



JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)

An International Scholarly Open Access, Peer-reviewed, Refereed Journal

TRANSPORTABLE CAMERA BASED TEXT TO SPEECH CONVERTER

¹Rajesh S, ²Swetha S, ³NandaKumar M, ⁴Sujish Kumar M

¹Assistant Professor, ²Student, ³Student, ⁴Student

¹Computer Science and Engineering,

¹Sri Ramakrishna Institute of Technology, Coimbatore, India

Abstract: This paper describes a framework for camera based mostly helpful device to assist the blind to scan text labels and merchandise packaging from hand-held objects. The projected system aims at planning an innovative system that is extremely useful for the physically challenged like blind persons to acknowledge the text from captured pictures. This paper has introduced an innovative, economical and period price helpful technique that permits user to listen to the contents of text pictures rather than reading through them. It combines the idea of Optical Character Recognition (OCR) and Text to Speech Synthesizer (TTS). This type of system helps visually impaired individuals to act with computers effectively through vocal interface. Text Extraction from color pictures may be a difficult task in pc vision. Text-to-Speech conversion may be a methodology that scans and reads English alphabets and numbers that square measure within the image mistreatment OCR technique and dynamic it to voices. This paper describes the planning, implementation and experimental results of the device. This device consists of 2 modules, image process module and voice process module.

Index Terms – Optical Character Recognition (OCR), Text to Speech Synthesizer (TTS), Image process Module, Voice process Module, Visually Impaired individuals.

I. INTRODUCTION

Every year the number of blind individuals square measure increasing because of eye diseases polygenic disorder, traffic accidents and alternative causes. Therefore, applications that give support to the blind individuals became a vital tool. Recent developments in pc vision, digital cameras, and pc create it potential to help these persons by developing camera-based merchandise that merge computer vision technology with alternative existing helpful merchandise like optical character recognition (OCR) systems. Text Reader for visually impaired person mistreatment camera module making certain movability is that the example created mistreatment the OCR and TTS Module and Python to scan the text from the handheld objects of the blind. This paper proposes an improved approach for text localization and extraction for detection of text areas within the pictures. The text size is a vital issue whose dimension ought to be properly elective to form the tactic additional general and insensitive to numerous font shapes and sizes. The Region of Interest is extracted from the blurred background then the text localization algorithmic program is applied to find and extract the text. Once extracting the text from the ROI, it's regenerated it into speech. It works additional with efficiency with Optical Character Recognition. Convolutional Recurrent Neural Network is projected for coaching the words severally. The experiment and coaching square measure performed on Synth 90k word dataset. Finally, mistreatment OCR and CRNN algorithmic program a projected model has been developed. Optical Character Recognition are often used wide in aid applications to help blind individuals. We've got projected a system that is employed to assist blind individuals to scan books, perceive totally different text on banners, photos and enormous ad boards. OCR includes primarily 3 parts the camera to capture the photographs, the programmable system to convert the captured camera into no matter format we wish and eventually the output system to point out the output of OCR. To assist the individuals additional effectively, have interaction with native and/or remote services text-to-speech (TTS) were 1st developed to help the visually impaired by giving a computer-generated spoken voice that might "read" text to the user Text detection and recognition in natural scene will provide valuable info for several applications. During this work, AN approach has been tried to extract and acknowledge text from scene pictures and convert that recognized text into speech. A text-to-speech (TTS) system converts traditional language text into speech. A Text-to-speech (TTS), as it's referred to as, is typically meant to assist sand-blind individuals. This task will undoubtedly be an empowering force in an exceedingly sand-blind person's life and may be validating in relieving them of their frustration of not having the ability to scan no matter they need therefore enhancing the standard of their lives.

II. PROBLEM STATEMENT

There are totally 285 million visually impaired individuals within the world. Of those populations, virtually half are blind. This range is increasing apace because of abnormalities attributable to genetic hereditary problems. The approach is largely to sense real time objects with utmost accuracy and offers output within the speech format. Recent advancements in pc vision and image process fields have created this situation a reality. So, this paper deals with serving to blind users with the most recent technologies. There's

an excellent ought to develop a text to speech convertor tool with easy human pc interface to satisfy desires of visually impaired individuals and additionally facilitate individuals with upset and to place foundation for facet applications. The TTS conversion tool will effectively address the wants of visually impaired individuals and customarily everyone within the country. Text-to-Speech system converts the text into speech output. The output is fed to an output device reckoning on the user's alternative. Output will be detected through headphones connected or through speakers of the system.

III. SCOPE OF THE PROJECT

- Increased net presence-Websites with TTS technology attract a number of the 774 million individuals worldwide with attainment problems and also the 285 million individuals with visual impairments. Also, speech enabling web page doesn't interfere with usability for those while not disabilities. It really aids all different populations, together with older users and foreign/nonnative speakers.
- Optimized development and maintenance- TTS technology that supports the foremost widespread platforms, is accessible each within the cloud and on premises, and is climbable in step with actual business desires, saves development and maintenance efforts.
- Save time and cash- With TTS technology that's web- or cloud-based on a SaaS (Software as a Service) platform, on-line content will quickly and simply be speech enabled, and maintenance is least.

IV. LITERATURE SURVEY

[1]. Improved text-detection methods for a camera-based text reading system for blind persons. Authors: Nobuo Ezaki, Marius Bulacu, Lambert Schomaker. IEEE in Proceedings of Eighth International Conference on Document Analysis and Recognition, pp 257 - 261 Vol. 1 ISSN:1520-5263, 2005. Automatic text recognition from natural images receives a growing attention because of potential applications in image retrieval, robotics and intelligent transport system. Camera-based document analysis becomes a real possibility with the increasing resolution and availability of digital cameras. Our research objective is a system that reads the text encountered in natural scenes with the aim to provide assistance to visually impaired persons. In the case of a blind person, finding the text region is the first important problem that must be addressed, because it cannot be assumed that the acquired image contains only characters. This paper describes a new text detection method geared for small text characters. This method uses Fisher's Discriminant Rate to decide whether an image area should be binarized using local or global thresholds. Fusing the new method with a previous morphology based one yields improved results.

[2]. Text Detection and Recognition with Speech Output for Visually Challenged Person: A Review. Authors: Ms.Rupali D. Dharmale Int. Journal of Engineering Research and Applications ISSN : 2248-9622, Vol. 5, Issue 3, (Part -3) March 2015, pp.84-87. Reading text from scene, images and text boards is an exigent task for visually challenged persons. This task has been proposed to be carried out with the help of image processing. Since a long period of time, image processing has helped a lot in the field of object recognition and still an emerging area of research. The proposed system reads the text encountered in images and text boards with the aim to provide support to the visually challenged persons. Text detection and recognition in natural scene can give valuable information for many applications. In this work, an approach has been attempted to extract and recognize text from scene images and convert that recognized text into speech. This task can definitely be an empowering force in a visually challenged person's life and can be supportive in relieving them of their frustration of not being able to read whatever they want, thus enhancing the quality of their lives.

[3]. Speech to text and text to speech recognition systems-A review. Authors: Ayushi Trivedi, Navya Pant, Pinal Shah, Simran Sonik and Supriya Agrawal. IOSR Journal of Computer Engineering (IOSR-JCE) e-ISSN: 2278-0661, p-ISSN: 2278-8727, Volume 20, Issue 2, Ver. I (Mar.- Apr. 2018), PP 36-43. In present industry, communication is the key element to progress. Passing on information, to right person, and in the right manner is very important, not just on a corporate level, but also on a personal level. The world is moving towards digitization, so are the means of communication. Phone calls, emails, text messages etc. have become an integral part of message conveyance in this tech-savvy world. In order to serve the purpose of effective communication between two parties without hindrances, many applications have come to picture, which acts as a mediator and help in effectively carrying messages in form of text, or speech signals over miles of networks. Most of these applications find the use of functions such as articulatory and acoustic-based speech recognition, conversion from speech signals to text, and from text to synthetic speech signals, language translation amongst various others. In this review paper, we'll be observing different techniques and algorithms that are applied to achieve the mentioned functionalities.

[4]. Using speech and touch to enable blind people to access schematic diagrams science direct, Journal of Network and Computer Applications, 2019N. Authors: P. Blenkhorn, D.G. Evans, Spandana and K. Prabhakara Rao et al (2016). This paper describes a framework for camera based assistive device to help the blind to read text labels and product packaging from handheld objects. The portable system which captures the images and text written which are placed in front of the Pi camera can be read out or announced out using earphones or speakers interfaced to the Raspberry Pi processor. The Raspberry Pi processor takes the responsibility for authentication of the image which is fed as input to it and also alerts the blind person through voice messages using earphones or speakers. The proposed system aims at designing an innovative system which is very helpful for the physically challenged like blind persons to recognize the text from captured images and also from messages reading, which are texted from android mobile and raspberry.

[5]. Wearable obstacle avoidance electronic travel aids for blind. Authors: D. Dakopoulos and N. G. Bourbakis, A survey in IEEE transactions on systems. They have proposed a smart spec for blind persons which performs text detection and the detected text is converted into voice output. A spectacle having an inbuilt camera is used to capture the text image from the printed text. Tesseract, an open-source optical character recognition (OCR) engine is used to extract the text from the image. The detected text is converted into speech output using compact open- source software called e Speak and the output is fed to an audio amplifier before it is read out. Raspberry pi is used as an interface between the software's, keyboard and the image processing result.

[6]. Texture Based Text Detection in Natural Scene Images: A Help to Blind and Visually Impaired Persons. Authors: Shehzad Muhammad Hani, Lionel Prevos. Conference & Workshop on Assistive Technologies for People with Vision & Hearing Impairments Assistive Technology for All Ages CVHI, 2007. Their proposed method involves four steps detection of an object, localization of the text extraction of the text and text to speech conversion. The Region of Interest is extracted from the cluttered background and then the text localization algorithm is applied to locate and extract the text. After extracting the text from the ROI, it is converted it into speech. To support requirement of TensorFlow python 3.5 is taken into account. Raspberry pi with camera attached to goggles and earphones connected and a battery to make raspberry on are equipment's that makes overall project portable. As soon as the raspberry pi is turned on the program will start. Indication for user to place an object near camera is first step. An image will be captured and using OCR for pre-processing plus CRNN for training the textual part is passed over it. The recognized text output is turned into speech through pyttsx3.

[7]. Improved text-detection methods for a camera-based text reading system for blind persons. Authors: Nobuo Ezaki, Marius Bulacu, Lambert Schomaker. IEEE in Proceedings of Eighth International Conference on Document Analysis and Recognition, pp 257 - 261 Vol. 1 ISSN:1520-5263, 2005. Automatic text recognition from natural images receives a growing attention because of potential applications in image retrieval, robotics and intelligent transport system. Camera-based document analysis becomes a real possibility with the increasing resolution and availability of digital cameras. Our research objective is a system that reads the text encountered in natural scenes with the aim to provide assistance to visually impaired persons. In the case of a blind person, finding the text region is the first important problem that must be addressed, because it cannot be assumed that the acquired image contains only characters. This paper describes a new text detection method geared for small text characters. This method uses Fisher's discriminant Rate to decide whether an image area should be binarized using local or global thresholds. Fusing the new method with a previous morphology based one yields improved results.

[8]. Image recognition for visually impaired people by sound. Authors: Krishnan K.G., Porkodi C.M., Kanimozhi K. IEEE International Conference on Communications and Signal Processing (ICCSP), Melmaruvathur. Page(s):943 – 946, 3-5 April 2013. The technique used here is TTS, STT and IVR. They suggested that for STT conversion the audio message should first be recorded and then be converted to text form and for TTS conversion the text should be translated to the audio and then play the audio message to the user. The proposed idea of an email system based on voice, makes use of 3 modules, namely: STT conversion, TTS conversion and IVR method. The artificial production of speech-like sounds has a long history, with documented mechanical attempts dating to the eighteenth century. The synthesis of speech by electronic means is obviously much more recent, with some early work by Stewart [1] that allowed the production of static vowel sounds. Text-analysis for speech synthesis has a much shorter history.

[9]. Efficient Portable Camera Based Text to Speech Converter for Blind Person. Authors: Trupti Shah, Sangeeta Parshionkar, International Conference on Intelligent Sustainable Systems (ICISS 2019), IEEE Xplore Part Number: CFP19M19-ART; ISBN: 978-1-5386-7799-5. Text Reader for Blind Person using camera module ensuring portability is the prototype made using the Raspberry pi 3b and Python to read the text from the handheld objects of the blind person. This paper proposes a better approach for text localization and extraction for detection of text areas in the images. The text size is an important factor whose dimension should be properly elected to make the method more general and insensitive to various font shapes and sizes. The proposed method involves four steps detection of an object, localization of the text extraction of the text and text to speech conversion. The Region of Interest is extracted from the cluttered background and then the text localization algorithm is applied to locate and extract the text. After extracting the text from the ROI, it is converted it into speech. It works more efficiently with Optical Character Recognition. Convolutional recurrent neural network is proposed for training the words separately. The experiment and training are performed on Synth 90k word dataset. Finally using OCR and CRNN a proposed model has been developed.

[10]. Assisting the Speech Impaired People Using Text-to-Speech Synthesis. Authors: Ledisi G, Kabari Ledum F. Atu. International Journal of Emerging Engineering Research and Technology, Volume 3, Issue 8, August, 2015, PP 214-224 ISSN 2349-4395 (Print) & ISSN 2349-4409. Many people have some sort of disability which impairs their ability to communicate, thus with all their intelligence they are not able to participate in conferences or meeting proceedings and consequently inhibit in away their contributions to development of the nation. Work in alternative and augmentative communication (AAC) devices attempts to address this need. This paper tries to design a system that can be used to read out an input to its user. The system was developed using Visual Basic6.0 for the user interface, and was interfaced with the Lem out & Hauspie True Voice Text-to-Speech Engine (American English) and Microsoft agent. The system will allow the users to input the text to be read out, and also allows the user to open text document of the ich Text Formant (.rtf) and Text File format (.txt) that have been saved on disk, for reading.

V. PROPOSED SYSTEM

The planned system consists of 2 main modules, the image process module and also the voice process modules

5.1 Image to Text Conversion Module: OCR

OCR analyses the patterns of light and dark that compose the letters and numbers to show the scanned image into text. OCR systems have to be compelled to acknowledge characters in numerous fonts, therefore rules area unit applied to assist the system match what it sees within the image to the proper letters or numbers. whereas early OCR systems were designed to figure with one specific font, that was specially created for the aim, some fashionable OCR systems will even acknowledge people's handwriting. This technology is termed Intelligent character recognition (ICR). For OCR to figure optimally, it's vital that you just scan the clearest potential version of the document. Blurred text or marks on the copy will produce errors. OCR programs acknowledge the text character by character however the end result is therefore quick on be fast. you'll check for errors as you go or at the tip of the method, and a few programs have automatic error detection.

5.2 Text to Voice Conversion Module: pyttsx3

Speech synthesis is that the artificial production of human speech. A computing system used for this purpose is termed a speech pc or speech synthesizer, and may be enforced in package or hardware merchandise. A text-to-speech (TTS) system converts traditional language text into speech. Associate in Nursing application invokes the `pyttsx3.init()` plant perform to urge a relevance a `pyttsx3`, engine instance. throughout construction, the engine initializes a `pyttsx3.driver`. Driver Proxy object to blame for loading a speech railroad engineer implementation from the `pyttsx3.driver's` module. After construction, an application uses the engine object to register and unregister event callbacks; produce and stop speech; get and set speech engine properties; and start and stop event loops.

5.2 Block Diagram

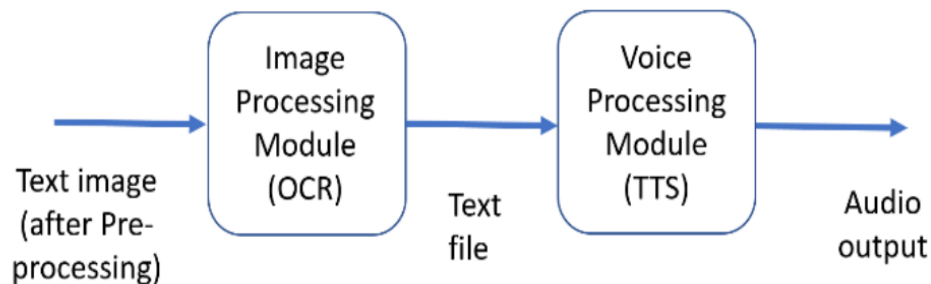


Fig 5.2 Block diagram of proposed System

VI. METHODOLOGY

Before feeding the image to OCR, the form is changed to binary image to extend the image recognition accuracy. Image binary conversion is finished by Image magic software package, that is another opensource tool for image manipulation. The output of OCR is that the text, that is hold on in an exceeding file(speech.txt). It wants some supporting condition so as get the marginal defect. Voice process module changes the text into sound and processes it with specific characteristics in order that the sound will be understood.

Initially the step is pre-processing the image before text will be force, the image has to be massaged in sure ways in which to create extraction easier and a lot of doubtless to succeed. This can be known as pre-processing, and completely different software package solutions use different mixtures of techniques.

Secondly, processing the image. First thing is, the OCR method tries to determine the baseline for each line of text within the image (or if it absolutely was zoned in pre-processing, it'll work every zone one at a time). Every known line of characters is handled one by one. For every line of characters, the OCR software package identifies the spacing between characters by searching for vertical lines of non-text pixels (which ought to be obvious with correct binarization). Every chunk of pixels between these non-text lines is marked as a "token" that represents one character. Hence, this step is termed tokenization.

Finally, `pyttsx3` could be a text-to-speech conversion library in Python. In contrast to various libraries, it works offline and is compatible with each Python two and three. Associate degree application invokes the `pyttsx3.init()` plant operate to urge a respect to a `pyttsx3`.

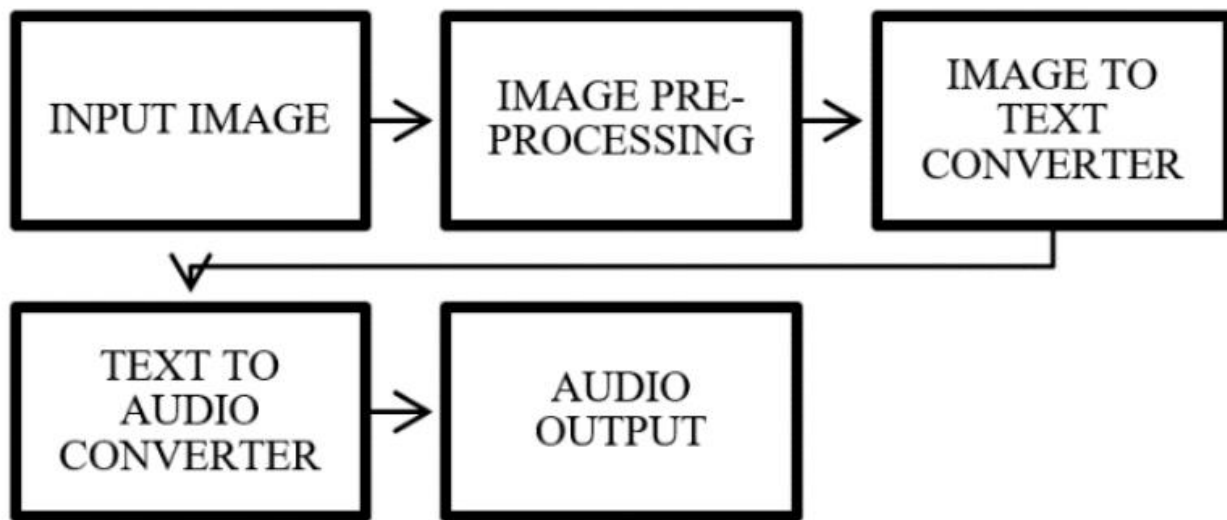


Fig 6.1 Text to Audio Conversion Methodology

VII. APPLICATIONS AND ADVANTAGES

7.1 Advantages

- It bridges the gap between normal people and visually impaired people.
- It normalizes and enhance their day-to-day life activities.
- Provides nearly natural sound.
- It reduces human effort in case of any application.

7.2 Applications

- Converts text into spoken voice output.
- It is mainly designed for blind people.
- It helps small kinds to improve word recognition, remember information while reading.
- It can be used as announcing device in schools/colleges and in other places.

VIII. REQUIREMENT SPECIFICATIONS

8.1 Hardware Requirements

- RAM: 8gb
- Webcam: 1080p HD
- Flash Drive Users: USB
- Speakers: USB speakers

8.2 Software Requirements

- Operating System: Windows
- Text Files: vs code
- Language: python

IX. RESULTS

The following figures describes the Input images and the corresponding recognized text.

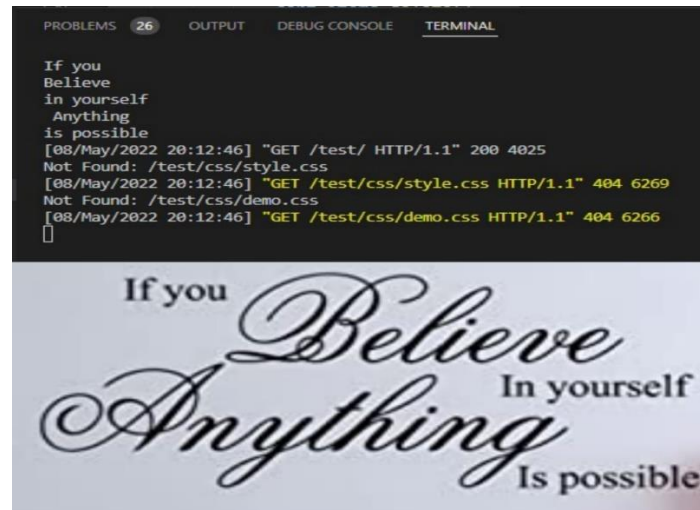


Fig 9.1 Input with Output Image



Fig 9.2 Input with Output Image



Fig 9.3 Input with Output Image

X. CONCLUSION AND FUTURE WORK

This paper suggests an approach to achieve portable camera-based text reading system. By using integration of image processing, speech processing it is possible to develop cost effective and user-friendly real-time application. Combination of OCR and CRNN overcomes drawback of individual and gives better result which successfully recognizes text image and convert it into speech for blind person. This proposed system presents the development of the work by implementing OCR and pyttsx3 using for Blind People, considering OCR and TTS stages, to create an application that was gradually improved and refined over the work. An analysis was

made regarding the OCR and TTS technologies that were used in the development of the application, in order to know the methods behind those, and to understand in greater detail the mechanisms that perform optical character recognition on images and speech synthesis of texts. The work consisted of the construction of an application composed by several parts, integrating the system of image capture by the web camera, which is used by an OCR framework for recognition of its text, which is then synthesized through a process of TTS. Optimizations carried out for improving outcomes resulted in a more efficient application, capable of responding to the challenge set by the theme of the work a camera that reads texts for the blind.

Through translation tools, one converts the text to the desired language and then again by using the TTS speech recognition tool can convert that changed text into audio. This idea will tend to anew technology that involves neuron based visual system which can give information directly to damaged neurons in mind in disabled people.

REFERENCES

- [1] Dr.Anto Bennet, M , Sankaranarayanan S, Ashokram S ,Dinesh Kumar T R, “Testing of Error Containment Capability in can Network”, International Journal of Applied Engineering Research, Volume 9, Number 19 (2014) pp. 6045-6054.
- [2] Ayushi Trivedi, Navya Pant, Pinal Shah, SimranSonik and Supriya Agrawal, “Speech to text and text to speech recognition systems”, IOSR Journal of Computer Engineering (IOSR-JCE) e-ISSN: 2278-0661, p-ISSN:2278-8727, Vol-20, Issue-2, Ver-I, March.April-2018.
- [3] Kamesh DBK, Nazma SK, Sastry JKR, Venkateswarlu S. Camera based text to speech conversion, obstacle and currency detection for blind persons. Indian Journal of Science and Technology. 2016 Aug; 9(30).
- [4] Krishnan K.G., Porkodi C.M., Kanimozhi K. “Image recognition for visually impaired people by sound”. IEEE International Conference on Communications and Sign Processing (ICCSP), Melmaruvathur. Page(s):943 – 946, 3-5 April 2013.
- [5] Hussain Rangoonwala, Vishal Kaushik, P Mohith and Dhanalakshmi Samiappan “Text to speech module International Journal of Pure and Applied Mathematics” Vol-115 No. 6 2017.
- [6] S. Venkateswarlu¹, D. B. K. Kamesh¹, J. K. R. Sastry and Radhika Rani, “Text to speech conversion”, Indian Journal of Science and Technology”, Vol-9(38), October2016
- [7] Gordana Laštovička-Medin, Itana Bubanja, “Hardware approach of text-to-speech in embedded applications” Kaladharan N, An English text to speech conversion system, International Journal of Advanced Research in Computer Science and Software Engineering, Vol-5, Issue-10, October-2015.
- [8] Poonam. S. Shetake, S. A. Patil and P.M Jadhav “Review of text to speech conversion methods”, International Journal of Industrial Electronics and Electrical Engineering, Vol-2, Issue-8, August-2014.
- [9] Kaveri Kamble and Ramesh Kagalkar, “Translation of text to speech conversion for Specific language”, International Journal of Science and Research”, Impact Factor (2012): 3.358 Vol- 3 Issue-11, November 2014.
- [10] Kumar Patra, Biplab Patra, Puspanjali Mohapatra, “Text to speech conversion with phonematic concatenation”, International Journal of Electronics Communication and Computer Technology “(IJECCCT) Vol- 2 Issue-5, September-2012, ISSN:2249-7838.
- [11] M. Jaderberg, K. Simonyan, A. Vedaldi, and A. Zisserman. Reading text in the wild with convolution neural networks. IJCV, 2015. 1,2,3,6,7.
- [12] K Kim, K Jung, et al. Texture-based approach for text detection in images using support vector machines and continuously adaptive mean shift algorithm, IEEE Trans. Pattern Anal. Mach. Intell, December 2003; 25: 1631–1639.
- [13] B. Su and S. Lu.Accurate scene text recognition based on recurrent neural Network. in ACCV,2014.
- [14] Pere Pujol Marsal, Susagna Pol Font, Astrid Hagen, H. Bourlard, and C. Nadeu, Comparison and Combination Of Rasta-Plp And Ff Features In A Hybrid Hmm/Mlp Speech Recognition System, Speech and Audio Processing, IEEE Transactions on Vol.13, Issue: 1, 20 December 2004.
- [15] Mathias De Wachter, Mike Matton, Kris Demuyne, Patrick Wambacq, Template Based Continuous Speech Recognition, IEEE Trans. On Audio, Speech Language Processing, vol.15, issue 4, pp 1377-1390, May 2007.