

Embedded Product Label Reading System using SIFT and Line Detection Algorithm

V. Ramya

Department of Computer Science
and Engineering
Annamalai University,
Annamalainagar, India

M. Kokila Sundari

Department of Computer Science
and Engineering,
Annamalai University,
Annamalainagar, India

C. Gayathri Devi

Department of Computer Science
and Engineering,
Annamalai University,
Annamalainagar, India

Abstract- In order to help the physically challenged visually impaired people to read text labels on product packaging, a product label reading system is proposed. To increase the clear visibility of the text with low time complexity, Scale Invariant Feature Transform (SIFT) and Line detection algorithm is applied. The proposed work is carried out at different levels, such as, preprocessing, segmentation, feature extraction, text extraction, and text recognition, Audio output. The input image is preprocessed to improve the quality of that picture. Median filter is applied to enhance the image, borders of the image are observed using the Sobel edge detection method. Segmentation is performed to segment the images. To extract the particular text region of the image scale invariant feature transform is performed and text extraction is employed to draw out parts by one another one. The extracted product label text is recognized by optical character recognition. The audio output of the recognized product label is provided for the benefit of the visually impaired.

Key words- SIFT Line detection, Median filter, Sobel edge detection, and optical character recognition.

I. INTRODUCTION

The visually impaired people worldwide are around 314 million, were 45 million are blind [1]. In today's developing world reading is really important, printed text is everywhere in day to day life, such as getting a product package, to find a manufacture date and expiry date of the product, receipts, bank statement and so on. There are several systems designed to help the visually impaired people, but to handle product labeling is found difficult. The existing system to find the product is by using bar code reader, where the user identifies the product [2] from the large database using speech. The drawback is that the visually impaired people find it hard to position the barcode [3].

II. PROPOSED SYSTEM

The overview of the proposed system is designed in Figure 1. The image of the product is captured using a camera, in proposed work the laptop camera is used. The input picture is captured lively, it is then preprocessed to improve quality of the image and by using a median filter image is

enhanced. The Sobel edge detection method is applied to detect the edges. The segmentation is performed to segment the image. To extract the particular text region of the image scale invariant feature transform is performed and text extraction is employed to draw out parts by one another one. Optical character recognition is used to binarize and recognize the text regions that are localized by the text characters.

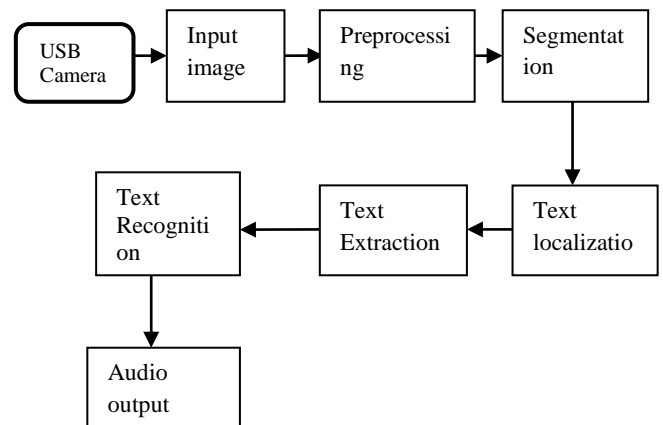


Fig 1. Block diagram of the proposed system

III. SYSTEM DESIGN

A. Preprocessing

Preprocessing of the image starts with elimination of noise and image enhancement, the captured image may contain a large amount of noise which must be removed for smooth processing, the captured color image is preprocessed to gray scale image in order to reduce the computational time and image is enhanced to improve the interoperability of information included in the image for human viewers. Both noise elimination and image enhancement is performed by median filter [4, 5].

1) Median Filter

In signal processing, noise reduction is often desirable on an image or signal. Noise reduction is a typical

preprocessing step which improves the results of edge detection on an image. To remove a noise median filter is used, which is a nonlinear digital filtering technique. The idea behind the median filter is used to run through the pixel of the image and replacing the pixel value with the median if neighboring pixel matches. Median filter also helps in preserving edges during the noise removal process.

2) Sobel Edge Detection

Sobel is a popular edge detection method which uses the derivative approximation to find edges. It returns edges at those points where the gradient of the considered image is maximized. The horizontal and vertical gradient matrices whose dimensions are 3X3 for the Sobel method has been broadly used in the edge detection operations. Each direction of Sobel mask is utilized to an image, and then two new images are made. One image shows the vertical response and other image shows the horizontal response. The value of the threshold is used to detect edge pixels.

B. Segmentation

The processes of partitioning a digital image into multiple segments (i.e. sets of pixels) are segmented. The concept of the segmentation is to modify the representation of an image into meaningful and easier to analyze. The region growing segmentation technique is performed.

1) Region Growing

This method mainly relies on the neighboring pixels with similar pixel values in one region. If the similarity in pixel is identified, cluster is made as one or more of its neighbors.

2) Histogram

The range of data is split into equal size bins to construct a histogram. For each bin, the number of points of data set is considered. Frequency is plotted along the vertical axis, while response variable is plotted along the horizontal axis.

C. Feature Extraction

To observe and describe local features in the images, Scale Invariant Feature Transform (SIFT) algorithm is introduced. The SIFT is invariant to translations, rotations and scaling in the image area and robust to moderate perspective transformations. There are primarily four steps involved in SIFT algorithm.

1) Detection of Scale Space Extrema

In order to detect larger corners, large windows are needed, for this purpose scale-space filtering is used. Laplacian of Gaussian is found in the image with various σ values. $G(x,y,\sigma)$ with input image $I(x,y)$ If the difference of Gaussian is found images are searched for local extrema over scale and space. If it's local extrema it is a potential key point. The key point is best represented in that scale.

2) Keypoint Localization

The potential keypoint locations have to be refined to get accurate results. If the strength of the extrema is less than a threshold value, it is eliminated. The deviation of Gaussian has higher response edges which also necessitate to be taken out. The interest points remain strong by eliminating low contrast keypoints and edge keypoints.

3) Orientation assignment

An orientation is set apart for each key point in order to achieve invariance of image rotation.

4) Keypoint Descriptor

By now keypoint descriptor is created. The region around the keypoint is assumed. It is divided into subblocks 4X4 size, 8 bin orientation of histogram is created, and an aggregate of 128 bin values is achieved. It represents a vector to form keypoint descriptor.

D. Text Extraction

Line Detection algorithm using character extraction.

1) Line Detection Algorithm

- A line is a connected group of aligned edge pixels, and a corner is a group of edges that form together two jointive non-parallel lines [12, 13]. Line-detection algorithms based on contour extraction usually involve the following stages of image processing: smoothing, edge detection, edge thinning, edge linking, chain straightening and line correction.
- While edges (i.e. boundaries between regions with relatively distinct gray levels) are by far the most common type of discontinuity in an image, instances of thin lines in an image occur frequently enough that it is useful to have a separate mechanism for detecting them.
- Hough transform can be applied to detect lines; yet, in that instance, the end product is a parametric description of the line in an image.

Equation of a line

$$Y=mx+c$$

E. Text Recognition

Text recognition is performed by off-the-shelf Optical character recognition (OCR)[1]. OCR prior to output of informative words from the localized text regions. A text region labels the minimum rectangular area for the accommodation of characters inside it, so the border of the text region contacts the edge boundary of the text character. However, the experiments show that OCR generates better performance if text regions are first assigned proper margin areas and binaries to segment text characters from background.

F. Audio output

The given product is recognized by using SIFT and Line detection algorithm. The recognized product is embedded unit which generated by P89C51, APR9600 and speaker. In micro controller is a program with the 10 product name and corresponding voice is stored in APR9600. A micro controller receive the signal from the pc through the RS232. Based on that received signal audio output played through Speaker under the control of microcontroller.

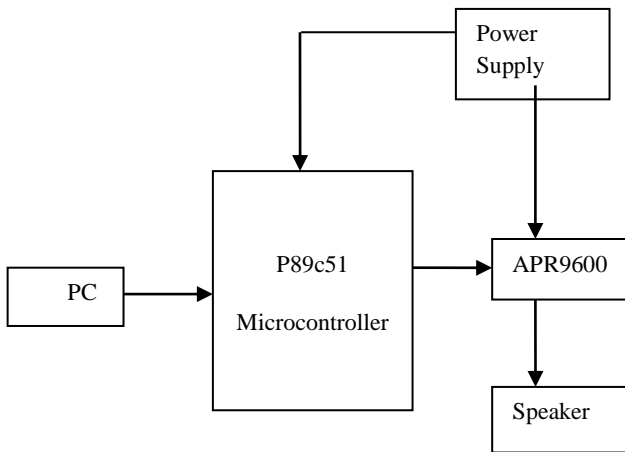


Fig 2. Block diagram of vision based Assistive System

IV. HARDWARE DESCRIPTION

A) Component description

- P89C51 Micro Controller
- APR9600
- Speaker

1) P89C51 Micro Controller

A Single-Chip 8-Bit Microcontroller manufactured in advanced CMOS process. On-chip Flash Program Memory with In-System Programming (ISP) and In-Application Programming (IAP) capability. The device also has four 8-bit I/O ports, three 16-bit timer/event counters. The added features of the P89C51 make it a powerful microcontroller for applications that require pulse width modulation, high-speed I/O and up/down counting capabilities such as motor control.

2) APR9600

APR device is a single chip voice recording. The playback capability for 40 to 60 seconds. The device is support both random and sequential access of multiple messages. The sample rates are user selectable, allowing designer to customize their design for unique quality. The device is ideal for use in portable voice recorders, many other consumer and industrial application.



Fig 3 Image of APR9600

3) Speaker

Speakers are one of the common devices. The purpose of speakers is to produce audio output that can be heard by the listener. Speakers are transducers that convert electromagnetic waves into sound waves. The speakers receive audio input from a device such as an audio receiver.

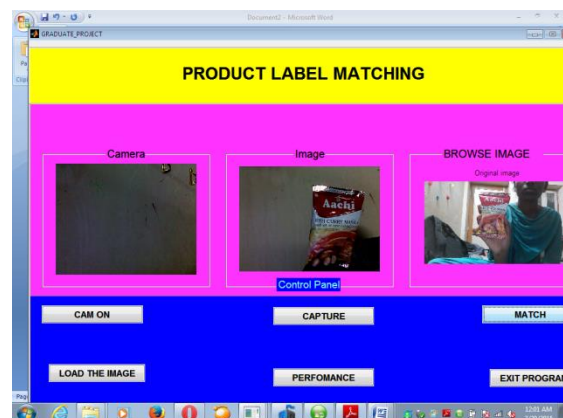


Fig 4 .Speaker

V. RESULT AND DISCUSSION

Step 1:

This is the first stage of the proposed work and it is used for identifying the input image. This phase consists of three modules, first one is a camera, it is employed to check that the camera will be acting right or not, and second stage is to capture the product image, and third stage is used to browse the image that is already stored.

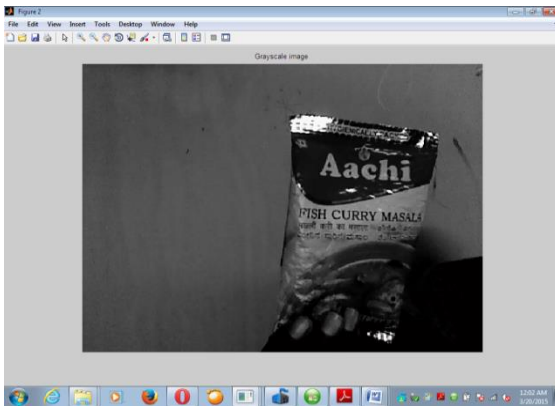


Step 2:

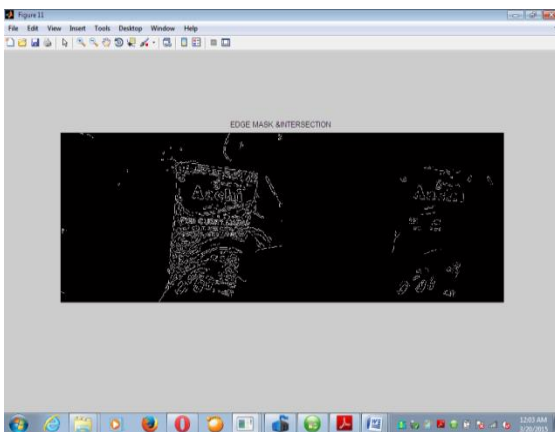
In step 2, the original image will be input to the system for preprocessing,



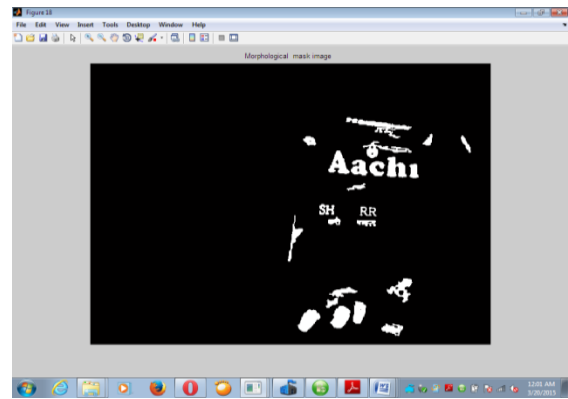
Step 3:
The original picture is converted into a grayscale picture for preprocessing, because the color image requires a lot of computational time.



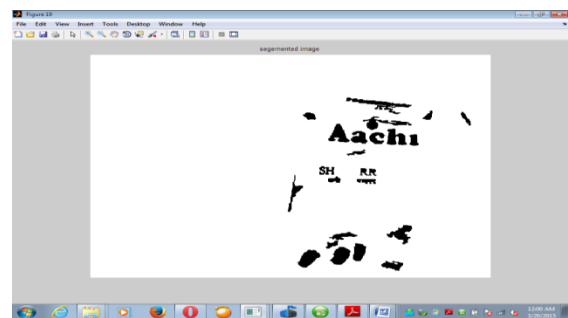
Step 4:
The text is filtered from the edge mask image



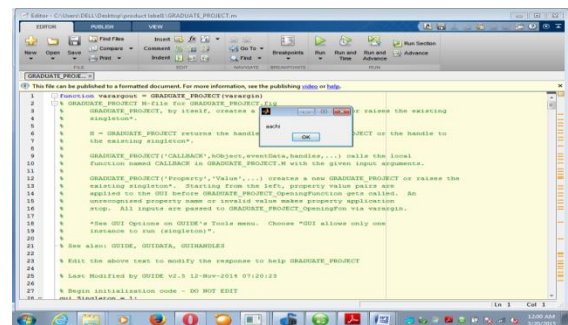
Step 5:
The region based segmentation is used for segmenting the image. This method detects region directly because it is used to extract the image region from the location of starting point and also segmented the image pixel by pixel



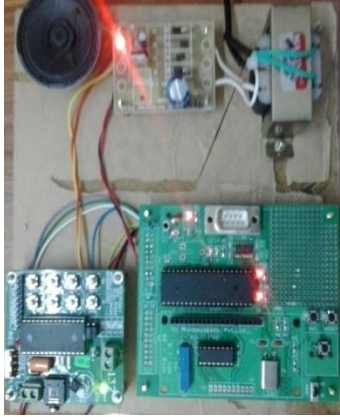
Step 6:
This step is used to detect the text.



Step 7:
After text localization the text is displayed in the message text box.



Step 7:
After text recognition the text codes are send to micro controller through RS232. corresponding voice is stored in APR9600.A micro controller receive the signal from the pc through the RS232.Based on that received signal audio output played through Speaker under the control of microcontroller.



CONCLUSION

The proposed system to read printed text on hand held objects for assisting visually impaired people. The input image is taken live and it is preprocessed to improve the quality of the image and to enhance the image using median filter, and then detect the edges using the Sobel edge detection method. At the next stage, segmentation is performed to segment the image and then scale invariant feature transform algorithm is used to localize text in the camera based image, and then text extraction is used to extract the character one by one. The optical character recognition is used to perform word recognition on the localized text regions and transform into audio output to blind users in speech.

REFERENCES

- [1]. C. Yi and Y. Tian, "Assistive text reading from complex background for blind persons," in *Proc. Int. workshop Camera-Based Document Anal. Recognit.*, 2011, vol. LNCS-7139, pp. 15–28.
- [2]. 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.
- [3]. S. Shoval, J. Borenstein, and Y. Koren, "Auditory guidance with the Nav-belt: A computerized travel for the blind," *IEEE Trans. Syst., Man, Cybern. C. Appl. Rev.*, vol. 28, no. 3, pp. 459–467, Aug. 1998.
- [4]. J. Zhang and R. Kasturi, "Extraction of Text Objects in Video Documents: Recent Progress," In *IAPR Workshop on Document Analysis Systems*, 2008.
- [5]. N. Nikolauou and N. Papamarkos, "Color Reduction for Complex Document Images," *International journal of Imaging system and Technology*, Vol.19, pp.14-26, 2009.
- [6]. N. Otsu, "A threshold selection method from gray-level histograms," In *IEEE Trans on system, man and cybernetics*, pp. 62-66, 1979.
- [7]. L. Ma, C. Wang, B. Xiao, "Text detection in natural images based on multi-scale edge detection and classification," In *the Int. Congress on Image and Signal Processing (CISP)*, 2010.
- [8]. B. Epshtein, E. Ofek and Y. Wexler, "Detecting text in natural scenes with stroke width transform," In *CVPR*, pp. 2963-2970, 2010.
- [9]. S. Kumar, R. Gupta, N. Khanna, S. Chaudhury, and S. D. Joshi, "Text Extraction and Document Image Segmentation Using Matched Wavelets and MRF Model," *IEEE Trans on Image Processing*, Vol. 16, No. 8, pp. 2117-2128, 2007.
- [10]. C. Stauffer and W. E. L. Grimson, "Adaptive background mixture models for real-time tracking," presented at the IEEE Comput. Soc. Conf. Comput. Vision Pattern Recognit, Fort Collins, CO, USA, 1999.
- [11]. X. Chen and A. L. Yuille, "Detecting and reading text in natural scenes," In *CVPR*, Vol. 2, pp. II-366 – II-373, 2004.
- [12]. X. Chen, J. Yang, J. Zhang and A. Waibel, "Automatic detection and recognition of signs from natural scenes," In *IEEE Transactions on image processing*, Vol. 13, No. 1, pp. 87-99, 2004.