International Journal of Advanced Information in Engineering Technology (IJAIET) ISSN: 2454-6933
Vol.3, No.5, May 2016

Camera Based Text Detector for Visually Impaired

Bibin Thomas
Dept. of Electronics and Communication
Government Engineering College
Barton Hill, Trivandrum

Afsal S
Dept. of Electronics and Communication
Government Engineering College
Barton Hill, Trivandrum

Linomon Joseph
Dept. of Electronics and Communication
Government Engineering College
Barton Hill, Trivandrum

***Abstract*—The popularity of digital image and video is increasing rapidly. With Recent developments in computer vision, digital cameras, portable computers and wearable devices we are able to assist Visually Impaired individuals by developing camera based products that combine computer vision technology with other commercial products to help them read small text labels on hand held object in daily life or detect currency he is holding or different colours.To isolate the object of interest from backgrounds or other surrounding objects in the camera view a motion based method to define a region of interest (ROI) in the video by asking the user to shake the object. Thus we can extract the hand held object that he is holding. In the extracted ROI, text localization and recognition are conducted to acquire text information. To automatically localize the text regions from the object ROI, text localization algorithm is uses by using gradient features of stroke orientations and distributions of edge pixels . Text characters in the localized text regions are then binarized and recognized using optical character recognition.The recognized text codes are output to blind users in speech.**

Keywords—Computer Vision,Region of interest, text recognition,

I. INTRODUCTION

The World Health Organization (WHO) estimated in 2010, [1] the number of people visually impaired was estimated to be 285 million, of whom 39 million were blind. Over the last decades, visual impairment and blindness caused by infectious diseases have been greatly reduced (an indication of the success of international public health action), but there is a visible increase in the number of people who are blind or visually impaired from conditions related to longer life expectancies. The great majority of visually impaired people are aged 65 years or older. It is estimated that there is a per-decade increase of up to 2 million persons over 65 years with visual impairments. This group is growing faster than the overall population.

Advances of technology and better knowledge in human psycho-physiological 3D world perception permit the design and development of new powerful and fast interfaces assisting humans with disabilities. For the blind, research on supportive systems has traditionally focused on two main areas: information transmission and mobility assistance. More recently, computer access has been added to the list .Problems related to information transmission concern reading, character recognition and rendering graphic information about 2D and 3D scenes. The most successful reading tool is the Braille dot code. There are few camera based blind-assistive systems which are able to help blind or visually impaired people find their daily necessities. Initial blind assistant system focuses more on navigation, path planning and blind person tracking.DimitriosDakopoulos and Nikolaos G. Bourbakis[2] introduce a "Wearable Obstacle Avoidance Electronic Travel Aids for Blind"this assist visually impaired people during navigation in known or unknown, indoor or outdoor environments. Afterwards some blind assisting system for object findings was introduced. Chucai Yi [3] proposes a prototype system of blind-assistant object finding by camera-based network and matching-based recognition. Dataset of daily necessities are collected and apply Speeded Up Robust Features (SURF) and Scale Invariant Feature Transform (SIFT) feature descriptors to perform object recognition. Along with this blind assistant system a number of reading assistants were also designed specifically for the visually impaired. MajidMirmehdi [4] describes about reading text from wearable computing, robotic vision or as an aid for visually handicapped people. It present an automatic text reading system using an active camera focused on text regions. Then a number of images are captured over the text region to reconstruct a high-resolution mosaic of the whole region. This magnified image of the text is good enough for reading by humans or for recognition by OCR. Even with a low resolution camera we obtained very good results. But this system cannot read text from challenging patterns and background found on many everyday commercial products .A better system was introduced by Chucai Yi,[5][6] which is also a camera-based assistive text reading framework to read text labels and product packaging from hand-held objects in their daily lives. In assistive reading systems for blind persons, it is very challenging for users to position the object of interest within the center of the cameras view. To make sure the hand-held object appears in the camera view, [5] use a camera with sufficiently wide angle to accommodate users with only approximate aim. This may often result in other text objects appearing in the camera view .To extract the hand-held object from the camera image, [5] develop a motion-based method to obtain a region of interest (ROI) of the object. Then, it perform text recognition only in this ROI. It is a challenging problem to automatically localize objects and text ROIs from captured images with complex backgrounds, because text in captured images is most likely surrounded by various background outlier noise, and text characters usually appear in

multiple scales, fonts, and colors. For the text orientations, [1] assumes that text strings in scene images keep approximately horizontal alignment.

Text detection and extraction is another challenging problem. Michael R. Lyu,[7] does a detailed analysis of multilingual text characteristics, including English and Chinese. He proposes a comprehensive, efficient video text detection, localization, and extraction method, which emphasizes the multilingual capability over the whole processing. The method is also robust to various background complexities and text appearances. The text detection is carried out by edge detection, local thresholding, and hysteresis edge recovery. Limitation of this method is that it cannot detect motion texts due to the assumption of stationary text. Dutta,&Shivakumara [8] speaks about a Gradient based Approach for Text Detection in Video Frames. First the gradient of the Image in calculated and then enhance the gradient information. Later binarized the enhanced gradient image and select the edges by taking the intersection of the edge map with the binary information of the enhanced gradient image. Here canny edge detector is used for generating the edge map. The selected edges are then morphologically dilated and opened using suitable structuring elements and used for text regions. Then perform the projection profile analysis to identify the boundary of the text region. At the end, implement a false positive elimination methodology to improve the text detection results . After detecting this text region we need to convert it to black and white image to convert it to text .There we require a threshold selection because the different regions will have different background an so to separate the ROI from the background the threshold should be chosen carefully.

II. FRAMEWORK AND ALGORITHM OVERVIEW

The system framework consists of four functional components: audio input, scene capture, data processing, and audio output. The system is trained with certain voice commands for recognizing text currency or colour.The system works according to specific command. Soon after the system recognize the voice commend the scene capture component collects scenes containing objects of interest in the form of images or video.

Fig. 1. Snapshot of the demo system

It corresponds to a camera attached to a pair of sunglasses. The object of interest is first extracted from the captured image. This is done by asking the user to shake the object that he is holding. Multiple images are taken while the user shake what he is holding and this images are used to detect the object of interest from the wide frame which is captured using the camera. The data processing system performs according to the voice command. For text recognition the processing unit search for text regions in the captured image. For currency detection certain processing is done to detect the currency. After detection processing next is recognition .The detected data is recognized and converted to text. This is further converted to speech and fed to the visually impaired one through a head phone. A laptop is used for the data processing unit. Fig 2 depicts a work flowchart of the system.

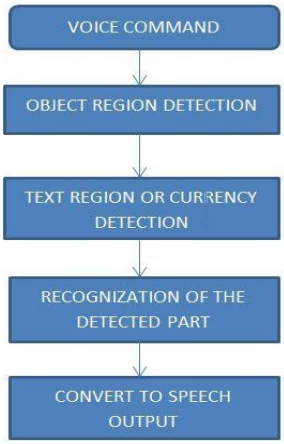

Fig. 2. Flowchart of the framework

III. VOICE COMMAND RECOGNITION

The system doesn't know what the user want to detect. Sotheuser needs to tell the system what it wants to detect using Voice commands. This is done by employing a voice command recognition system. Voice command recognition make the system more interactive to the user and more focused on what it is meant to be done. Voice command recognition is done on artificial neural network.1st the net file is created for each voice commands. Each net file is created by training the network with the corresponding command as positive commends and the remaining command as negative. Each commend is recorded for a period of one second and Mel frequency cepstral coefficients (MFCCs) is computed for each recorded audio commends and this values are used for training the net file for each voice commands. During recognition also one second duration audio is recorded and MFCC is computed for the recorded audio which is then simulated with each net file. The output of each net file is compared to recognize the input voice command.

IV. OBJECT REGION DETECTION

To extract hand held object of interest from other object in camera view an image subtraction method is employed here.ie, the user is asked to shake the hand-held objects containing the text that they wish to identify. Multiple images are taken while the user shake the object. This images are used to localize the region of interest by subtracting this images

from each other which highlights the moving region which is our region of interest.

Steps

1) We take multiple images of the object while the user shake it
2) We make the difference images with different combinations and take the magnitude and add all the image to get a net difference image
3) Now from the net difference image we create a black and white image
4) Current image will have noise which appears as small regions to remove this we remove regions smaller than 800 pixels
5) Further there will be small moving objects around to remove this
6) 1st we add each line horizontally and if it is below a threshold it is all made zero
7) This is repeated for all the horizontal lines and then for the vertical lines as well
8) This will give an image which highlights the moving object
9) Now the area corresponding to this moving object is cropped out to get the object of interest

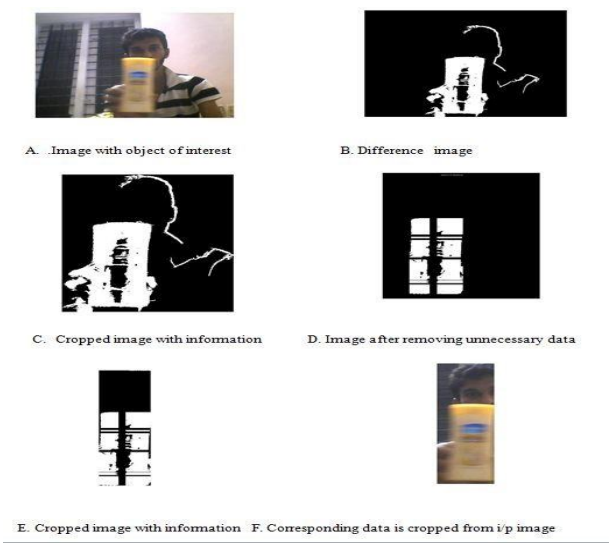

Fig. 3. ROI detection

V. TEXT REGION DETECTION

Text detection consists of two steps. The first step involves detecting text regions in a given image, while the second step retrieves text information from these regions using OCR or other technologies. We here concentrate on the first step of text region detection in natural scenes. Here we constructs a strong classifier from a combination of weak classifiers. Weak classifiers are build using different features. different features which can be used for text are gradient feature maximally stable extremal regions (MSER) feature, canny edge feature, strock width feature..etc

A. Maximally stable extremal regions

Since text characters usually have consistent color, we begin by finding regions of similar intensities in the image using the MSER feature. Maximally stable extremal regions (MSER) are used as a method of blob detection in images. Blob detection refers to mathematical methods that are aimed at detecting regions in a digital image that differ in properties, such as brightness or color, compared to areas surrounding those regions. Informally, a blob is a region of a digital image in which some properties are constant or vary within a prescribed range of values; all the points in a blob can be considered in some sense to be similar to each other

B. Edge Detector

Texts are typically placed on clear background, so it tends to produce high response to edge detection. So edge feature can be used to obtain text regions in an image. Both MSER and edge detector can be combined to get a far better result so we combine this MSER and canny edge feature to obtain text regions with better clarity. Still there can be non text regions some of this non text region can be removed by gradient feature. We combine the current image with the gradient image.

C. Region Filtering

Region properties such as area, eccentricity and solidicity can be used to remove further non text region .This is done by finding the connected components and removing the connected regions with eccentricity greater than .995 area between 50 and 3000 and solidicity less than .1.This numerical values may vary for different font image size or language. Here the value is fixed using trial and error method done on different images.

D. Strock Width Feature

Stroke width feature is a very useful feature for the extractionof text regions in an image. Characters in most languages havea similar stroke width or thickness throughout therefor strokewidth of each region in an image can be Computed and analyzed to see whether it belongs to the text category. It is therefore useful to remove regions where the stroke width exhibits too much variation. For this we compute the stroke width of all regions in an image. For each region we compute the coefficient of stroke width variation(α).

$$\alpha = std(strockwidth)/mean(strockwidth)$$

If stroke width variation is greater than a threshold we remove that region from the image .After trial and error method a

threshold of .78 is fixed here to distinguish the text region from the non-text regions.

Fig. 4. mser, canny edge ,canny and mser, strock filtering

Fig. 5. detected text regions

VI. TEXT IMAGE TO TEXT CONVERTION

This is done next to text region extraction from the detected object.
Steps

1) Initially we need to create templates for different characters .Multiple templates for a single character will increase the readability of the text
2) Templates are made of dimension 24x42 pixels

Fig. 6. Template for A

Fig. 7. Template for 8

3) To start, text regions are converted into black and white image
4) Now we find the sum of pixel values in a line, when this is zero we assume it as the separation between two lines thus we separate each line of text image
5) Now from each line we extract the regions in that image in which 1st region corresponds to 1st letter in the line 6) Now each of this region is resized to 24x42 pixels
7) This is now correlated with each template that we have created
8) Maximum value of the correlated value is computed and if it is greater than a threshold that region is identified as the corresponding letter in that template
9) This is now repeated for each regions in all the lines that we have identified

Example;

Correlation result

M in Malayalam with different templates

Corrilation between the above two gives a result of .5040 which is greater than the rest of the correlated result thus we can confirm it as the letter M

Here we use 37 template

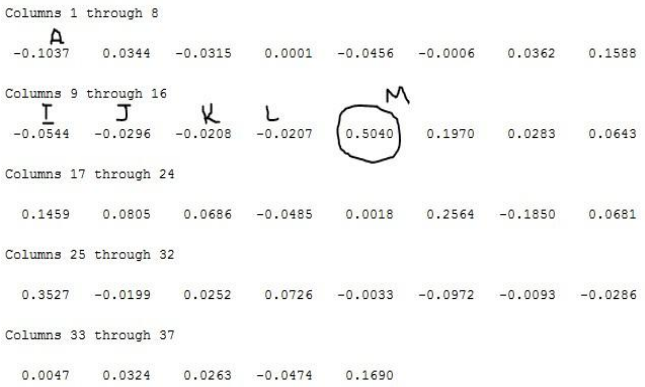

Fig. 8. Result of correlation with 37 templates

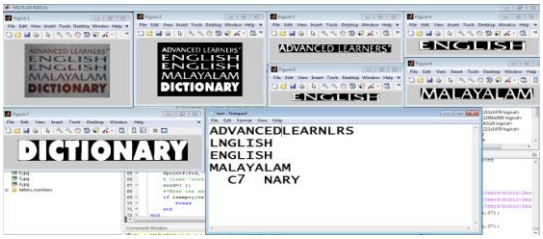

Fig. 9. Image converted to text

The output can be made much better by including more no of tamplates for each character .

VII. CONCLUSION

In this a system to read printed text on hand-held objects for assisting visually impaired persons in introduced. In order to solve the common aiming problem for blind users, a motion based method to detect the object of interest is used, while the blind user simply shakes the object for a couple of seconds. This method can effectively distinguish the object of interest from background or other objects in the camera view. To extract text regions from complex backgrounds, text localization algorithm based on models of stroke orientation and edge distributions and maximally stable extremal regions are used. The corresponding feature maps estimate the global structural feature of text at every pixel. The text region thus detected is given to an optical character recognizer which convert the text region to simple text file. This text file can be easily converted to speech with advanced off the shelf text to speech converter. The system as a whole is bulky it cannot be carried by a the visually impaired persons. But the idea implemented here can be imported to android platform like an application with some advancement

VIII. FUTURE SCOPE

If we are able to import this to the newly developed stuffs like Google glass the system will become very simple and handy because every hardware needed to implement this is already available in it like the camera, head phone, microphone etc. Along- with text reader different functions like currency recognition,colourrecognition ,bar-code reader face recognition system etc can be integrated to the current system to make it a complete virtual assistant

REFERENCES

[1] World Health Organization. (2009). 10 facts about blindness and visualimpairment [Online]. Available: www.who.int/features/factfiles/blindness/blindnessfacts/en/index.html

[2] DimitriosDakopoulos and Nikolaos G “Wearable Obstacle AvoidanceElectronic Travel Aidsfor Blind: A Survey”IEEE TRANSACTIONS ONSYSTEMS, MAN, ANDCYBERNETICS-PART C: APPLICATIONSAND REVIEWS, VOL. 40, NO. 1, JANUARY 2010

[3] Chucai Yi, Roberto W. Flores, Ricardo Chincha, and YingLiTian“Finding Objects for Assisting Blind People”

[4] MajidMirmehdi, Paul Clark, and Justin Lam “Extracting Low ResolutionText with an Active Camera for OCR”

[5] Chucai Yi and YingliTian, “Portable Camera-Based Assistive Textand Product Label Reading From Hand-Held Objects for Blind Persons”IEEE/ASME TRANSACTIONS ON MECHATRONICS, 2013

[6] C. Yi and Y. Tian, “Assistive text reading from complex background forblind persons”in Proc. Int. Workshop Camera-Based Document Anal.Recognit, 2011, vol. LNCS-7139, pp. 15-28.

[7] Michael R. Lyu,“A Comprehensive Method for ultilingualVideoTextDetection, Localization, and Extraction”IEEERANSACTIONS ONCIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, VOL. 15,NO. 2, FEBRUARY 2005

[8] A. Dutta, U. Pal, P. Shivakumara “Gradient based approach for TextDetection in Video Frames”