International Journal of Emerging Trends & Technology in Computer Science (IJETTCS)
Web Site: www.ijettcs.org Email: editor@ijettcs.org, editorijettcs@gmail.com
Volume 11, Issue 4, July – August 2022 ISSN 2278-6856

Image to Speech Converter using Machine Learning Techniques for Divyanga

Mallamma V Reddy[1], NaveenKumar T R[2] and Mallikarjun Hosur[3]

[1,2,2]Rani Channamma University P.G Campus, Vidyasangama, Belagavi
591156, Karnataka

Abstract: ***This paper focuses on assisting blind, dumb and deaf people. OCR and GTTSC are used which helps to hear the voice output by uploading text images and extracts the text letters from the images.***

Keywords: Image Processing, Optical Character Recognition (OCR), Text-to-Speech (TTS), Voice Processing.

1. INTRODUCTION

An image is a visual representation of a picture, real object, image consist of rectangular array of dots called pixel. An image that contains only alphabets, words, sentences is called textual image. Speech is a type of computing technology that enables an electronic device to recognize, analyze and understand spoken word or audio. This is typically done by inputting digital sound signals.

The primary goal of this project is to develop communication for those one who are blind, deaf, & dumb development of information and communication technology we now live in a digital age. For those who are deaf, and/or blind, sign language is an expressive means of communication between the two groups. Braille is the most common kind of reading material for the blind. Just to read, a person must learn Braille. 285 million individuals in the globe have some kind of vision impairment. In addition to the 39 million people who are visually impaired, there are 246 million people who have poor vision. About 9.1 billion individuals in the world are deaf and dumb. In their everyday lives,they encounter a number of communication difficulties because of their limitations, those who are deaf or dumb are mostly excluded from participating in society. The rest of society unintentionally treats them in a strange manner. Normal and stupid persons alike may express themselves naturally and expressively using sign language. In order to facilitate communication between the deaf and mute, the sign language translation system translates conventional sign language into spoken language.
To remedy this, we've created a brand-new technology called "Smart communication." So that persons with and without disabilities may converse with one other without difficulty as a result, our assistive reading gadget for the blind is an absolute requirement. Here, you'll learn how to utilize the System as a primary unit with a built-in camera that can scan any written document and use Optical Character Recognition (OCR) to turn the picture into a digital text. Once the digital text has been converted into a synthetic voice, we employ a text-to-audio technology (Image to Speech). As soon as the other person talks, it is presented as a message for the deaf individuals to hear. i.e. (speech to text) and also Kannada language, and also, we had to include hand motions that deaf people can easily comprehend well, in order to make it more accessible to them. Rather of using sign language, the deaf and dumb use textualcommunication. This leads to the complexity and is not realistic. They have a hard time communicating face-to-face. Though many individuals are dumb or deaf, there is virtually little study being done to break through the communication barrier. We have developed a technology that makes it easier for hearing and deaf individuals to converse with one another. We have developed a new method for blind individuals who utilize the braille system, and they'llbe able to effortlessly read any normal-sized textbooks using it. In order to communicate with the outside world, persons who are deaf or hard of hearing have a hard time since they do not understand sign language. They may effortlessly communicate their words or information to another individual by using our technology.

2. LITERATURE SURVEY

Blindness is a medical term for the inability to perceive objects in the environment owing to a variety of physiological or neurological issues. Partial blindness occurs when the optic nerveor visual center of the eye does not fully integrate throughout its development. The lack of light perception in absolute blindness is called total blindness. Using a virtual eye that is simple, inexpensive, and user-friendly, blind and visually impaired people mobility may be enhanced in certain areas. For those who are blind or visually handicapped in third-world nations, this technology offers a novel, low-cost, and simple-to-use option. A camera-based assistive text reading framework is suggested and developed to enable blind people read text labels and product packaging from hand-held items in their everyday lives. An effective and efficient motion-based approach based on learning gradient features of stroke orientations and using an Ada boost model for the distribution of edge pixels is initially described. The binarized text characters in the localized text sections are identified using off-the-shelf OCR [1] software. The detected text codes are spoken to the person who is blind.

People with disabilities are having a difficult time keeping up with the rapid pace of technological change in our culture. Handicapped individuals must have access to communication technology. People who are deaf or dumb often communicate by sign language, but they often have trouble doing so with others who do not know the sign language. Normal and visually impaired, dumb, deaf persons can communicate naturally and expressively using sign language (information majorly conveyed through the hand gesture). As a result, we need the services of a translator in order toeffectively interact with them. As a result of the sign language translation system, it is now simpler for persons who are deaf or hard of hearing to communicate with others. To address this, we'll need to create a system that translates the voice of a typical person into text, and then displays the matching motion on the screen. Deaf-mute [2] people and non-disabled people might communicate with each other using a two-way gadget.

Opinions, and expressions between two or more people is called communication. Non-verbal communication, such as hand gestures, is just as important as verbal communication when it comes to human interaction. Sign language is a more formalized type of hand gesture communication. Each letter of the English alphabet has a unique symbol in this language. This language is used by those who are visually impaired such as the deaf and the dumb. Designing a system to communicate with the outside world is the goal of this project. [3] A real-time system to recognize significant shapes created by hands is being developed in this project, which takes into account the fact that, in most situations, everyone has the same hand shape with four fingers and a thumb.

This application makes use of pre-existing techniques OCR, gTTS, that work well. The real-time gesture to text conversion is the system's most important function. Gestural data is extracted from dataset, matched to a speaker and converted into speech. Various image processing methods, including histogram matching, bounding box calculation, skin color segmentation, and region growing are all used in the process of extracting gestures from images. Feature point matching and correlation-based matching may be used for Gesture matching. On the basis of the technologies utilized for gesture extraction and matching, we have come up with four possible strategies. They are also evaluated for their efficiency and accuracy in a comparative study.[4] Text to gesture conversion and text-to-voicing are also included in this software.

Most persons who are deaf or hard of hearing utilize sign language for communication, but they have a hard time speaking with others who do not. As a result, communication between the two groups is hindered. Efforts are underway to reduce this barrier to communication. The primary goal of the proposed project is to create a low-cost system that uses Smart Gloves to provide voice to the voiceless. The proposed work utilizes a flex sensor and microcontroller to translate sign language image into text [5] and audio. It implies that employing smart gloves, communication between two distinct cultures will not be a barrier.

People who are deaf or dumb have difficulty conversing with others and expressing themselves correctly. As a result, they encounter a wide range of challenges in this area. They utilize sign language to communicate with one other and convey their thoughts and feelings. This necessitates the use of a professional translation. There is a lot of researches going on out there for the deaf and dumb, and they are doing a good job at it. When it comes to getting to know them, an effective method must be put in place. Deaf and mute people may communicate using this framework [6]. It would use sign language as an input and output the results in text and audio formats, respectively.

3.METHODOLOGY

Using OCR, it extracts the text from the picture in this module, which accepts a text image as input. GTTS may be used to turn text into speech. An OCR is used in the system. We are utilizing a camera to take an image of the text. The Picture processing portion receives the image captured by the camera as shown in figure 3.1.

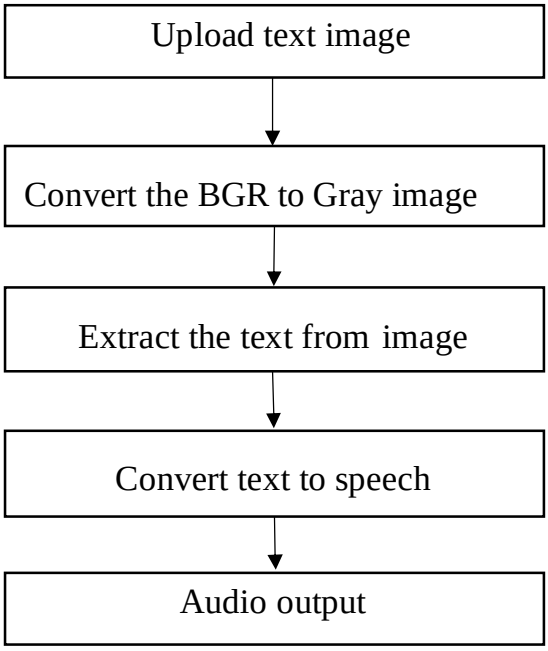


Figure 3.1 Block Diagram of Image to Speech

The processed image output is sent through a filter to remove noise signals from the picture. The edge reduction device receives the filtered output. The background separation unit receives the output. Eventually, the OCR

receives the output that it has been waiting for. The OCR's result is sent to the system for further processing. An audio signal is generated by the raspberry-pi based on the visual content. The figure 3.1 process with the following steps:

Capture image:
An active webcam may collect up to 30 frames per second of a picture at full resolution. Any video device, including USB camera's, capture cards that are attached to analog cameras, TV-boards, and camcorders that are connected to and from network cameras, may be used to capture the picture for this purpose. As soon as the software detects movement in a predetermined region, the application also has some extra features, such as the ability to add text captions and image logos to photographs, and the ability to alter the frame rate, picture size, and quality of each video frame.

Processing Image:
Web-cam refers to a video camera that transmits images in real time to a computer network through a computer connection. Computers may capture video streams and then preserve, watch or even transfer the "caught" images to other networks through systems like the internet or by email as attachments.

It is possible that the video stream may be preserved and watched or delivered to a distant place when the it is sent there. In contrast to IP cameras, which need an Ethernet or Wi-Fi connection, Webcams may either be connected through a USB cable or built into computer hardware such as laptops and desktops.

Filtering:
The method of altering or improving a picture via the use of filters. Example: the filter is the sole way to change an image's appearance by emphasizing or removing particular elements. Image processing is where filtering, which smooths, sharpens, and enhances the image's edge. An algorithm is applied to the values of the nearby pixels and those values decide which pixels in the output picture corresponds to those in the input image. The pixels in the immediate vicinity of a given pixel are defined by its position in relation to those pixels. When pixels in the input's immediate vicinity are used to create the value of an output pixel, this method is known as linear filtering.

Edge Detection:
Edge detection is a key procedure in image processing. In order to limit the quantity of data (pixels) that must be processed, edge detection is quite beneficial. Edgedetection methods based on the gradient (Sobel - first order derivatives) and the Laplacian (second order derivatives) (2nd order derivative, so it is extremely sensitive to noise). The eventual aim of both techniques is edge detection, which is accomplished by convolution of both schemes. The computer has a tough time reading the picture correctly since the pixel density only fluctuates bruptly at the image's boundaries. The pixel intensity in the edge's is the pixel intensity on the border between the two parts with differing pixel intensities.

Optical Character Recognition:
Optical character recognition is used to extract the text from the picture (OCR). Pattern recognition, artificial intelligence, and computer vision all play a role in OCR research. A digital text or computer format text is created by converting images of typed, handwritten, or printed text into digital text. In the past, OCR had to be taught for each character in a text with its own typeface.

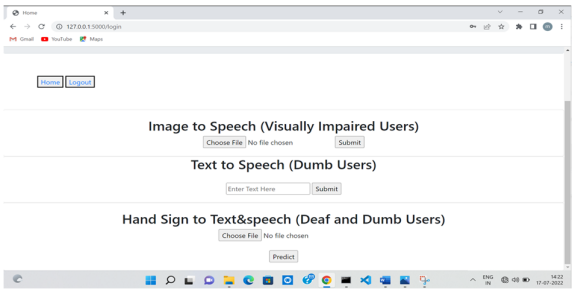
Speech Output:
Speech synthesis refers to the technique of using a computer to turn text into audio. Speech synthesis is accomplished by the use of a text-to-speech system (TTS). The front end and the rear end make up a TTS. Text is transformed into a symbol, such as a number, by the front end. It is given a phonetic to each symbol. The phonetic is then turned into sound at the backend. Festival TTS has been utilized in our research paper. For free source TTS is best technology. A broad range of accents and languages are supported, including American English, english language has been used.

ALGORITHM FOR IMAGE-TO-SPEECH

Step 1: Start
Step 2: Upload the required image.
Step 3: Convert image to text using Tesseract OCR.
Step 4: Split the text into paragraph.
Step 5: Text is displayed on the screen.
Step 6: Convert text to speech using e-speak synthesizer.
Step 7: Audio Signal Voice is generated.
Step 8: Stop

4. EXPERIMENTAL SETUP AND RESULT

The Application is built by using front as Python Programming Language and backend as file system.A home page for communication between a normal person and a person which is for blind or deaf or dumb is depicted in the below "Fig. 2" and "Fig. 3", which includes three modules: Image-to-speech (ITS), Text-to-speech (TTS), and hand sign intern classified as Hand sign to text and speech conversion for the deaf and dumb user.



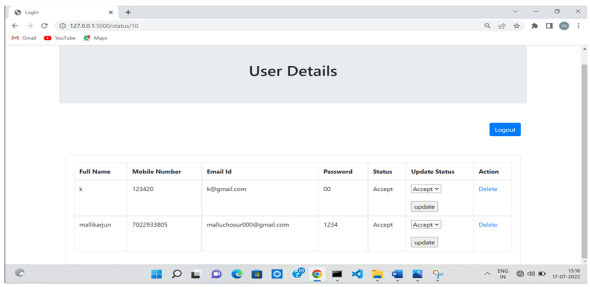


Figure 2. Smart communication for Blind, Deaf and Dumb Home Page and Admin page

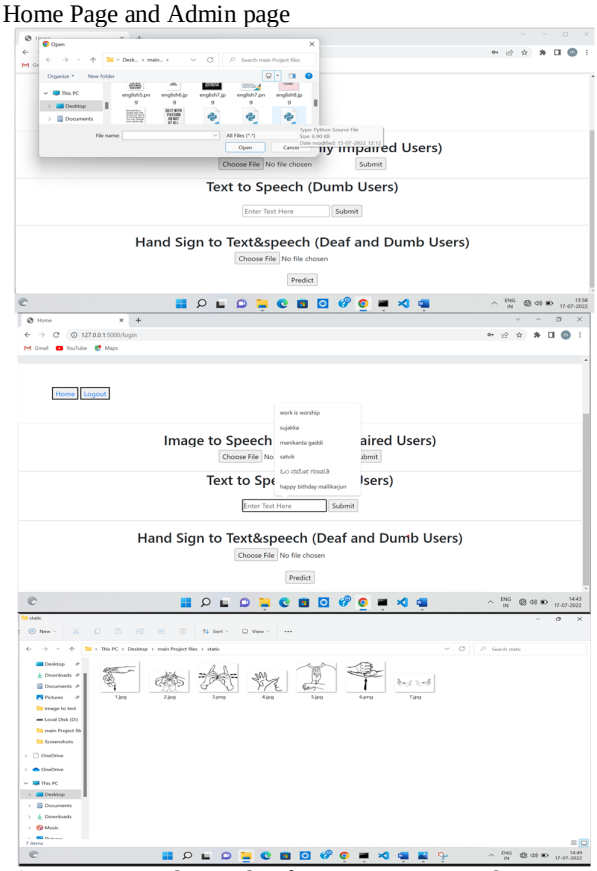


Figure 3. Sample results for Image to Speech, Text to Speech and Hand Gesture Data sets

5. CONCLUSION

This research focuses on visually impaired people that may read literature that is not printed in braille without feeling like they are losing out on the experience. It is possible to get a high-quality OCR input by doing picture preprocessing and then removing the relevant text from the background. The result of the OCR is transferred to the TTS engine, which generates the speech. A battery may be used to power up the system in order to make the gadget more portable. Instead of relying on still photos, the next step might be to build gadgets that can identify and extract text from movies.

Using this technology, the deaf and dumb people will be more connected to the restof society, making it possible for them to lead more normal lives. For the blind, deaf, and dumb, the technology converts text and images into voice, speech to text, and hand movements into text (in the form of Kannada words). For the blind, deaf, and dumb persons, we created a single, small prototype. It is small size and light weight, this gadget has the benefit of being readily taken about. It is a language-independent system that may be used as a smart assistant for persons with disabilities to communicate with others.

6. FUTURE ENHANCEMENT

The following are some potential future developments that may be linked to this research paper: Gesture control of the alphabets and numerals may be added to the system in the future and it is also possible to enter video, which is broken down into frames and then turned into text. Next Sign language may also have grammatical structure added to it and the system may be made more convenient if it is integrated with a mobile device. In order to help the visually impaired, we have developed a tool that translates any hand-written information into an audio signal that these individuals can hear.

REFERENCES

[1] Asha G. Hagargund: Title Image to SpeechConversion for Visual Impaired (International Journal of Latest Research in Engineering and Technology Volume-2, Issue-1, March 2017.)

[2] Shraddha R. Ghorpade, Surendra K. Waghmare, “Full Duplex Communication System for Deaf & Dumb People,” International Journal of Emerging Technology and Advanced Engineering (IJETAE), Volume 5, Issue 5, May 2015, ISSN 2250-2459.

[3] Er. Aditi Kalsh, Dr. N.S. Garewal, “Sign Language Recognition System,” International Journal of Computational Engineering Research (IJCER), Volume 03, Issue 6, June 2013.

[4] Ashish Sethi, Hemanth’s, Kuldeep Kumar, Bhaskar Rao N, Krishnan R, Sign Pro-An Application Suite for Deaf and Dumb, IJCSET —May 2012— Vol 2, Issue 5,1203-1206.

[5] Pallavi Varma, Shomi. S. L, Richa Priyadarshini,” Design of Communication Interpreter for Deaf and Dumb person” in Jan-2015.

[6] Anchal Sood, Anju Mishra “AAWAAZ: A Communication System for Deaf and Dumb” International Journal of Emerging Technology and Advanced Engineering. ISSN 2250-2459, Volume 2, Issue 5, May 2012.

[7] Prashant G. Ahire, Kshitija B. Tilekar, Tejaswi A. Jawake, Pramod B. Warale “Two Way Communicator between Deaf and Dumb People and Normal People “International Journal of Computer Technology and Electronics Engineering (IJCTEE) Volume 2, Issue 1.

[8] Kavita Sharma and Prateek Hakar "Fingertip Detection for Hand Pose Recognition," International Journal on Computer Science and Engineering (IJCSE), Vol. 4No. 03 March2012, ISSN: 0975-3397.
[9] Puneet Kaur, Bhupender Singh and Neha Kapur Video Gesture Classification Using Fourier Descriptors and General Fuzzy Min Max Neural Network, International Journal of Emerging Technology and Advanced Engineering. ISSN 2250-2459, Volume 2, Issue 5,2012.
[10] Kuldip K. Paliwal Sign Language Recognition for Deaf and Dumb International Journal of Advanced Research in Computer Science and Software engineering Volume 3, Issue 9, September 2013 ISSN: 2277 128.