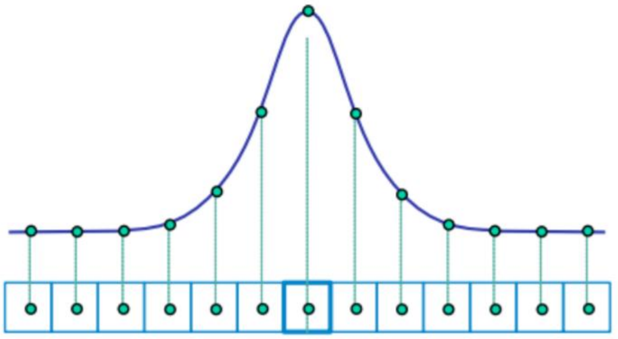
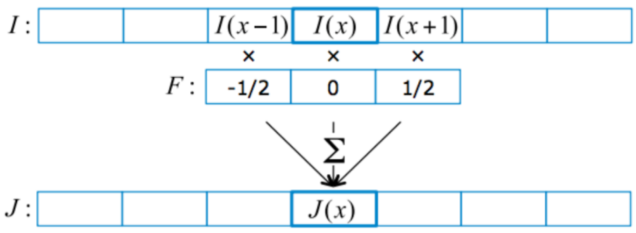


Fig. 7.Normalized Gaussian Filter

σ - Amount of smoothening



$$J(x) = \frac{I(x+1) - I(x-1)}{2}$$

C. Speech Synthesis

Speech synthesis is an artificial expression of human speech. The computer program that has been used this feature is called a speech computer as well as speech synthesizer that can be implemented on the basis of software or hardware primitives. The text-to-speech system (TTS) converts a standard language text into a speech; some programs provide figurative language presentations such as typed text in speech. The text-to-speech system (or "engine") is made up of two parts: the front end and the back end. The front-end has two main functions. First, it converts raw text containing symbols such as numbers and abbreviations into the same word and text. This process is often referred to as text typization, pre-processing, or embedding. The front-end then assigns phonetic writing to each word, then splits and tags the text into prosodic parts, such as phrases, phrases and sentences. The back-end is often called a synthesizer and turns the symbolic language representation into sound.

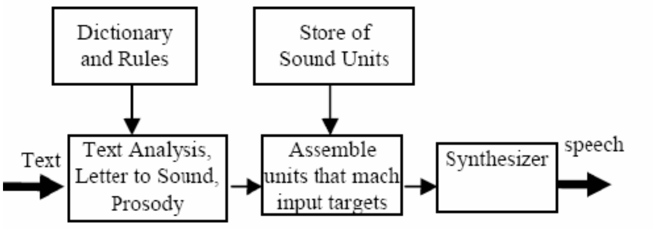

Dictionary and Rules
Store of Sound Units
Text
Text Analysis, Letter to Sound, Prosody
Assemble units that mach input targets
Synthesizer
speech

Fig. 8 Block Diagram of Speech Synthesizer

$$P_\theta(x) = \sum_{t=1}^{T} p(x_t|x_1,\dots,x_{t-1})$$

$P_\theta(x)$ is phonetic parameter, t-T is the time domain, $p(x_t|x_1,\dots,x_{t-1})$ is the frequency domain with respect to time. Most text-to-speech systems (TTS) do not produce representations of their input text, as the procedures for doing so are unreliable, poorly understood, and ineffective. Because of this, various guessing techniques are used to guess the correct way to trick home pages, such as checking out neighboring words and using frequency statistics. Recently TTS systems have begun to use HMMs (discussed above) to produce "speech components" to help disrupt home computers. This method is most effective in many cases such as "read" should be called "red" which means the past tense, or as "reed" meaning the present tense. Standard error rates when using HMMs in this way are usually less than 5 percent. These programs work well in many European languages, although access to the necessary training companies is always difficult in these languages.

D. Flow Chart

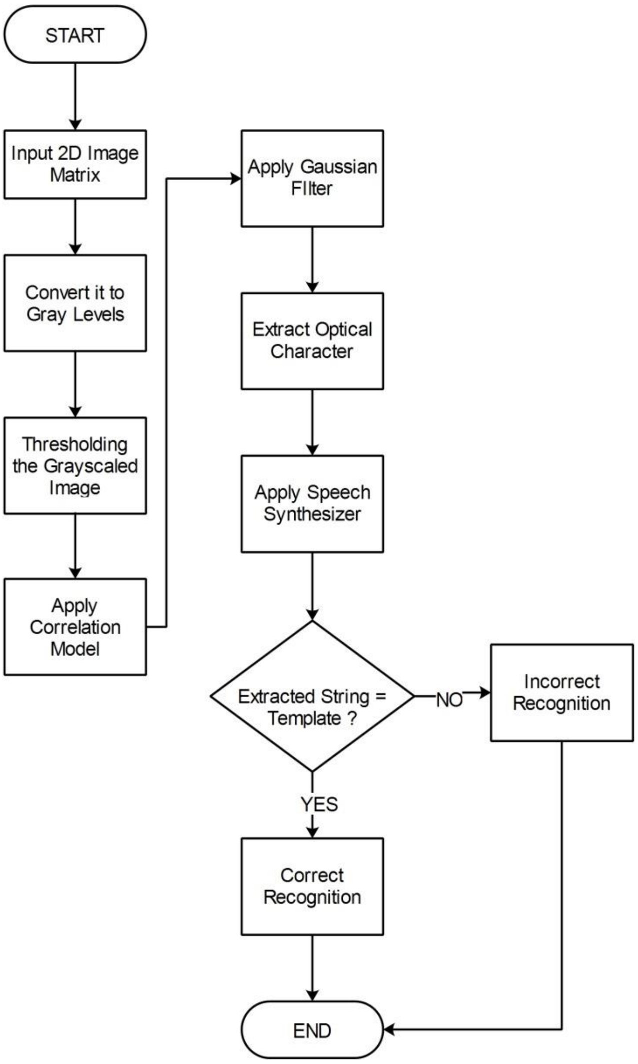


Fig. 9 Flow Chart of Proposed Work

In flow chart, first of all a 2D image matrix has to be browsed and input for preprocessing the data. Then it will convert it into gray levels as scaled color mapping and segment the image by binarization. Then Correlation filter has been applied once the frame is pre-processed. Then Gaussian filter has been applied for proper smoothening the data for background elimination then optical character can be extracted and speech synthesis can be applied for text to speech conversion.

E. Correlational & Speech Synthesizer Algorithm

Input: 2-D Image Matrix

Output: Convolutional Matrix

Step 1: Input 2-D Image Matrix

Step 2: Convert 2-D Image Matrix to Gray Levels with Scaled Color Mapping

Step 3: Thresholding the Gray Scaled Image to Binary Codes

Step 4: Initialize Correlation Function F

$$F \circ I(x) = \sum_{i=-N}^{N} F(i)I(x+i)$$

Where o (circle) denotes correlation, F has 2N+1 element with –N to N limit and F(0) is the center element.

Step 5: Apply Gaussian Filter

$$Log\,(x,y) = -\frac{1}{\pi\sigma^4}\left[1-\frac{x^2+y^2}{2\sigma^2}\right]e^{-\frac{x^2+y^2}{2\sigma^2}}$$

The value of σ is 1 for 3x3 matrix and 2 for 5x5 matrix, (x, y) are native pixels.

Step 6: Convert extracted optical character to speech synthesis i.e. grapheme-to-phoneme

$$P_\theta(x) = \sum_{t=1}^{T} p(x_t|x_1,\dots,x_{t-1})$$

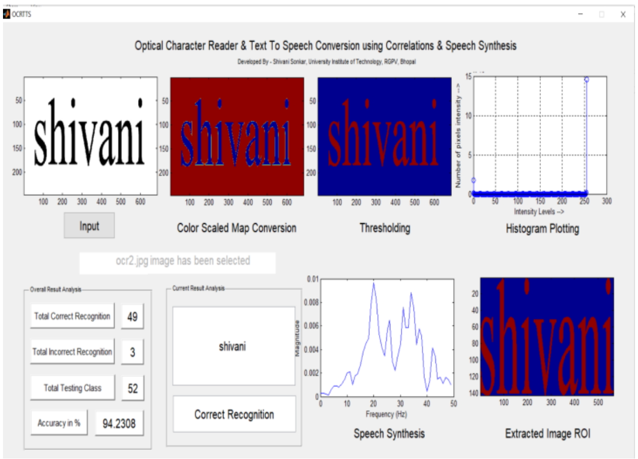
$P_\theta(x)$ is phonetic parameter, t-T is the time domain, $p(x_t|x_1,\dots,x_{t-1})$ is the frequency domain with respect to time.

Step 7: Plot energy signal for phonetic time-freq domains

Step 8: End

IV. RESULT ANALYSIS

There are total no. of 52 frames have been tested and results are obtained accordingly. If a system is able to recognize character that matches with the template then it will be considered as correct recognition otherwise it will consider as incorrect recognition. System obtained 49 correct recognition attempt and 3 as incorrect recognition. So, the accuracy has been computed on the basis of correct recognition and incorrect one which is 94.23 %.



OCRTTS

Optical Character Reader & Text To Speech Conversion using Correlations & Speech Synthesis

Developed By - Shivani Sonkar, University Institute of Technology, RGPV, Bhopal

ocr2.jpg image has been selected

Overall Result Analysis

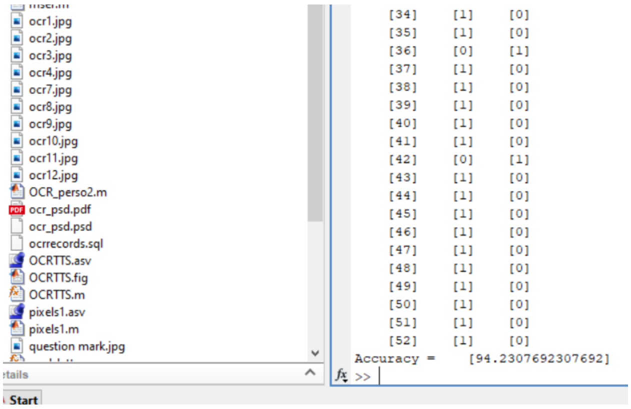
Total Correct Recognition	49
Total Incorrect Recognition	3
Total Testing Class	52
Accuracy in %	94.2308

Current Result Analysis

shivani

Correct Recognition

Fig. 10 Proposed GUI



```
  ocr1.jpg
  ocr2.jpg
  ocr3.jpg
  ocr4.jpg
  ocr7.jpg
  ocr8.jpg
  ocr9.jpg
  ocr10.jpg
  ocr11.jpg
  ocr12.jpg
  OCR_perso2.m
  ocr_psd.pdf
  ocr_psd.psd
  ocrrecords.sql
  OCRTTS.asv
  OCRTTS.fig
  OCRTTS.m
  pixels1.asv
  pixels1.m
  question mark.jpg
```

```
      [34]    [1]    [0]
      [35]    [1]    [0]
      [36]    [0]    [1]
      [37]    [1]    [0]
      [38]    [1]    [0]
      [39]    [1]    [0]
      [40]    [1]    [0]
      [41]    [1]    [0]
      [42]    [0]    [1]
      [43]    [1]    [0]
      [44]    [1]    [0]
      [45]    [1]    [0]
      [46]    [1]    [0]
      [47]    [1]    [0]
      [48]    [1]    [0]
      [49]    [1]    [0]
      [50]    [1]    [0]
      [51]    [1]    [0]
      [52]    [1]    [0]
Accuracy =     [94.2307692307692]
>>
```

Fig. 11 Simulation Result

Fig. 10 shows the graphical user interface of the proposed system with all pre-processing modules and approaches that have been applied over that. Fig. 11 shows the cosole result of the system that accquired 94.23 % of accuracy on the basis of 52 testing samples with different font, color, background, noisy image and many more. The obtained outcome is better than the previous one with less false recogniton rate.

CR- Correct Recognition, IR- Incorrect Recognition, TTC – Total Testing Class

Table- I: Result Analysis

	Proposed
CR	49
IR	3
TTC	52
Accuracy	94.23 %

$$\text{Accuracy} = \frac{TTC-IR}{TTC} * 100\ \%$$

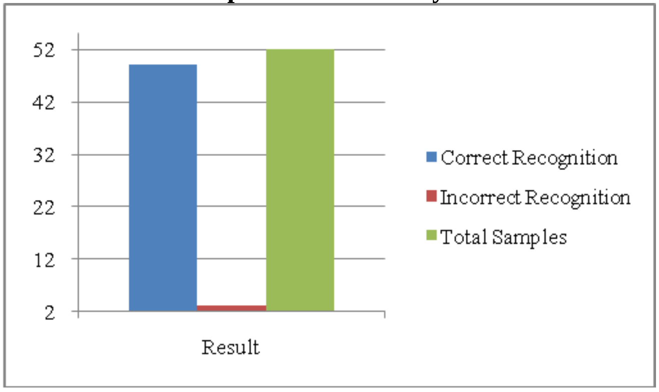
$$= \frac{52-3}{52} * 100\ \%$$

$$= 94.23\ \%$$

Table- II: Result Comparison

	Shimona Gupta [4]	**Proposed**
Method	Adaptive Neuro Fuzzy	Correlation and Speech Synthesizer
Target Language	English	English
Error Rate	7.10	6.77
Accuracy	92.9	94.23

Graph- I: Result Analysis



V. CONCLUSION & FUTURE SCOPE

The systems which have been proposed till now are not reliable because those systems are not efficient to extract the useful information. The information is missing the sensitive edges that trail the accuracy in order to achieve the correct recognition. The proposed system is capable enough to efficiently recognize the optical characters with text to speech conversion at real time with high level of accuracy. System uses correlation model and speech synthesizer for implementing the approach for better enhancement of the system. Accuracy is very important with respect to the correct and incorrect recognition for better outcomes. The technique can be enhanced in future where accuracy depends. Machine learning approach can be used in future for better analysis.

REFERENCES

1. Google Patents, OCR (Optical Character Recognition) and TTS (Text To Speech) based low-vision reading visual aid system , https://patents.google.com/patent/CN104966084A/en, Accessed- 07 July 2015.
2. Mathur, Geetika & Rikhari, Suneetha. (2017). ISSN: 2454-132X Impact factor: 4.295 A Review on Recognition of Indian Handwritten Numerals.
3. S. C. Madre and S. B. Gundre, "OCR Based Image Text to Speech Conversion Using MATLAB," 2018 Second International Conference on Intelligent Computing and Control Systems (ICICCS), Madurai, India, 2018, pp. 858-861, doi: 10.1109/ICCONS.2018.8663023.
4. S. Gupta and S. Gupta, "Character Recognition and Speech Synthesis using Adaptive Neuro Fuzzy Inference System," 2018 International Conference on Advances in Computing, Communication Control and Networking (ICACCCN), Greater Noida (UP), India, 2018, pp. 1091-1096, doi: 10.1109/ICACCCN.2018.8748742.
5. J. J. Mullani, M. Sankar, P. S. Khade, S. H. Sonalkar and N. L. Patil, "OCR BASED SPEECH SYNTHESIS SYSTEM USING LABVIEW : Text to Speech Conversion System using OCR," 2018 Second International Conference on Computing Methodologies and Communication (ICCMC), Erode, 2018, pp. 7-14, doi: 10.1109/ICCMC.2018.8487731.
6. N. K. Sawant and S. Borkar, "Devanagari Printed Text to Speech Conversion using OCR," 2018 2nd International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC)I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC), 2018 2nd International Conference on, Palladam, India, 2018, pp. 504-507, doi: 10.1109/I-SMAC.2018.8653685.
7. A. Mathur, A. Pathare, P. Sharma and S. Oak, "AI based Reading System for Blind using OCR," 2019 3rd International conference on Electronics, Communication and Aerospace Technology (ICECA), Coimbatore, India, 2019, pp. 39-42, doi: 10.1109/ICECA.2019.8822226.
8. Qixiang Ye and David Doermann,," Detection and Recognition in Imagery: A Survey", IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGNCE VOL. 37, NO. 7, JULY 2015.
9. Amit Choudhary, Rahul Rishi, Savita Ahlawat," Off-Line Handwritten Character Recognition using Features Extracted from Binarization Technique", 2013 AASRI Conference on Intelligent Systems and Control.
10. Pratik Madhukar Manwatkar and Shashank H. Yadav, "Text Recognition from Images", EEE Sponsored 2nd International Conference on Innovations in Information,Embedded and Communication systems (ICIIECS)2015.
11. Kumuda T and L Basavaraj," Edge Based Segmentation Approach to Extract Text from Scene Images", 2017 IEEE 7th International Advance Computing Conference.
12. Xiaoming Huang, Tao Shen, Run Wang, Chenqiang Gao," Text Detection and Recognition in Natural Scene Images", 2015 International Conference on Estimation, Detection and Information Fusion (ICEDIF 2015).
13. Jagruti Chandarana, Mayank Kapadia," Optical Character Recognition", International Journal of Emerging Technology and Advanced Engineering 2014.
14. Gupta Mehula, Patel Ankita, Dave Namrata, Goradia Rahul, and Saurin Sheth," Text-Based Image Segmentation Methodology", 2nd International Conference on Innovations in Automation and Mechatronics Engineering, ICIAME 2014.
15. Rodolfo P. dos Santos, Gabriela S. Clemente, Tsang Ing Ren and George D.C. Calvalcanti," Text Line Segmentation Based on Morphology and Histogram Projection", 2009 10th International Conference on Document Analysis and Recognition.
16. Anitha Mary M.O. Chacko and P.M. Dhanya," A Comparative Study of Different Feature Extraction Techniques for Offline Malayalam Character Recognition", Springer India 2015 Computational Intelligence in Data Mining - Volume 2, Smart Innovation, Systems and Technologies 32, DOI 10.1007/978-81322-2208-8_2.
17. Mustain Billah, Sajjad Waheed, Abu Hanifa, "An Optical Character Recognition System from Printed Text and Text Image using Adaptive Neuro Fuzzy Inference System", International Journal of Computer Applications Volume 130 - No.16, November 2015.

AUTHORS PROFILE

Dr. Avinash Rai has done B.E. (Electronics), M.E. (VLSI Design), Ph.D. (Wireless Sensor Network) and is currently working as Assistant Professor in Department of Electronics & Communication Engineering at University Institute of Technology, Rajiv Gandhi Proudyogiki Vishwavidyalaya, Bhopal (M.P). Dr. Avinash Rai does researches in Wireless Sensor Network, Telecommunication Engineering and Electronics Engineering. His most recent publication is "Enhancing Energy of Sensor Node to Increase Efficiency in Wireless Sensor Network"

Shivani Sonker has done B.E. (Electronics & Communication) and is currently an M.E. scholar (Digital Communication) in the Department of Electronics & Communication Engineering at University Institute of Technology, Rajiv Gandhi Proudyogiki Vishwavidyalaya, Bhopal (M.P). Shivani Sonker has undertaken research work presented in this paper under the guidance of Dr. Avinash Rai which is required in the partial fulfillment for the award of M.E. in digital communication.