International Journal of Early Childhood Special Education (INT-JECSE)
DOI:10.9756/INTJECSE/V14I5.43 ISSN: 1308-5581 Vol 14, Issue 05 2022

CONVERSION OF IMAGE TO TEXT TO SPEECH USING OCR AND TTS SYSTHESIS

Dharshini V,
Student, Department of Computer Science & Engineering, Prince Dr. K. Vasudevan College of Engineering and Technology,Chennai,India

Keerthana P,
Student, Department of Computer Science & Engineering, Prince Dr. K. Vasudevan College of Engineering and Technology,Chennai,India

Manju T,
Professor,Department of Computer Science & Engineering, Prince Dr. K. Vasudevan College of Engineering and Technology, Chennai, India

Deepa R
Head of Department, Department of Computer Science & Engineering, Prince Dr. K. Vasudevan College of Engineering and Technology,
Chennai, India
ermanjucs@gmail.com

Abstract: - The main purpose of this project is to overcome the problems facing by the blind people and illiterates. Because the blind people and illiterates can be easily manipulated, this leads to misuse. To overcome this problem we are proposing a device which helps in conversion of hard copy of text which is inserted into the device will be converted to speech. Most of these applications find the use of functions such as articulators, conversion from text to synthetic speech signals, language translation amongst various others. In our project, we'll be executing OCR and TTS Synthesis that are applied to achieve the concept of the project. Here OCR is used to extract the words in the given object as a text from help of a webcam and the TTS synthesis is used to convert the extracted text into speech by a speaker. The aim of the project is to convert an image into speech .An image is processed to identify the text in the image. Then the text is converted into speech. Here we are going to use OCR to extract the text from the image and then we will use Text To Speech(TTS) synthesis to convert the text into speech.

Key Words— ***Conversion, Open CV, OCR, TTS Systhesis,.***

I. INTRODUCTION

Languages are the oldest way of communication between human beings whether they are in spoken or written forms. In the recent era, visual text in natural or manmade scenes might carry very important and useful information. Therefore, the scientists have started to digitize these images, extract and interpret the data by using specific techniques, and then perform text-to-speech synthesis (TTS).[1] It is done in order to read the information aloud for the benefit and ease of the user. Text extraction and TTS can be utilized together to help people with reading disabilities and visual impairment to listen to written information by a computer system.

To help this type of technology we are proposing a device which helps in conversion of hard copy of text

DOI:10.9756/INTJECSE/V14I5.43 ISSN: 1308-5581 Vol 14, Issue 05 2022

which is inserted into the device will be converted to speech. We used OCR process which is nothing but called as the Optical Character Recognition.WE use OCR to extract the text from[2]

The image and the we will use the TTS Systhesis which is nothing but it is called as text to speech systhesis,by using this tts we will convert the text which is extracted from the image using OCR into the speech default in system. Most of these applications find the use of functions such as articulators, conversion from text to synthetic speech signals, language translation amongst various others. [3,4,]In our project, we'll be executing OCR and TTS Synthesis that are applied to achieve the concept of the project. Here we are going to use OCR to extract the text from the image and then we will use Text To Speech(TTS) synthesis to convert the text into speech.

OUR proposed system is tested on images representing different scenes ranging from documents to natural scenes. Promising results have been reported which prove the accuracy and robustness of the proposed algorithm and encourage its practical implementation in real world scenarios.

Rest of the paper is structured as follows: Section II covers the background of the research problem addressed in this paper and related methods. The proposed algorithm is presented in Section III followed by experimental analysis in Section IV. Results of the experimental analysis are discussed in section V. Conclusion and future prospects of this research are summarized in Section VI. Acknowledgement is in section VII and section VIII contains the References

II. LITERATURE SURVEY

2.1 Image Text to Speech Conversion Using OCR Technique in Raspberry Pi

Author: N. Kumari, J. MeghanaReddy

In this paper an innovative, efficient and real-time cost beneficial technique that enables user to hear the contents of text images instead of reading through them as been introduced. It combines the concept of Optical Character Recognition (OCR) and Text to Speech Synthesis (TTS) in Raspberry pi. This kind of system helps visually impaired people to interact with computers effectively through vocal interface. Text Extraction from color images is a challenging task in computer vision. Text-to-Speech is a device that scans and reads English alphabets and numbers that are in the image using OCR technique and changing it to voices. This paper describes the design, implementation and experimental results of the device. This device consists of two modules, image processing module and voice processing module. The device was developed based on Raspberry Pi v2 with 900 MHz processor speed. Text Extraction from color images is a challenging task in computer vision. Text-to-Speech is a device that scans and reads English alphabets and numbers that are in the image using OCR technique and changing it to voices

2.2 Design and Implementation of Speech Generation System using MATLAB

Author: Pooja Chandran, Aravind S, Jisha Gopinath and Saranya S S

In this paper the main idea is to recognize the text character and convert it into speech signal. The text contained in the page is first pre-processed. The pre-processing module prepares the text for recognition. Then the text is segmented to separate the character from each other. Segmentation is followed by extraction of letters and resizing them and stores them in the text file. These processes are done with the help of MATLAB. This text is then converted into speech

2.3 TEXT TO SPEECH SYSTEM USING OCR

Author: J. Gopinath, S. Aravind, P. Chandran

There are about 45 million blind people and 135 million visually impaired people worldwide. Disability of visual text reading has a huge impact on the quality of life for visually disabled people. Although there have been several devices designed for helping visually disabled to see objects using an alternating sense such as sound and touch, the development of text reading device is still at an early stage. Existing systems for text recognition are typically limited either by explicitly relying on specific shapes or colour masks or by requiring user assistance or may be of high cost. Therefore we need a low cost system that will be able to automatically locate and read the text aloud to visually impaired persons. The main idea of this project is to recognize the text character and convert it into speech signal. The text contained in the page is first pre-processed. The preprocessing module prepares the text for recognition. Then the text is segmented to separate the character from each other. Segmentation is followed by extraction of letters and resizing them and stores them in the text file. These processes are done with the help of MATLAB. This text is then converted into speech

2.4 OPTICAL CHARACTER RECOGNITION BASED TEXT TO SPEECH SYNTHESIS

Author: Yamini D.Patil, Nandkishor C.Patil.

Text to speech, there are many systems which convert normal language text in to speech. This thesis aims to study on speech synthesis technology using image recognition technology (Optical Character Recognition) to develop a cost effective user friendly image to speech conversion system using MATLAB for blind person. In this work we tried to make a system by which we can get the text through image and then speech through that text using MATLAB. The primary motivations are to provide users with a friendly vocal interface with the computer and to allow people with certain handicaps (such as blindness, dumbness, poor vision, visual dyslexia) to use the computer or to read any type of documents

2.5 A Image text to speech conversion in the desired language by translating with Raspberry Pi

Author: H. Rithika ; B. Nithya Santhoshi

In this paper the device basically can be used by people who do not know English and want it to be translated to their native language. The novelty component of this research work is the speech output which is available in 53 different languages translated from English. This paper is

based on a prototype which helps user to hear the contents of the text images in the desired language. It involves extraction of text from the image and converting the text to translated speech in the user desired language. This is done with Raspberry Pi and a camera module by using the concepts of Tesseract OCR [optical character recognition] engine, Google Speech API [application program interface] which is the Text to speech engine and the Microsoft translator. This relieves the travelers as they can use this device to hear the English text in their own desired language. It can also be used by the visually impaired. This device helps users to hear the images being read in their desired language.

IV SYSTEM ARCHITECTURE

System Architecture is a generic discipline to handle objects (existing or to be created) called "systems" in a way that supports reasoning about the structural properties of these objects. The system architecture is a response to the conceptual and practical difficulties of the description and the design of complex systems and the reliability of the this system is complex and stochastic [9,10,11,12,13,14]

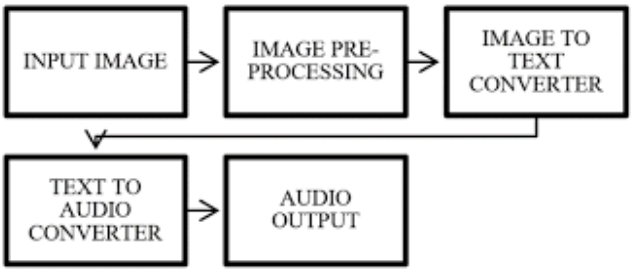

Fig 3 System architecture

V PROJECT DESCREPTION

Our project we proposed is for helping the people who are visually challenging. In our project we have proposed a system which can convert an image into text into speech .The OCR will help to recognize the character from an image from the back ground, then it extracted text is converted into speech by TTS(Text to speech) Synthesis And this process uses the incremental text-to-speech (TTS) method that performs synthesis in small linguistic units while maintaining the naturalness of output speech. Incremental TTS is generally subject to a trade-off between latency and synthetic speech quality. It is challenging to produce high-quality speech with a low-latency setup that does not make much use of an unobserved future sentence. To resolve this issue, we propose an incremental TTS method that uses a pseudo lookahead generated with a language model to take the future contextual information into account without increasing latency.

VI CONCLUSION AND FUTURE SCOPE

In this paper, the presented methodology is effectively able to help the blind people and the illiterates. Our project can be developed as many different models such as we can innovate the camera and the camera can be portable in the future.

And also our project can be innovated into the size of miniature so that the blind people can use the system freely. We hope this will help the blind people and in the future many innovations can be done and excute for the innovative people in

future.

The goal of speech recognition [15,16,17,18,19] is to transform the input speech feature vector sequence into a sequence of words using phonetic and linguistic information. Now a days these are progressing at an exponential rate in different sectors, such as banking, healthcare, government, security, construction, retail, etc. For the last few years, context-based understating is a breakthrough in voice assistance technology, which is now becoming a key component in people's lives

Nowadays, there is increasing demand of text information extraction from image. So, many extracting techniques for retrieving relevant information have been developed. Moreover, extracting text from the color image takes time that leads to user dissatisfaction e have proposed a method to extract the text from image which extracts text more accurately in this project[20-24].

VII ACKNOWLEDGEMENT

I am using this opportunity to express my gratitude to everyone who supported me throughout the course of research project report. We Would like to thank all our faculties, friends and family members who have directly and indirectly helped in easy completion of our project. Special thanks to Mrs. M. Anitha, Associate Professor, Department of Computer Science Engineering, Prince Dr. K. Vasudevan College of Engineering and Technology Chennai, Tamil Nadu, India for guiding us in completing this project successful.

VIII REFERENCES

1. S. Bangalore, V. K. Rangarajan Sridhar, P. Kolan, L. Golipour, and A. Jimenez, “Real-time incremental speech-to-speech translation of dialogs,” in Proc. Conf. North Amer. Chapter Assoc. Comput. Linguist. Human Lang. Technol., Montreal, Canada, Jun. 2012, pp. 437–445
2. K. Sudoh et al., “Simultaneous speech-to-speech translation system with neural incremental ASR MT, and TTS” 2020, arXiv:2011.04845.
3. M. Ma et al., “Incremental text-to-speech synthesis with prefix-to-prefix framework,” in Proc. Empirical Methods Natural Lang. Process., Online, Nov.

2020, pp. 3886–3896

4. J. Shen et al., "Natural TTS synthesis by conditioning WaveNet on mel spectrogram predictions," in Proc. IEEE Int. Conf. Acoust., Speech Signal Process., Apr. 2018, pp. 4779–4783.
5. Y. Cong, R. Zhang, and J. Luan, "PPSpeech: Phrase based parallel endto-end TTS system," 2020, arXiv:2008.02490.
6. K. Tokuda, H. Zen, and A. W. Black, "An HMM-based speech synthesis system applied to english," in Proc. IEEE Workshop Speech Synthesis, Santa Monica, U. S. A., Sep. 2002, pp. 227–230.
7. H. Zen, K. Tokuda, and A. Black, "Statistical parametric speech synthesis," Speech Commun., vol. 51, no. 11, pp. 1039–1064, 2009.
8. S Hemalatha, T Sunder Selwyn, Computation of mechanical reliability for Sub-assemblies of 250 kW wind turbine through sensitivity analysis, Materials Today: Proceedings, Vol. 46, pp 3180-3186, 2021.
9. T Sunder Selwyn, S Hemalatha, Condition monitoring and vibration analysis of asynchronous generator of the wind turbine at high uncertain windy regions in India, Materials Today: Proceedings, Vol. 46, pp3639-3643, 2021.
10. T Sunder Selwyn, S Hemalatha, Experimental analysis of mechanical vibration in 225 kW wind turbine gear boxMaterials Today: Proceedings, Vol. 46, pp 3292-3296, 2021
11. T. Sunder selwyn, R Kesavan, Vibration Analysis of a Constant Speed and Constant Pitch Wind Turbine, Springer-Lecture Notes in Mechanical Engineering-https://doi.org/10.1007/978-81-322-1007-8_40, 2012, pp 429- 443.
12. T Sunder Selwyn, R Kesavan, Reliability analysis of sub assemblies for wind turbine at high uncertain wind, Advanced Materials Research 433,2012, 1121-1125
13. T. Sunder Selwyn, R. Kesavan, Computation of availability and performance of wind turbine with markov analysis in India, Adv. Mater. Res. J. 488–489 (2012) 1702–1707
14. P.Bhagya divya, S.Shalini, R.Deepa, Baddeli sravya Reddy" Inspection of suspicious human activity in the crowd sourced areas captured in survillence cameras", International Research Journal of Engineering and Technology(IRJET),Vol-4,Issue-12,pp-802-806, Dec(2017), eISSN:2395-0056
15. Niruban, R.,SreeRenga Raja, T and Deepa,R (2015), "Similarity and Variance Of Color Difference Based Demosaicing", TELKOMNIKA Indonesian Journal of Electrical Engineering, DOI:10.11591/telkomnika.v13i2.7048, Vol.13, No. 02, pp. 238-246, February (Scopus Indexed). e-ISSN: 2302-9293.
16. Baddeli sravya reddy, R.Deepa, S.Shalini, P.Bhagya divya," A Novel Machine Learning Based Approach For Detection And Classification Of Sugarcane Plant Disease By Using Dwt", International Research Journal of

Engineering and Technology(IRJET),Vol-4,Issue-12,pp-843-846, Dec(2017),eISSN:2395-0056.

17. R. Niruban, R. Deepa, G.D. Vignesh, (2020), ”A Novel Iterative Demosaicing Algorithm Using Fuzzy Based Dual Tree Wavelet Transform”, Journal of Critical Reviews, Vol 7, Issue 9, pp.141-145, May, ISSN: 2394-5125.
18. Deepigka. M. S Deepa. R, Ashlin Lifty. S, ,Recognization and Systematization of MR Imagesusing K Means Clustering and DNN, International Journal of Innovative Technology and Exploring Engineering (IJITEE), Vol.9, Issue.6, pp 924-927.
19. K. K. Senthilkumar, G. Kalanandhini, A. R. Aravind, G. Vijayalakshmi, and J. Gayathri , "Image fusion based on DTDWT to improve segmentation accuracy in tumour detection", AIP Conference Proceedings 2393, 020120 (2022) https://doi.org/10.1063/5.0074183
20. J. Gayathri, K. K. Senthilkumar, G. Vijayalakshmi, A. R. Aravind, and G. Kalanandhini , "Multi-purpose unmanned aerial vehicle for temperature sensing and carbon monoxide gas detection with live aerial video feeding", AIP Conference Proceedings 2393, 020124 (2022) https://doi.org/10.1063/5.0074193
21. Subburam, S., Selvakumar, S. & Geetha, S. High performance reversible data hiding scheme through multilevel histogram modification in lifting integer wavelet transform. Multimed Tools Appl 77, 7071–7095 (2018). https://doi.org/10.1007/s11042-017-4622-0
22. Rajesh, G., Mercilin Raajini, X., Ashoka Rajan, R., Gokuldhev, M., Swetha, C. (2020). A Multi-objective Routing Optimization Using Swarm Intelligence in IoT Networks. In: Peng, SL., Son, L.H., Suseendran, G., Balaganesh, D. (eds) Intelligent Computing and Innovation on Data Science. Lecture Notes in Networks and Systems, vol 118. Springer, Singapore. https://doi.org/10.1007/978-981-15-3284-9_65
23. Kathiresan, S., & Mohan, B. (2020). Multi-Objective Optimization of Magneto Rheological Abrasive Flow Nano Finishing Process on AISI Stainless Steel 316L. Journal of Nano Research, 63, 98–111. https://doi.org/10.4028/www.scientific.net/jnanor.63.98
24. G. Indira, A. S. Valarmathy, P. Chandrakala, S. Hemalatha, and G. Kalapriyadarshini , "Development of an efficient inverter for self powered sand sieving machine", AIP Conference Proceedings 2393, 020144 (2022) https://doi.org/10.1063/5.0074347