ISSN 2348-1196 (print)
International Journal of Computer Science and Information Technology Research ISSN 2348-120X (online)
Vol. 3, Issue 2, pp: (897-904), Month: April - June 2015, Available at: www.researchpublish.com

Product Label Reading System For Visually Challenged People

[1]NAGARATHNA, [2]SOWJANYA V M

[1]Asst Professor, Dept of CS&E, PESCE Mandya, [2]M.tech Final Year Student, PESCE Mandya

Abstract: **This paper presents a camera-based label reader to help blind persons to read names of labels on the products. Camera acts as main vision in capturing the video of the product then video captured is split into frames. From sequence of frames one frame is taken and processed internally to separate label from frame using text detection algorithm. The received label image is converted to text by using OCR (Optical Character Recognizer). Once the identified label name is converted to text and converted text is output to blind users in speech.**

Keywords: **Assistive devices, blindness, distribution of edge pixels, hand-held objects, optical character recognition (OCR), stroke orientation, text reading.**

1. INTRODUCTION

OF the 314 million visually impaired people worldwide, 45 million are blind. Even in a developed country like the U.S., the 2008 National Health Interview Survey reported that an estimated 25.2 million adult Americans (over 8%) are blind or visually impaired. This number is increasing rapidly as the baby boomer generation ages. The number of blind persons in India in 2010 was estimated to be 24.7 million**.** India shoulders the largest burden of global blindness, about 3.5 million across the country with 30000 new cases being added each year.Recent developments in computer vision, digital cameras, and portable computers make it feasible to assist these individuals by developing camera-based products that combine computer vision technology with other existing commercial products such as optical character recognition (OCR) systems.The development of an effective visual information system will significantly improve the degree to which the visually impaired can interact with their environment.

Reading is obviously essential in today's society. Printed text is everywhere in the form of reports, receipts, bank statements, restaurant menus, classroom handouts, product packages**,** instructions on medicine bottles, etc. and while optical aids, video magnifiers, and screen readers can help blind users and those with low vision to access documents, there are few devices that can provide good access to common hand-held objects such as product packages, and objects printed with text such as prescription medication bottles.

Today, there are already a few systems that have some promise for portable use, but they cannot handle product labeling. For example, portable bar code readers designed to help blind people identify different products in an extensive product database can enable users who are blind to access information about these products through speech and braille. But a big limitation is that it is very hard for blind users to find the position of the bar code and to correctly point the bar code reader at the bar code. Some reading-assistive systems such as pen scanners might be employed in these and similar situations. Such systems integrate OCR software to offer the function of scanning and recognition of text and some have integrated voice output. However, these systems are generally designed for and perform best with document images with simple backgrounds, standard fonts, a small range of font sizes, and well-organized characters rather than commercial product boxes with multiple decorative patterns. Most state-of-the-art OCR software cannot directly handle scene images with complex backgrounds. K Reader Mobile runs on a cell phone and allows the user to read mail, receipts, fliers, and many other documents. However, the document to be read must be nearly flat, placed on a clear, dark surface (i.e., a no cluttered background), and contain mostly text. Furthermore, K Reader Mobile accurately reads black print on a white background, but has problems recognizing colored text or text on a colored background. It cannot read text with complex backgrounds, text printed on cylinders with warped or incomplete images (such as soup cans or medicine bottles). Furthermore, these systems require a blind user to manually localize areas of interest and text regions on the objects in most cases.

The physical object that displays information will be helpful for the blind. Although a number of reading assistants have been designed specifically for the visually impaired, to our knowledge, no existing reading assistant can read text from the kinds of challenging patterns and backgrounds found on many everyday commercial products. The main aim is to design a system for blind persons to recognize the hand held objects or products.

2. LITERATURE SURVEY

In [1] a vision based assistive system for label detection with voice output is discussed. A camera based assistive text reading framework help blind persons read text labels and product packaging from hand-held object in their daily resides. To isolate the object from cluttered backgrounds or other surroundings objects in the camera view, 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. In the extracted ROI, text localization and recognition are conducted to acquire text information. To automatically localize the text regions from the object ROI, a novel text localization algorithm by learning gradient features of stroke orientations and distributions of edge pixels in an Adaboost model is applied. Text characters in the localized text regions are then binarized and recognized by off-the shelf optical character recognition software. The recognized text codes are output to blind users in speech.

[2] Discusses an intelligent system. Physically invisible people experience difficulty and inconvenience using computers through a keyboard and mouse. The purpose of this system is to provide a way the blind people population can easily control many functions of a computer via speech. When blind people speak, the audio voice input is sent to the speech Browser .solenoid plated are very useful to convert this web search into braille. Many applications running on this purpose but not all the applications able to fulfil over it and this system has better aspects in future for normal people as well as blind people. This application is firstly embedded on raspberry pi and Qt creator is the software which is being useful to interface this GUI with the hardware connected to Pi.

[3] Presents Darshan a Navigation System for blind people to navigate safely and quickly, in the system obstacle detection and recognition is done through ultrasonic sensors and USB camera. The proposed system detects the obstacles up to 300 cm via ultrasonic sensors and sends feedback in the form of beep sound via earphone to inform the person about the obstacle. USB webcam is connected with Raspberry Pi Embedded board which captures the image of the obstacle, which is used for finding the properties of the obstacle (Human Being). Human presence is identified with the help of human face detection algorithm written in Open CV. The constraints coming while running the algorithm on Embedded System are limited memory and processing time and speed to achieve the real time image processing requirements. The algorithm is implemented in Open CV, which runs on Debian based Linux environment.

3. PROPOSED SYSTEM

This project is to design and develop a system to find products or objects with voice announcements. The proposed system mainly consists of three functional components:

- Video capturing
- Text detection and
- Audio output

The video capturing component refers to camera and it is used to capture video of the hand held object or product. The text detection component refers to raspberry pi board and it is used for deploying text detection and recognition algorithm. The third component is audio output it refers to ear phone used to give the voice output of the recognized text. The figure 1 show hardware set up of the proposed system.

ISSN 2348-1196 (print)
International Journal of Computer Science and Information Technology Research ISSN 2348-120X (online)
Vol. 3, Issue 2, pp: (897-904), Month: April - June 2015, Available at: www.researchpublish.com

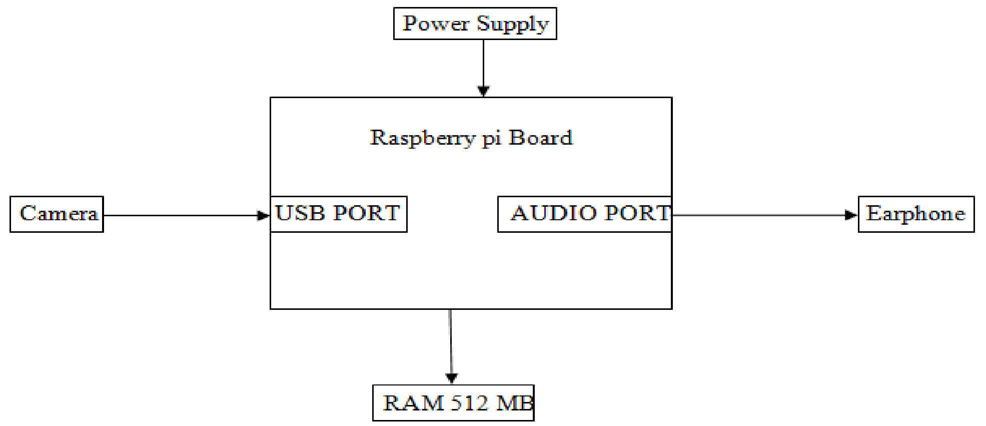


Fig 1 Hardware set of the proposed system

3.1 VIDEO CAPTURING:

The video capturing component refers to camera. The hand held product or objects video is captured using camera. To identify the product label in the capture video it is passed on to text detection component. The text detection component receives the hand held object video as the input. The figure 2 shows the detailed flow of the label reading system.

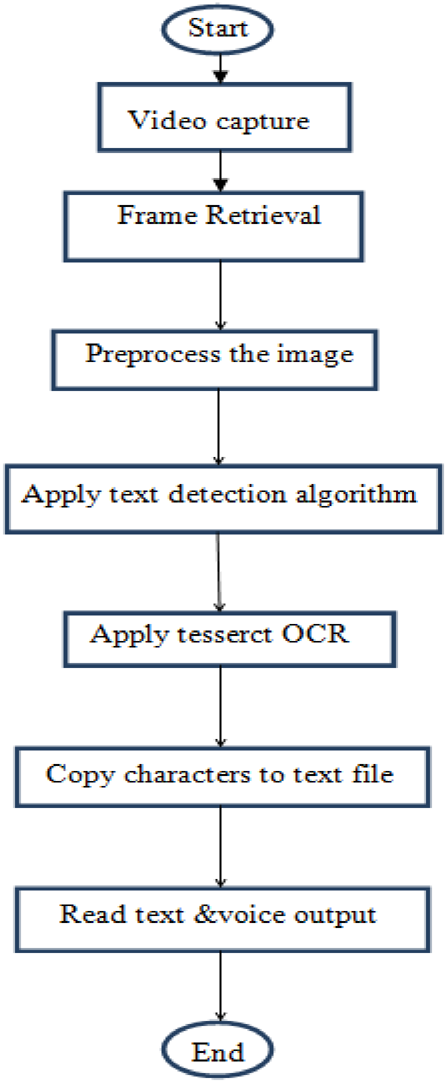


Fig 2.The detailed flow of the label reading system

3.2 TEXT DETECTION:

The raspberry pi board is used for deploying text detection and recognition algorithm. The Raspberry Pi Model B+ is a credit-card sized computer board that's up and running when a keyboard, mouse, display, PSU and MicroSD card with installed OS are added.The secret sauce that makes this computer so small and powerful is the Broadcom BCM2835, a System-on-Chip that contains an ARM1176JZFS with floating point, running at 700MHz, and a Videocore 4 GPU. The GPU provides Open GL ES 2.0, hardware-accelerated OpenVG, and 1080p30 H.264 high-profile decodes and is capable of 1Gpixel/s, 1.5Gtexel/s or 24 GFLOPs of general purpose compute. The Model B+'s FOUR built-in USB ports provide enough connectivity for a mouse, keyboard, or anything else that the RPi needs. Powering the Raspberry Pi is easy just plug any USB power supply into the micro-USB port. There's no power button so the Pi will begin to boot as soon as power is applied, to turn it off simply remove power.

To identify the text from the video, first the video captured is segregated into frames. The frame will be in the JPEG format. Text detection algorithm takes frames and finds the text in the frame.

Steps involved in detecting text in video frame:

Preprocessing the frame:

- Obtain gray scale of image.
- Perform canny edge detection on gray scale image.

Apply SWT algorithm on the canny image. It has 3 major steps:

1. The stroke width transforms.
2. Grouping the pixels into letter candidates based on their stroke width.
3. Grouping letter candidates into regions of text

- Binarize (thresholding) resulting image.

3.2.1 Gray scaling:

A grayscale image is an image in which the value of each pixel is a single sample, that is, it carries only intensity information. Grayscale is a range of shades of gray without apparent color. The darkest possible shade is black, which is the total absence of transmitted or reflected light. The lightest possible shade is white, the total transmission or reflection of light at all visible wavelengths. Intermediate shades of gray are represented by equal brightness levels of the three primary colors (red, green and blue) for transmitted light, or equal amounts of the three primary pigments (cyan, magenta and yellow) for reflected light.

3.2.2 Canny edge detector:

The Canny edge detector is an edge detection operator that uses a multistage algorithm to detect a wide range of edges in images The main steps involved in converting a gray image to edge image are:

- Filter out any noise. The Gaussian filter is used for this purpose.
- Find the intensity gradient of the image
- Non-maximum suppression is applied. This removes pixels that are not considered to be part of an edge. Hence, only thin lines (candidate edges) will remain.
- Hysteresis: Canny does use two thresholds (upper and lower). If a pixel gradient is higher than the upper threshold, the pixel is accepted as an edge, if a pixel gradient value is below the lower threshold, then it is rejected and if the pixel gradient is between the two thresholds, then it will be accepted only if it is connected to a pixel that is above the upper threshold.

3.2.3 Stroke Width Algorithm:

A stroke in the image is a continuous band of a nearly constant width. An example of a stroke is shown in figure 3(a). The Stroke Width Transform (SWT) is a local operator which calculates for each pixel the width of the most likely stroke containing the pixel.

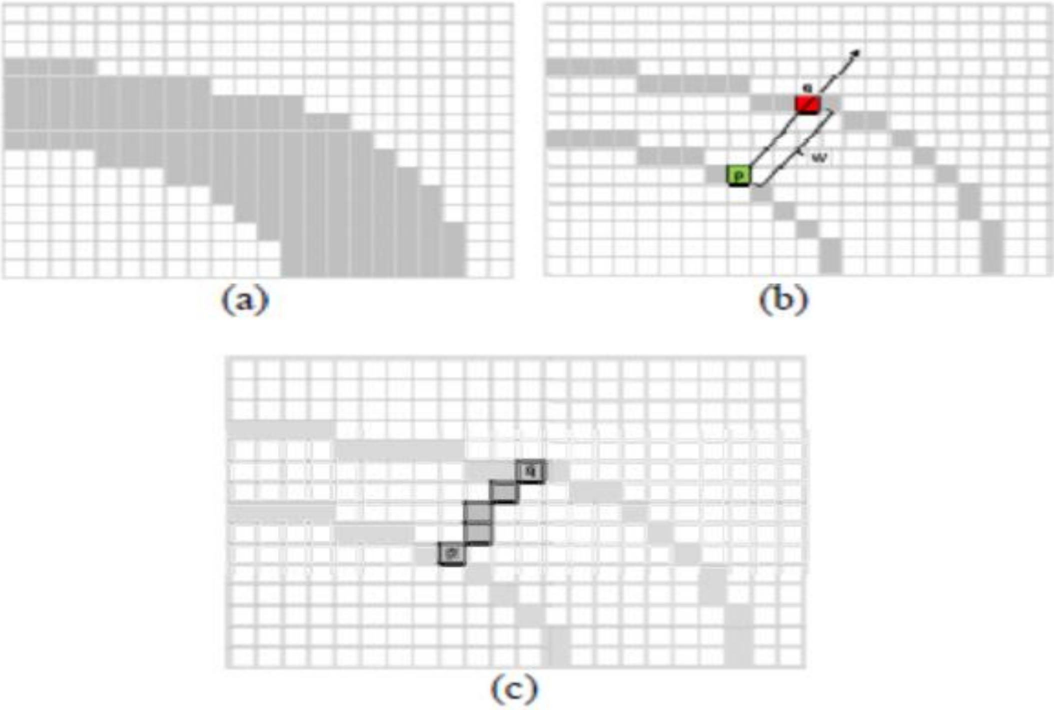


Fig 3 a) A typical stroke b) p is pixel on the boundary of the stroke. Searching in the direction of the gradient at p, leads to finding q, the corresponding pixel on the other side of the stroke c) Each pixel on the ray is assigned by the minimum of its current value and the found width of the stroke [6].

All the edges considered as possible stroke boundaries, and should find the width of such stroke. If p is an edge pixel, the direction of the gradient is roughly perpendicular to the orientation of the stroke boundary. Therefore, the next step is to calculate the gradient direction gp of the edge pixels, and follow the ray r=p+n*gp (n>0) until another edge pixel q is found. If the gradient direction gq at q is roughly opposite to gp, then each pixel in the ray is assigned the distance between p and q as their stroke width, unless it already has a lower value. If, however, an edge pixel q is not found, or gq is not opposite to gp, the ray is discarded. In order to accommodate both bright text on a dark background and dark text on a bright background, the algorithm should be applied twice: once with the ray direction gp and once with –gp. After the first pass described above, pixels in complex locations might not hold the true stroke width value as shown in figure 3(b). For that reason, it has to be passed along each non-discarded ray, where each pixel in the ray will receive the minimal value between its current value, and the median value along that ray.

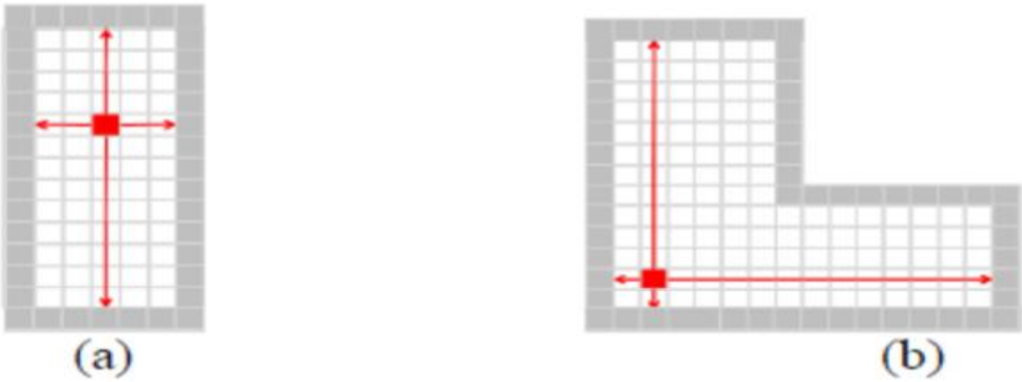


Fig 4 filling the pixels with SWT value a) An example red pixel is filled with minimum between the lengths of vertical and horizontal rays passing through it. Proper stroke width value is stored b) An example red pixel stores the minimum between the two rays lengths; this is not the true stroke width; this shows the necessity of second pass [6]

3.2.4 Finding letter candidates:

Obtain the map of the most likely stroke-widths for each pixel in the original image. The next step is to group these pixels into letter candidate. This will be done by first grouping pixels with similar stroke width, and then applying several rules to distinguish the letter candidates. The grouping of the image will be done by using a Connected Component algorithm. In order to allow smoothly varying stroke widths in a letter, two pixels will be grouped together if their SWT ratio is less than 3.0. Detect the connected components which can pass as letter candidates, by applying a set of fairly flexibly rules. These rules are as follows:

-The variance of the stroke-width within a component must not be too big. This helps with rejecting foliage in natural images, which are commonly mistaken for text.

-The aspect ratio of a component must be within a small range of values, in order to reject long and narrow components.

-The ratio between the diameter of the component and its median stroke width to be less than a learned threshold. This also helps reject long and narrow components.

-Components whose size is too large or too small will also be ignored. This is done by limiting the length, width, and pixel count of the component.

3.2.5 Grouping letter candidates into text regions:

Text on a line is expected to have similarities, including similar stroke width, letter width height and spaces between the letters and words based this letters will be grouped into text regions. The block diagram of the text detection algorithm is shown in fig 5.

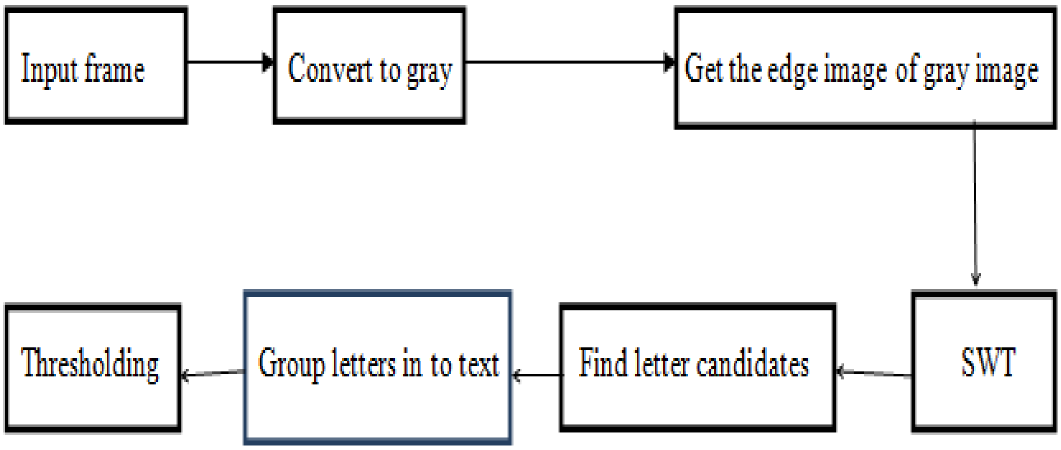


Fig 5 Block diagram of text detection algorithm

3.2.6 Thresholding:

Thresholding is a process of converting a grayscale image to a bi-level image by using an optimal threshold. Thus the objective of binarization is to mark pixels that belong to true foreground regions with a single intensity and background regions with different intensities. After grouping the letters into text the image is applied thresholding.

3.3 TEXT RECOGNITION AND AUDIO OUTPUT:

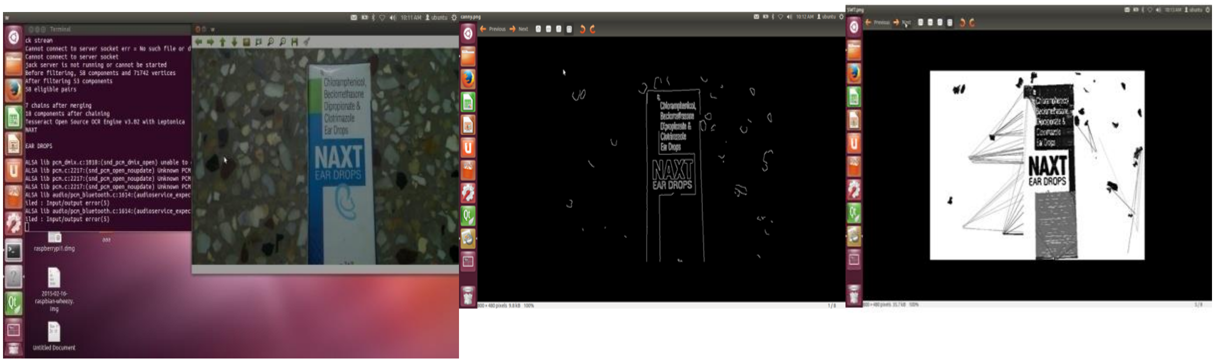
Feed the output of text detection to tesseract OCR for recognizing the text. Tesseract OCR (Optical Character Recognition) is used to extract text from an image or a scanned document. The recognized text codes are recorded in text files. Then, the e-speak engine load these files and gives the audio output of text information through the ear phone connected to the audio jack port of raspberry pi.

4. RESULTS

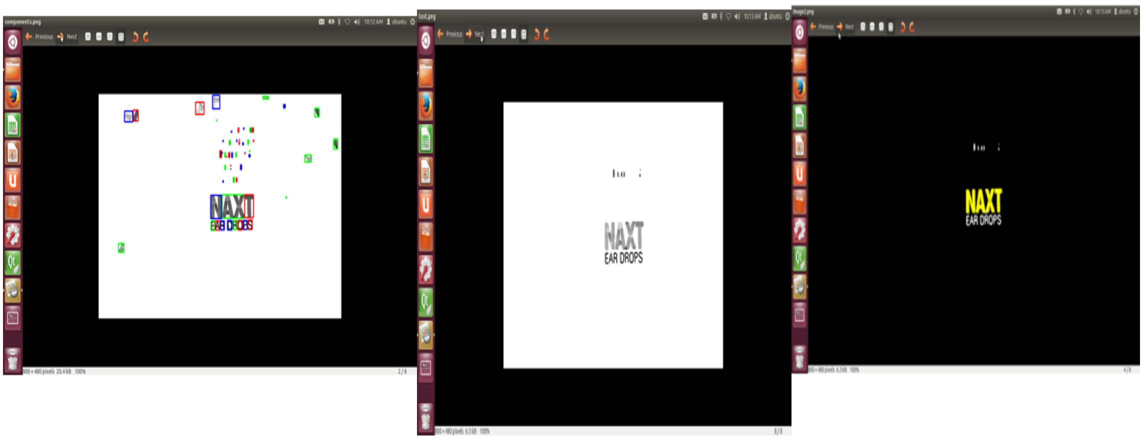
The Fig 6 depicts the entire original setup of how the connectedsystem appears. The Laptop acts as a remote to invoke the program. Linux is the platform used where OPEN_CV library is used. The blue wire depicts the LAN interfacing between the OS loaded in the memory card present in the kit and the laptop (Remote).

Fig 6.Prototype of our proposed system

The fig7 shows the result of the proposed system. The proposed system ensures to read printed text on hand-held objects for assisting blind persons. The text detector does not handle round and curved letters and curved lines of text produce weak results. Another weakness was that small and close letters tend to be grouped together in the SW labeling phase. Since a group of letters behaves differently than a single letter, these groups may be dismissed in the 'finding letter candidates' phase, the improvement can be found in the 'Finding Letter Candidate' phaseand in the 'Grouping Letters into Regions' phase.



(a) **(b)** **(c)**



(d) **(e)** **(f)**

Fig 7 Stages of text detection algorithm (a) Input frame (b) Canny edge detection (c)Stroke width detection(d) Connected components (e) Detected text (f)Thresholding

5. CONCLUSION

This paper presented a camera-based label reader to help blind persons to read names of labels on the products. Camera acts as main vision in detecting the label image of the product then image is processed internally and separates label from image by using open CV library and received label image is converted to text by using tesseract OCR. The converted text is converted to voice to hear label name as voice through ear phones connected to audio jack port using e speak engine

In many ways the result of the project are both surprisingly good and surprisingly bad. For images without definite edges the program may not work properly. But it will work perfectly for image texts which have prominent edge. For the product with fancy font, transparent text, text that is too small, blurred text, and for nonplanar surface it will not work properly.

The labeling algorithm needs to be improved.A better labeling method of components could improve the detection of characters. This could get better results for circular text, which tends to be dismissed as noise due to the grouping of the letters.

REFERENCES

[1] “Vision based assistive system for label detection with voice output” by Vasanthi.G and Ramesh Babu, International journal of innovative research in science, engineering and technology, volume 3, special issue 1, january 2014.

[2] “Intelligent system based on speech recognition with capability of self-learning” by Ms. Sneha K. Upadhyay, and Mr. Vijay N, International journal for research in applied science and engineering technology.

[3] ”Darshan: Electronics guidance for the navigation of visually impaired person” by MarutTripathi, Manish Kumar, Vivek, and Warshakandlikar, National institute of electronics and information technology, Aurangaba.

[4] “Portable, robust and effective text and product label reading, currency and obstacle detection for blind persons” by Asha Mohandas, Bagyalakshmi. G, and Manimala. G, International journal for technological research in engineering, volume 1, issue 9, may-2014 “Text information extraction in images and video: a survey” by Keechuljung, Kwang in Kim, Anil k. jain.

[5] “Portable camera-based assistive text and product label reading from hand-held objects for blind persons” by Chucaiyi, student member, IEEE, Vinglitian, senior member, IEEE, and Aries Arditi.

[6] “Detecting Text in Natural Scenes with Stroke Width Transform” by Boris Epshtein, EyalOfek, Yonatan Wexler Microsoft Corporation.