A Prototype of Assistive Text Label Reading System for Blind – An Overview

Ms. P. H. Shah[1] Mr. G.P.Jain[2]
[1]M.E. Student [2]Assistant Professor
[1]Department of Electronics Engineering [2]Department of Electronics and Tele-Communication Engineering
[1,2]WIT, Solapur, India

Abstract— Self-Dependency for disabled persons is of prime importance. This paper presents a cost effective prototype system to help blind persons to shop independently. A camera based prototype system is proposed in which label of the product which a blind user wishes to buy will be extracted from the image and will be pronounced as voice. To isolate the object of interest from the captured image sequence or video, the user is asked to shake the product. Background subtraction will be used to detect the object of interest that is nothing but Region of Interest (ROI). Text extraction will be performed on ROI to obtain the label of product. The prototype system will be implemented on Raspberry Pi board.

Key words: Portable, Label Reading, Text Detection, Text Extraction, Raspberry Pi

I. INTRODUCTION

Person with disabilities are important part of society. Due to their disability they are unable to put social, economic impact on society. The National Census of India has estimated around 21.9 Million disabled people in the country. Out of which more than 15 million people in India are blind [1]. Generally, a significant number of individuals who are legally blind find large print or audio texts helpful, while only 8-10% use Braille as a reading medium. On a global scale, there are 37 million blind people; every second, a person becomes blind, every minute, a child becomes blind. This is considered to be the highest among all other disabilities. Blind people are an integral part of the society. However, their disabilities have made them to have less access to computers, Internet, and high quality educational software than the people with clear vision. Consequently, they have not been able to improve on their own knowledge, and have significant influence and impact on the economic, commercial, and educational ventures in the society. One way to narrow this widening gap and see a reversal of this trend is to develop a system, within their economic reach, and which will empower them to communicate freely and widely using the Internet or any other information infrastructure. Text information acts as an important scene indicator such as on medicine bottles, text signage in street, newspapers, instructions on various machines, bank notes, train tickets, product labels etc. Therefore, robust automatic text localizing and recognition is essential for automatic text reading in various systems to help visually impaired persons.

Many systems such as K Reader Mobile, Portable Scanner [2], [3] have been invented and implemented to assist disabled people. These systems integrate text recognitions along with text to speech conversion software. But those are not handy, are cumbersome and cannot be implemented in each and every situation. Even though many systems have proposed earlier, the problem of reading labels from product packages with different complex backgrounds and different fonts, styling and text color has not been addressed yet.

The remainder of the paper is organized as follows- Section II briefly overviews the text extraction and text recognition methods. Section III describes the proposed system in brief. Section IV concludes the paper.

II. TEXT EXTRACTION AND TEXT DETECTION:

Text Information Extraction is the first and important function of any assistive reading system for blind and this process determines the intelligibility of the output speech. Many algorithms were presented to localize text regions in scene images. These methods can be classified into four categories.

The first category is based on frequencies. (1) In this kind of approach, the text is extracted from the background in the frequency (e.g. wavelet) domain. This is time consuming, and the frequency representation may not be better than the spatial representation. Bong-Kee Sin, Seon Kyu Kim and Beom Joon Cho used frequency features in both time and frequency domain to locate text strings in natural scene image [5].The second category is based on textures(2) and assumes that texts have specific texture patterns. It is more time-consuming and can fail when the background is cluttered with text. Kwang In Kim, Keechul Jung, and Jin Hyung Kim proposed a novel texture-based method for detecting texts in images. [6]

The third category is based on edges (3) which assume high contrast differences between the text and background. It is fast and can have high recall rate. However, it often produces many false alarms since the background may also have strong edges similar to the text. The fourth category uses connected component analysis (CCA), in which pixels with similar colors are grouped into connected components, and then into text regions. CCA is fast but, it fails when the texts are not homogeneous and text parts are not dominant in the image.

Recently, there is a lot of interest in using pattern classification techniques (such as AdaBoost, support vector machines) for text localization. A support vector machine (SVM) is used to analyze the textural properties of texts. Xilin Chen, Jie Yang, Jing Zhang, and Alex Waibel presented an approach for automatic detection and recognition of signs from natural scenes, and its application to a sign translation task. The team has applied the approach in developing a Chinese sign translation system, which can automatically detect and recognize Chinese signs and translate the recognized text into English. [7] Chucai Yi and YingLi Tian proposed a new method- The adjacent character grouping method calculates the sibling groups of

each character candidate as string segments and then merges the intersecting sibling groups into text string [8].

With the help of elaborately designed features that incorporate various properties of the text (such as geometry, color, texture and frequency), these techniques are often successful indiscriminating text from its background.

In the proposed system, initially the product, from which label has to be identified, has to be detected from the image captured by the camera, since the image captured may contain various products e.g. Image taken at a Grocery shop. Examples of the products are shown in figure 1.

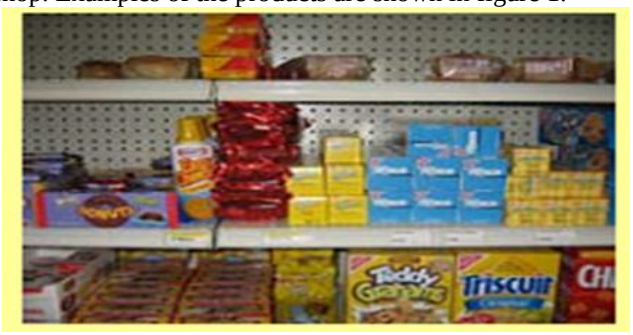

Fig. 1: Proposed System

The hand held product will be moved by the user and with help of motion based background subtraction method, the object of interest can be found. The part of the image, containing the product can be thus now isolated and is the region of interest. (ROI)

The process of text recognition from the image consists of steps such as capturing image, detecting text and non-text regions, grouping the text characters together and recognizing the text with the help of off-the-shelf optical character recognition software (OCR). Many algorithms have been proposed to automatically locate the text strings in scenes. These algorithms can be classified as Rule based algorithms and Learning based algorithms. Rule based algorithms apply pixel level image to text information from predefined text layouts such as character size, aspect ratio, edge density, character structure, color uniformity of text string, etc. Learning based algorithms on the other hand model text structure and extract representative text features to build text classifiers.

III. PROPOSED FRAMEWORK

A. Block Diagram:

The process flow can be depicted in the figure no.2. Logitech™ webcam software supports motion detection which can be utilized for detecting object of interest. The process of detection of Region of Interest (ROI) from the image sequence can be done using moved object method The ROI region is confined within the rectangular area and it contains the text which has to be detected.

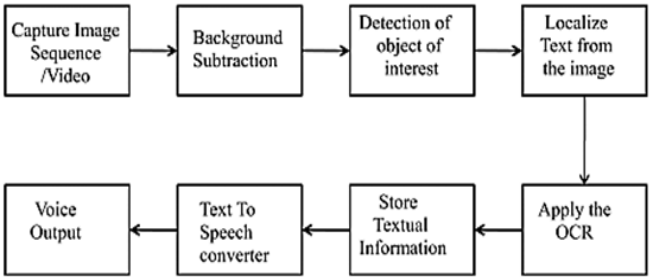

Fig. 2: Block Diagram

The live video is captured by using web cam and it can be done using OPENCV libraries. The image format from the webcam is in RGB24 format. The frames from the video is segregated and undergone to the preprocessing. Background subtraction (BGS) can be performed by many methods such as mixture of K –Gaussian; multiple Gaussian mixtures based BGS etc. It can be done by method of mean and variance by comparing each new frame with previous one and comparing number of pixels which are changed.

To extract text information from video,

$$S = \frac{1}{|V|}\sum i\, R(Vi, B)$$

Where Vi denotes the i^{th} frame in the captured sequence, |V | denotes the number of frames, B denotes the estimated background from motion- based object detection, an R represents the calculated foreground object at each frame and S denotes user's object of interest.

Once confined ROI is obtained rule based or learning based text localization algorithms can be used. Region based method focuses on high contrast between text and background. Canny edge detector thus can be used for obtaining contour of text strings within the cropped image. In learning based classifiers, text classifies are built up Haar based classifier can be used to train classifier.

B. Software Part:

OpenCV (open source computer vision library) will be used for image processing. Processing will include video capturing, segmenting the video into frames, background subtraction and detection of object of interest from the captured image, localizing the text within the image automatically, and recognizing the text using optical character recognition (OCR). Many different packages can be installed on Raspberry Pi for portable implementation of the proposed framework. Tessaract library can be used for text extraction, although Tessaract is designed for and preforms best with simple backgrounds, standard fonts and well organized characters, rather than commercial product boxes with multiple decorative patterns and fonts. Text codes output from OCR can be converted to speech using text to speech (TTS) software e.g. e-speak engine, Flite engine etc. Pico2wave library is installed which converts text file into '.wav' file format.

C. Hardware Part:

The Proposed portable system can be implemented on the 'On board Computer' Raspberry Pi board. Raspberry Pi is single board computer with on board 512 MB RAM and its System on Chip is a Broadcom BCM2835 .This contains an ARM1176JZFS (ARM11 using an ARMv6-architecture core) with floating point, running at 700 MHz, and a Video core 4 GPU. Out of many available single board computers, cost (just some 3000 bugs), portability and size (credit card size!) are attractive features of Raspberry Pi. For image capturing purpose, Logitech™ USB camera is used, although either any USB camera or Camera module of Pi can be used.

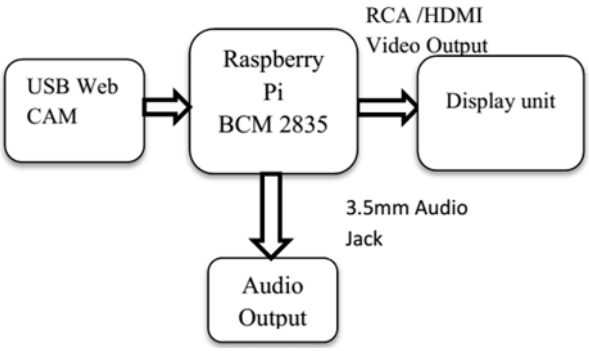

Fig. 3: Depicts the Block Diagram of the Proposed Framework.

Many operating systems can be installed on SD card in order to use raspberry pi as standalone portable device. Examples of operating system are RASPBMC, Raspbian, RISC OS, Pidora, OPENELEC etc. Raspbian is installed on SD card. Raspbian is a free operating system based on Debian optimized for the Raspberry Pi hardware. It comes with over 35,000 packages; pre-compiled software bundled in a nice format for easy installation on Raspberry Pi.

IV. Conclusion

The proposed system aims at developing a general framework to assist blind person to identify the label of product package from hand held product .The proposed system will help to improve the ability of people who are blind or have significant visual impairments to independently access, understand, and explore unfamiliar products.

V. Acknowledgment

I would like to seize the opportunity to express my sincere gratitude towards Head of Department Prof. Mr.S.R.Gengaje, project guide Prof. Mr. G P. Jain and other electronics department faculties for their invaluable co-operation and guidance.

References

[1] National Census of India- http://censusindia.gov.in/Census_And_You/disabled_population.aspx
[2] ScanTalker, Bar code scanning application to help Blind Identify over one million products. (2006)
[3] KNFB Reader -https://nfb.org/handheld-reader
[4] Chucai Yi, Student Member, IEEE, YingliTian, Senior Member, IEEE, and Aries Arditi "Portable Camera-Based Assistive Text and Product Label Reading from Hand-Held Objects for Blind Persons", IEEE/ASME Trans. on MECHATRONICS, vol. 19, No. 3, June 2014
[5] Bong-Kee Sin, Seon Kyu Kim, Beom Joon Cho "Locating characters in scene images using frequency features." IEEE 2002.
[6] K. Kim, K. Jung, and J. Kim, "Texture-based approach for text detection in images using support vector machines and continuously adaptive mean shift algorithm," IEEE Trans. Pattern Anal. Mach. Intell., vol. 25, no. 12,pp. 1631–1639, Dec. 2003.
[7] X. Chen, J. Yang, J. Zhang, and A. Waibel, "Automatic detection and recognition of signs from natural scenes," IEEE Trans. Image Process., vol. 13, no. 1, pp. 87–99, Jan. 2004.
[8] C. Yi and Y. Tian, "Text string detection from natural scenes by structure based partition and grouping," IEEE Trans. Image Process., vol. 20, no. 9, pp. 2594–2605, Sep. 2011
[9] www.raspberrypi.org
[10] opencv.org

