Camera-Based Assistive Text and Product Label Reading For Blind Person

Sachin Bansode
Department of Computer Engineering
Sinhgad Institute of Technology and Science
Pune, India

Bhagyashri Patil
Department of Computer Engineering
Sinhgad Institute of Technology and Science
Pune, India

Shivanand Kampali
Department of Computer Engineering
Sinhgad Institute of Technology and Science
Pune, India

Niranjan Dhumal
Department of Computer Engineering
Sinhgad Institute of Technology and Science
Pune, India

Prof. Mrs. B. D. Shendkar
Department of Computer Engineering
Sinhgad Institute of Technology and Science
Pune, India

***Abstract*—We propose a camera-based assistive text reading framework to help blind person or low visual power to read text and product label or document from hand-held objects in their day to day life. We have an efficient and effective method which is used to define region of interest (ROI). by a mixture-of-Gaussians-based background subtraction technique we subtract the text from the clattered background .To automatically focus the text regions from the object ROI, we offer a stroke width transform (SWT) algorithm this algorithm is help to recognize the character by their shape and width by calculating each pixel by their start to end point Is a local image operator which computes per pixel the width of the most likely a stoke containing the pixel. The output of SWT is an image of size equal to the size of the input image where each elements contains the width of the stroke associated with the pixel text reorganization is performed by off-the-shelf optical character reorganization (OCR) prior to output of information words from the localize text region. The recognized text code are recording Script file, then we employee the Free Text to speech (TTS) library to load these file and display the audio output of text information. Blind person can adjust speech rate, volume and tone according to their preference. We explore the user interface issues and robustness of the algorithm in extracting and reading text from different objects with complex backgrounds.**

***Keywords*—Assistive device, Visual impact, Blindness, Hand-held objects, Distribution of edge pixel , Optical character recognition(OCR), Stroke width transform(SWT), Text reading , Text region localization , Background subtraction, Text to speech(TTS), Audio output**

I. INTRODUCTION

Of the 285 million people are visually impaired worldwide, 39 millions are blind. Even in developing country like India, 2015 Blind people Association survey reported that a 12 million people are blind. Recent development in computer technology, computer vision, digital camera that makes it feasible to assist these technology by developing such a new system like portable camera based product which combines the computer technology with existing product such as optical character recognition (OCR).Reading is one of the important factor in today's society. The printed text is anywhere in the form of reports, bank document, receipts, hotel menus, classroom boards, product label, bottle label, etc. And using such system like video magnifiers, screen readers help blind person and those with low vision to access the documents and text. The ability of people who are blind or have low visual impairments to read printed labels and documents will enhance independent living and social self-sufficiency.

Today, there are many system that have promise to portable use, but they cannot provide the product labelling. Such system are bar code reader which help to blind person to identify the different product .Database can gives the permission to access the information for blind persons about the product through speech .But there is big limitation for blind person to find the position of bar code on the product. Some assistive systems like pane scanner which is used in some situations. Such systems integrated with OCR software having functions to scanning and recognition of the text and have integrated voice output. These system generally design to read the text from simple backgrounds, standard fonts and also small range of fonts. Some systems need only white background for scanning the text. These system cannot read the text from the complex background.

To assist blind person for reading the text from any hand held object, we conceived of a camera-based assistive text reading framework to track the object of interest with camera view and extract text information from the object. Our proposed algorithm can effectively and efficiently handle complex background and multiple pattern and extract the information from the object or any document as well as product label.

In assistive text reading for blind person, it is one of big challenge for user to position the object of interest within the camera's view to make sure that the object appears in the center of the camera' view with sufficient wide angle to confirm user with only approximation aim. To extract the object from the camera image we develop a motion based method known as region of interest (ROI) then we perform text recognition in this ROI.

It is challenging problem to localize text and object ROI's from captured image with complex background which is surrounded by 'noise' and text. Many algorithm is developed for text localization. We divide them into two categories: learning-based and rule-based. Learning-based algorithm, model text structure and extract text feature to produce text classifier. In other hand rule-based algorithm is apply for pixel-level image processing to extract information from text layouts such as font, character size etc.

To extract the text information from the complex background we here propose a text localization algorithm that combines rule-based and learning-based text classifier training, which is used to distinguish text character from background outliers.

II. LITERATURE SURVEY

Detecting text regions in natural scenes is an important part of computer vision. A novel text detection algorithm is proposed that extracts six different classes' features of text, and uses Modest Ada Boost with multi-scale sequential search. Experiments show that our algorithm can detect text regions with af= 0.70, from datasets which include images with text of various fonts, sizes, colours, alphabets and scripts.

A text detection method is proposed based on a feature vector generated from connected components produced via the stroke width transform. Several properties, such as variant directionality of gradient of text edges, high contrast with background, and geometric properties of text components jointly with the properties found by the stroke width transform are considered in the formation of feature vectors. Finally, the obtained text components are grouped and the remaining components are discarded. Since the stroke width transform relies on a precise edge detection scheme, we introduce a novel band let-based edge detector which is quite effective at obtaining text edges in images while dismissing noisy and foliage edges.

Text that appears in images contains important and useful information. Detection and extraction of text in images have been used in many applications. We propose a multi scale edge-based text extraction algorithm, which can automatically detect and extract text in complex images. The proposed method is a general-purpose text detection and extraction algorithm, which can deal not only with printed document images but also with scene text. It is robust with respect to the font size, style, color, orientation, and alignment of text and can be used in a large variety of application fields, such as mobile robot navigation, vehicle license detection and recognition, object identification, document retrieving, page segmentation, etc.

III. SYSTEM APPROACH

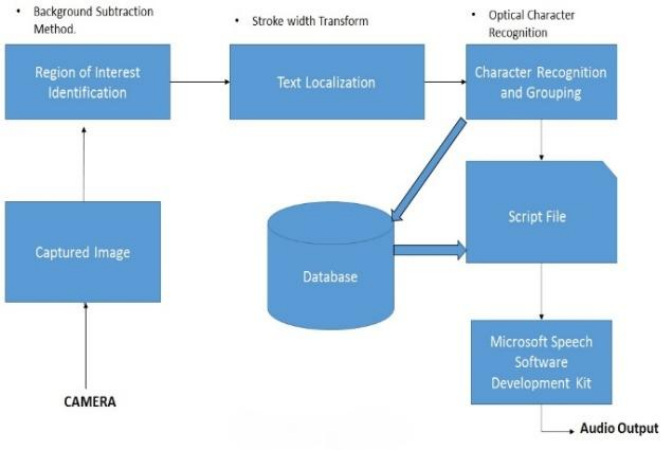

Fig. 1. Architecture Diagram.

A. *Camera*

Camera is one of the most important constituent in our system. Camera is used in order to scan particular product or a thing containing text labels. We have to hold the component having the label to read in front of the camera and just shake the particular component.

Fig. 2. The portable camera.

B. *Region Of Interest Identification (ROI)*

Region of interest is a particular area selected from the image which we want to focus on or in which we are interested. In our system we need to find the ROI to locate the text present on the surface so we use one of the most popular method called Gaussians mixture based Background

subtraction method. Our basic aim is to find the ROI having label text which is to be read.

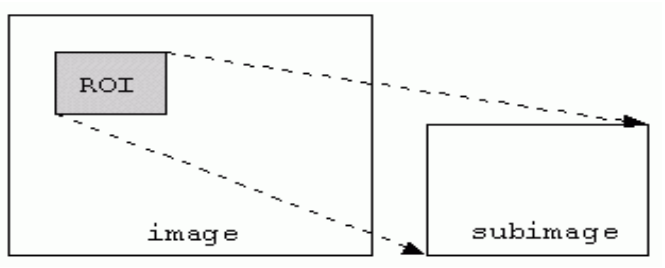

Fig. 3. Region of interest of a particular image.

C. Text Localization

Text localization is the step where we localize the text in the captured image. In this step we find the text in the ROI. In our system we are using the Stroke Width Transform Method (SWT) for text localization. It is one of the robust method used for text localization. Stroke Width Transform is an algorithm that uses Open CV to try to separate the text out of images that contain embedded text.

D. Character Recognition and Grouping

When we get the particular text we group them and form the sequence of text and try to recognize that text. For this we can use Optical Character recognition method (OCR). OCR converts the typed, handwritten or printed text into machine-encoded text.

E. Free Text To Speech(TTS)

Free Text To Speech (TTS) is a software that allows to build speech applications for Windows operating system. It is the Text-to-Speech (TTS) Engine that can convert written text to audio output.

F. Audio Output

Free TTS provides us and text in audio form. Using headsets or audio system person can here the output.

IV. PROPOSED SYSTEM

A. Mixture Of Gaussians Based Background Subtraction Method

To ensure that the hand-held object appears in the camera view, we used a camera with a reasonably wide angle in our purposed system. To extract the hand-held object of interest from other objects in the camera view, we need to shake the objects in front of camera. Which text they wish to identify and then apply a motion-based method to localize the text from cluttered background.

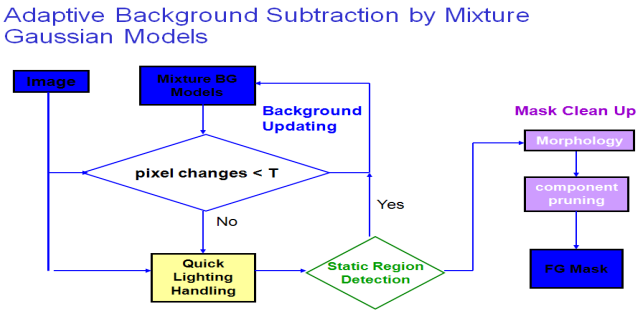

Fig. 4. Background subtraction method.

Background subtraction method is conventional method to detect moving objects in a dynamic scene, many background subtraction method can be available for subtract the background.

B. Text Localization Algorithm

Text character is formed with the group of pixel with variable as well as constant orientations. In our proposed system we are using a method dependent on stroke orientations, this method is used to describe the structure of a particular text character. According to the pixel level analysis of the text, stroke orientation is perpendicular to the gradient orientations at pixels of stroke boundaries, we propose a new operator to map a gradient feature of strokes to each pixel to model the text structure by stroke orientations. This extends the local structure of a stroke boundary into its neighborhood by gradient of orientations. We use this to develop a feature map in order to analyses global structures defined for specific text characters.

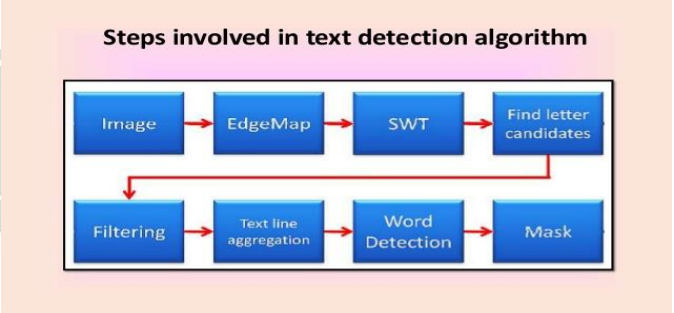

Fig. 5. Text detection algorithm.

C. Stroke Width Transformation

The Stroke Width Transform (SWT) is a one of the most efficient method which is a local image operator which computes per pixel width of the most likely stroke containing the pixel. SWT gives an output that is an image of equal size to the size of the input image where each element contains the width of the stroke associated with the respective pixel.

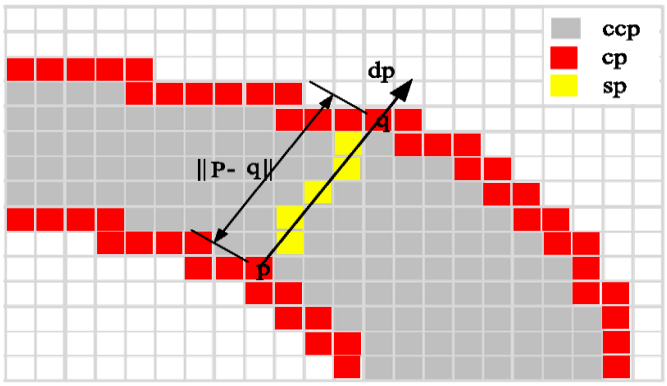

Fig. 6. Stroke width transformation.

D. Optical Character Recognition

For increasing the reliability of the algorithm, we consider groups of letters. Usually single letters do not appear in images, finding such group is a remarkable filtering mechanism and due to this unusually scattered noise can be removed. Basically the text appears in a linear form. There are some similarities expected from text on a line, which also includes similar stroke width, height, letter width and spaces between the letters and words. For the possibility of belonging to the same text line, each pair of letter candidates is considered. Two letter candidates should have similar stroke width. The height ratio of the letters must not pass 2.0. The distance between letters must not overreach three times the width of the wider one. The candidate pairs regulated above are clumped together into chains at the next step of the algorithm. Initially, each chain consists of a single pair of letter candidates. If two chains share one end and have similar direction, then they can be unified together. When no chains can be unified, the process ends. Each produced chain of adequate length is considered to be a text line. At last, using a heuristic that determines a histogram of horizontal distances between consecutive letters and estimates the distance threshold that separates intra-word letter distances from inter-word letter distances, the text lines are broken into separate words.

Optical Character Recognition (OCR) is a technology that can be used to convert different types of texts present in the documents, images, hand held products etc. captured by a digital camera into searchable data and editable data.

These systems are generally designed to perform best with document images with simple backgrounds as well as complex backgrounds, standard fonts, a small range of font sizes.

In our system we are using the latest version of OCR i.e. ABBYY Fine Reader supports adaptive recognition technology which is specifically designed for processing camera images.

This software is more robust than other ones.

E. Free Text To Speech (TTS)

The text codes which recognized in the previous steps are recorded in script files. Then, we load these script files in Free Text to Speech (TTS) Kit and obtain the audio output of recognized text information. Blind users are provided the functionality to adjust volume, tone and speech rate, according to their preferences.

CONCLUSION AND FUTURE SCOPE

In this paper, we have described a system which helps blind person to read the text from the hand-held object and different documents. In order to solve the common problem for blind users, we have describe a motion-based method to detect the object of interest, while the blind user or person simply shakes that object for a couple of seconds in front of camera. This method can effectively distinguish the object of interest from background or other objects in the camera view. To extract the text from background here we used stoke width transform algorithm based on edge destitution and Stoke orientation. OCR can be used to performed word character recognition and by sing the Microsoft development software kit we generate the output. Our future work is to recognize the text from the more complex background and different type of challenging background surfaces and also we will try to implement such system which can read the text from the object and translate output in different languages.

REFERENCES

[1] "Portable Camera-Based Assistive Text and Product Label Reading from Hand-Held Objects for Blind Persons" Chucai Yi, Student Member, IEEE, Yingli Tian, Senior Member, IEEE, and Aries Arditi IEEE/ASME TRANSACTIONS ON MECHATRONICS, VOL. 19, NO. 3, JUNE 2014

[2] International Workshop on Camera-Based Document Analysis and Recognition (CBDAR 2005, 2007, 2009, 2011). [Online]. Available: http://www.m.cs.osakafu-u.ac.jp/cbdar2011/

[3] 3.Shahab, F. Shafait, and A. Dengel, "ICDAR 2011 robust reading competition: ICDAR Robust Reading Competition Challenge 2: Reading text in scene images," in Proc. Int. Conf. Document Anal. Recognit., 2011, pp. 1491–1496

[4] R. Manduchi and J. Coughlan, "(Computer) vision without sight," Commun. ACM, vol. 55, no. 1, pp. 96–104, 2012

[5] Yi and Y. Tian, "Text string detection from natural scenes by structure basedpartitionandgrouping, "IEEETrans.ImageProcess., vol.20,no.9, pp. 2594–2605, Sep. 2011.