Portable Camera Based Assistive Text and Label Reading From Hand-Held Object for Visually Impaired Person

Gagan Vaidya[1], Praful Mool[2], Vaishnavi Khobragade[3], Harshali Sonkusare[4], Aishwarya Ghuguskar[5], Karuna Gedam[6], Pranjali Pongde[7], Sweta Fating[8]

[1,2,3,4,5,6,7,8] *KDK College Of Engineering Umrer*

Abstract- **Disability of visual text reading has a huge impact on the quality of life for visually impaired people. Day to day activities of visually impaired people may become challenging and complicated. Reading texts on product labels, book covers, images having complex background using conventional method have many problems. We propose a camera-based assistive text reading framework to help visually impaired persons read text labels and product packaging from hand-held objects in their daily lives. To isolate the object from cluttered backgrounds or other surrounding objects in the camera view, we first propose an efficient and effective motion based method to define a region of interest (ROI) in the video by asking the user to shake the object. This method extracts moving object region by a mixture-of-Gaussians-based background subtraction method. In the extracted ROI, text localization and recognition are conducted to acquire text information.**

I. INTRODUCTION

The national census of India has estimated 21.9 million disabled people in country. Out of which 48.5% people are visually impaired. Many day to day activities of visually impaired people may become challenging and complicated. One of the major problem they face is understanding visual text information which is not in Braille form. Their disabilities have made them less access to computers, Internets, documents etc. Printed texts are found everywhere in the form of reports, classroom handouts, instructions on medicine bottles, bank statements, product labels etc.

This framework helps visually impaired person to extract texts from hand held objects. The object is captured with the help of camera. It is given to computer where image is processed in Matlab. Fast and robust image processing is done to improve the quality of image. Classification of text and non text regions are done with the help of classifier. Extracted data is given to microcontroller which is interfaced with the speaker. Finally visually impaired person will get the output in audio form. Reading is essential in today's society. Printed texts are found everywhere in the form of reports, bank receipts, classroom handouts, instructions on medical bottles, product labels etc. It is very challenging for visually impaired people to read the printed texts because of their disabilities.

There exists some few technologies which help visually impaired people to read printed text and product labels but there exists some limitations of that. Visually impaired person can access the printed data using Braille scripts and audio tapes. But these media have limited assistance. Mobile barcode reader can read the product information from the extensive database but it is very difficult for visually impaired person to correctly find the position of barcode. Some reading-assistive systems such as pen scanners might be employed in these and similar situations. K-Reader, mobile runs on a cell phone and it allows the user to read mails, receipts, documents etc. But here the document to be read must be flat, placed on clear, dark surface and contains texts. It accurately reads black texts on white background. But it cannot handle colored backgrounds or colored texts. Many systems cannot handle the texts written on cylindrical objects or non horizontal The goal of this system is to read texts on hand held objects in images with complex background. It will help us to read horizontal and non horizontal text string.

LITERATURE REVIEW

A. Text Detection from Natural Scene Images
Every year, the number of visually impaired persons is increasing due to eye diseases diabetes, traffic

accidents and other causes. There are about 200,000 persons with acquired blindness in Japan. Therefore computer applications that provide support to the visually impaired persons have become an important theme. We have already developed a pen-based character input system for blind persons using a PDA. On this system, people with acquired blindness remember the shape and writing order of Japanese characters and they can use this system as a notepad and as an E-mail terminal anytime, anywhere. This application essentially works as communication tool. However, such a device does not solve all of the problems encountered by a blind person willing to go outside unaccompanied.When a visually impaired person is walking around, it is important to get text information which is present in the scene. For example, a 'stop' sign at a crossing without acoustic signal has an important meaning. In general, way finding into a man-made environment is helped considerably by the ability to read signs.

B. Ote-Ocr Based Text Recognition and Extraction from Video Frames
The goal of this paper is to provide a new method to detect and recognize the text from the video frames. The task performed is divided into three step approach that combines the text detection and text recognition from the video frame. The video frame creation involves in dividing the video into an individual frames. The individual frame is grabbed and passed to the rest two phases. The text detection is a two-step approach, which involves text localization phase and the text verification phase. The text recognition involves in text verification phase and the optical character recognition phase. The final outcome of this paper is the detection of the text from the video frames in a word file. Experimental results demonstrating the proposed approach was also included, which shows the accuracy level of Optical character recognition (OCR) in terms of text extraction.

C. Text Detection and Recognition in Natural Scene Images
Text detection and recognition in natural scene images plays an important role in content analysis of images. In this paper, based on the characteristics of scene text, we propose a robust text detection and recognition method using Maximally Stable Extremal Regions (MSER) and Support Vector Machine (SVM). Different from the end to end text recognition, we split the recognition problem into detection and recognition procedure. Firstly, in the detection stage, in order to extract potential text as much as possible, we use MS ER and color clustering to extract connected component. Then, for the obtained candidate connected component, we use visual saliency and some prior information to filter non-text regions. Finally, we can obtain word image by text line generation. In the recognition stage, we use vertical projection to segment word images, then recognize character in SVM based framework. The experiment results evaluated on standard dataset show that with a small amount of prior information and simple segment strategy, the proposed method has a better performance compared to conventional text detection and recognition method.

HARDWARE REQUIREMENTS

A. CP2102 Serial to USB Converter Board

Fig 1: CP2102 Serial to USB Converter Board

It is the small size USB serial module [19]. It is designed as per convenience of user. User can interface devices or modules which supports the UART (Universal Asynchronous Receiver and Transmitter) with the laptops or PCs through USB or serial port. This module comprises of standard USB type-B connector in order to reduce the size of the module. It provides platform for adding a serial or USB interface to product designs which are already existing. MAX232 IC behaves as transition stage between RS232 (Recommended Standard 232) and TTL (Transistor-Transistor Logic) voltage levels. Receivers (Rx) pin of MAX232 convert from RS-232 to TTL voltage levels and driver pin (Tx) converts from TTL logic to RS-232 voltage levels. Similarly, CP2102 is transition stage in between USB and TTL voltage levels which allows data to be read or write through USB port.

B.   Text to Speech module

Fig 2: TTS Module

In this paper has introduced an innovative, efficient and real-time cost beneficial technique that enables user to hear the contents of text images instead of reading through them. It combines the concept of Optical Character Recognition (OCR) and Text to Speech Synthesizer (TTS). 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 conversion is a method 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

C.   USB Camera

Fig 3: Logitech HD C207 Web Camera

A webcam is a video camera that feeds or streams its image in real time to or through a computer to a computer network. When "captured" by the computer, the video stream may be saved, viewed or sent on to other networks travelling through systems such as the internet, and e-mailed as an attachment. When sent to a remote location, the video stream may be saved, viewed or on sent there. Unlike an IP camera which connects using Ethernet or Wi-Fi, a webcam is generally connected by a USB cable, or similar cable, or built into computer hardware, such as laptops

D.   Earphone

Fig 4: Earphone (3.5mm phone jack)

Headphones are a pair of small loudspeaker drivers worn on or around the head over a user's ears. They are electro acoustic transducers, which convert an electrical signal to a corresponding sound. Headphones let a single user listen to an audio source privately, in contrast to a loudspeaker, which emits sound into the open air for anyone nearby to hear. Headphones are also known as ear speakers, earphones or, colloquially, cans. Circum aural ('around the ear') and supra-aural ('over the ear') headphones use a band over the top of the head to hold the speakers in place. Headphones connect to a signal source such as an audio amplifier, radio, CD player, portable media player, mobile phone, video game console, or electronic musical instrument, either directly using a cord, or using wireless technology such as Bluetooth, DECT or FM radio. The first headphones were developed in the late 19th century for use by telephone operators, to keep their hands free. Initially the audio quality was mediocre and a step forward was the invention of high fidelity headphones.

SOFTWARE

A.   MATLAB 8.5 (R2015a)

The MATLAB (Matrix Laboratory) is the high-level language for technical as well as numeric computation, visualization, programming and developing applications. It provides different computation methods for analyzing data, creating models and developing various algorithms. It is used for accessing the data from files, database and external device. We can also read the file from popular format such as Microsoft Excel. Using MATLAB, we can acquire data from the hardware devices using serial port or sound card, and measure data directly into MATLAB for visualization and analysis and can also communicate with signal analyzers.

1)   Computer Vision System Toolbox:

Computer Vision System Toolbox provides algorithms, apps, and functions to design and simulate computer vision and video processing system. We can perform different functions like object detection and tracking, feature detection, extraction, and matching, video processing and motion estimation. For 3-D computer vision, the system toolbox supports stereo vision, camera calibration, 3-D point cloud processing and 3-D reconstruction. With the help of machine learning based frameworks, we can train object recognition, object detection, and image retrieval systems. Algorithms are available as System objects, MATLAB functions, and Simulink blocks

2)   OS Video Generic Interface:

From version R2014a, each adaptor is separately available through the Support Package Installer. To use toolbox with adaptor, all support packages contain the necessary MATLAB files. It also contain third-party files like camera set-up utilities and drivers.For using the cameras with the toolbox, we need to install the support package for the adaptor that the camera uses. If we use multiple adaptors, then we need to install the support package for each one we use. For example, if we have a Webcam on our Windows system and a Matrox camera, then for webcam, we should install the OS Generic Video Interface support package for the win video adaptor and the Matrox Hardware support package for the matrox adaptor.
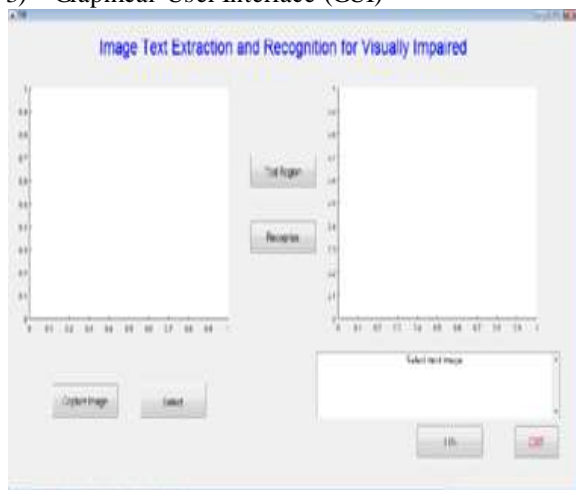
3)   Graphical User Interface (GUI)

Fig 5: GUI

GUI is the main part of software application for interaction between human and computer. User can interact with our system with the help of graphical user interface. Once the user get GUI, he can control it with the help of some keys on the keyboard. Here, 'Capture Image' button can be control by key 'F'. 'Text Region' button can be controlled by key 'G'. 'Recognize' button can be controlled by key 'H'. 'TTS' button can be controlled by key 'J'. We can also make GUI more effective by giving voice instructions on different buttons. GUI of this project is as shown above.

PROPOSED METHODOLOGY

A.   System Architecture

This framework mainly consists of three components. They are

1.   Image Capture
2.   Data Processing and
3.   Audio Output

1.   Image Capture:-The image capture component captures image with the help of web camera attached to the system. Here, Image is captured with the help of web cam (Logitech C207 HD). The image from the webcam is taken in RGB format.

2.   Data Processing:-The data processing component is used for processing captured image for extracting text regions. Fast and robust image processing techniques are used for improving image quality and to detect text region in an image. Text localization algorithm is used to obtain text regions in image. Text recognition is performed to transform text information into understandable codes. We can use laptop or desktop computer as a processing device.

3.   Audio Output:-The audio output component helps to inform the visually impaired user of recognized text codes in the form of audio or speech. Recognized texts are passed serially to text to speech conversion module through serial convertor. By using text to speech module, text information is converted into speech or audio. Visually impaired user can hear the audio output with help of headphones or speaker attached to the system.
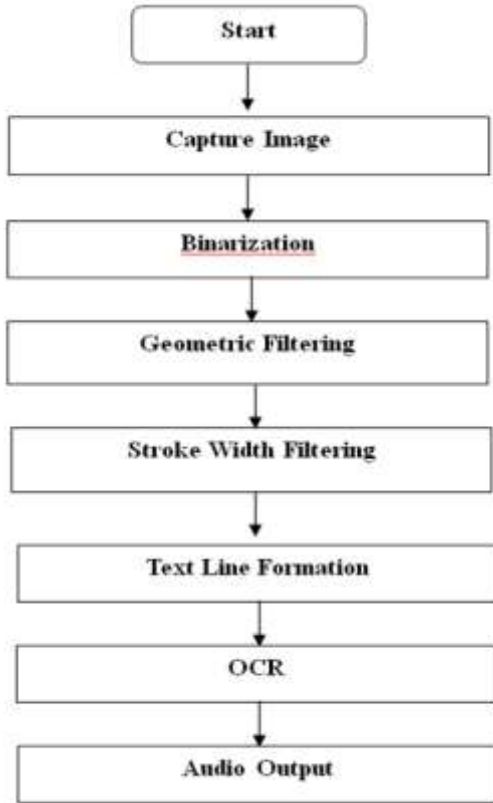
B.   Flowchart

Fig 6: Flowchart of Proposed Method

RESULT

To evaluate the effectiveness or efficiency of our proposed system, we apply this algorithm on various captured images. We have collected images of book covers, product labels, natural scene images, documents etc. Logitech C207 HD web camera with autofocus is connected to laptop with the help of USB connection. Figure 5.1 shows snapshot of experimental system.

Fig.5: Snapshot of Experimental Result

CONCLUSION

We have proposed a system to read text on book covers, documents, product labels, scene images etc. Non-text regions from cluttered or complex background are removed by applying geometric rules and stroke width filtering. Text recognition is done with OCR i.e. optical character recognition. User will get the output in speech or voice with the help of text to speech converter module. Performance of proposed method is degraded when deal with distorted or blur and low resolution or low contrast images.

APPLICATIONS

1) Wearable computing:
Wearable computing is one of the promising application of Text recognition from camera captured images. A camera attached to a pair of glasses embedded with OCR technology can capture infinite range of information. Visually impaired and blind people can get the benefit from such advanced technologies. Recognizing text in natural scenes images which is embedded with text to speech technology, can make systems to read street signs, book covers, bank receipts, name plates ,labels on office doors, medicine bottles etc. For visual disabled person, such devices provide great promises and boosts the confidence level to live their life independently.

2) Portable travelling aids:
Camera captured texts are automatically decoded with the help of cameras which are having translation capabilities (It is used for foreign language). Such types of technology will help people traveling abroad for understanding signs, destinations of buses and trains, names of streets etc.

3) Content based video indexing and retrieval:
Text detection in video has received growing attention because it provides an additional way of indexing the video. Words are having well defined and clear meaning. Therefore text information which can be extracted from video gives clear information about its content.

4) Robotic navigation:

Robots are helpful in automatic navigation especially in unknown environment for getting destination when maps are not available. When the system is integrated with OCR, road sign boards, offices, nameplates, buildings can be identified with the help of texts written on it. Text recognition finds important applications in robot navigation such as goal driven navigation, path planning etc.

ADVANTAGES

1. Visually impaired people can lead an independent life. They don't need to seek help of other person for their work.
2. To some extent we can help visually impaired people to cope up of equalize with the normal people.
3. The graphical user interface is created in a way such that the visually impaired people can interact with the application easily.
4. As we are not using desktop application, we can use this application in home or outside home also.
5. It can be used by the normal people who cannot read the language but understands it.
6. It also helps to learn the Pronunciation of the word in the written languages with the help of text to speech conversion.

FUTURE SCOPE

We have described a system to read printed text on various images for assisting visually impaired people. This method can effectively extract the text regions from images with complex background. Our future work will extend our algorithm to handle curved or non-horizontal text string. In future we will enhance this work with higher accuracy and large number of input samples. Also we will try to increase portability of system.

REFERENCES

[1] Shashank Shetty and Arun S Devadiya "Ote-Ocr Based Text Recognition and Extraction from Video Frames" 2014 IEEE 8th Proceedings International Conference on Intelligent Systems and Control (ISCO)

[2] Marcin Pazio, Ryszard Kowalik "Text Detection System for The Blind" 15th European Signal Processing Conference (EUSIPCO 2007), Poznan, Poland, September 3-7, 2007, copyright by EURASIP