International Journal of Research in Advent Technology (E-ISSN: 2321-9637) Special Issue
National Conference "NCPCI-2016", 19 March 2016
Available online at www.ijrat.org

Shopping Assistant Framework

Ravindra Patare[1], Umesh Jaiswal[2], Hitendra Chouhan[3], Shubham Pagare[4,] Shweta.A.Tiwaskar(Asst.Prof.)
Computer Department[1, 2, 3, 4],Vishwakarma Institute Of Information[1, 2, 3, 4]
Email:patare.ravindra@gmail.com[1],umesh.jaiswal20@gmail.com[2] ,
chouhan.hitendra9@gmail.com[3],shubhamp.pagare@gmail.com [4]

Abstract- Today, India is home to the world's largest number of blind people. Over 15 million blind people out of 37 million are from India. In India there are about 287 million illiterate people which is largest in the world. This is 37 per cent of the global total. Reading is obviously essential in today's society. Printed text is everywhere in the form of reports, receipts, restaurant menus, product packages, instructions on bottles and medicines, etc. And while devices such as optical aids, video-magnifiers and screen readers can help elderly or blind users and those with low vision to read product details, there are few devices that can provide good access to common hand-held objects. In this paper we present the design of a system to be used in supermarkets, conceived primarily for helping people, probably with mobility limitations, such as physically impaired or elderly, while shopping.

Index Terms— Android Device; Image Processing; Voice Output; Computer Vision.

1. I NTRODUCTION

Today, in India there is world's largest number of blind people. Over 15 million blind people out of 37 million are from India. In India there are about 287 million illiterate people which are largest in the world. This is 37 per cent of the global total. Reading the product details essential while shopping in today's society. Printed text is everywhere in the form of reports, receipts, restaurant menus, product packages, instructions on bottles and medicines, etc. Shopping is a part of day to day life. So there is a need to design and develop a system which will assist visually impaired and elderly people in shopping.

In this paper we present the design of a prototype system to be used in supermarkets, primarily for helping people, possibly with mobility limitations, such as visually impaired or elderly, during shopping. As an elementary overview of the operation of our system we can say that the user uses a mobile device. Then, upon arriving to the supermarket, the system will help users to identify the supermarkets shopping infrastructure. Then, user moves around in supermarket capturing the product pictures with the portable device.

Image processing is the mathematical treatment of an image and manipulation of the image according to the requirements of the user. Image processing refers to Digital image processing but optical and analog image processing is also possible. Image processing is closely related to computer graphics and computer vision. In computer graphics images are manually made from physical world object without acquiring them from some device like camera. Computer vision on the other hand is high level image processing and in this an image is acquired using a camera and various techniques are then applied to manipulate that image.

In our system we will be using the computer vision part of image processing. We have a camera that will acquire the image from a product and then using image processing algorithms modifications like background subtraction, background illumination would then be applied to create a final image to extract text.

The shopping assistant consists of a portable system with camera and speaker. The system can be used by people with physical limitations or those who are illiterate. The user will capture a product image using camera and the image will be processed to extract text. The extracted text will be then provided as a audio response for the user to understand the product. Considering blind people the system will have a wide angle capture of image.[2]

Computer vision is a field that includes methods for gathering, processing, analyzing and understanding an image basically manipulating a high-dimensional data from the real world and generating useful information from that. In more technical terms computer vision aims at creating artificial systems that extract valuable information from images. The image data could be in any format as it can be captured from multiple cameras, video sequence or a multi-dimensional data.

Considering today's shopping scenario it can be seen that there is increase in supermarkets and Malls. It requires a person to search and browse through the products to make purchases. People with some sort of disability like blindness or low vision or dyslexia avoid visiting such markets since it requires accomplice or the entire process takes a very long time. The Shopping assistant will help such people to browse these products and buy them.

2. L ITERATURE SURVEY

Different assistive system for Visually Impaired people based on Barcode Reader , RFID and computer vision were studied and pros and cons of each system is given below.

A. Barcode Reader

A barcode reader (or barcode scanner) is an electronic device which can read and output printed barcodes to a computer. Like a flat bed scanner, it consists of a source of light, a lens and a light sensor which translates optical impulses into electrical ones. Additionally, nearly all barcode readers contain decoder circuit which analyzes the barcode's image data provided by the sensor and sends the barcode's content to the scanner's output port. Trinetra aim to be cost-effective and assistive technologies to provide blind people a greater degree of independence in their daily activities. The overall objective is to improve the quality of life for collective capability of diverse networked embedded devices to support grocery shopping, transport, etc. Trinetra system, a barcode-based solution having components, such as an Internet and Bluetooth-enabled cell phone, text-to-speech software and a portable barcode reader. [6]

Advantages:

- Scan barcode with handheld device.
- Scan with device fitted on each trolley.

Drawbacks:

- Have to locate barcode.
- Requires specific devices.

B. RFID reader

Radio-frequency identification (RFID) is use of electromagnetic fields to transfer data in wireless manner . The tags contain electronically stored information to automatically identify and track the tags which are attached to objects. Some tags are powered by electromagnetic induction. Some types collect energy from the interrogating radio waves.
Other types have a battery as a local power source and may operate at hundreds of meters from the reader

Grozi solution makes use of new computer vision techniques for visual recognition of specific products inside of a store as specified in advance in a shopping list. There are techniques of complementary resources such as RFID, barcode scanning, and sighted guides.[4]

Advantages:

- RFID tags are fitted on products.
- RFID reader fitted on trolley.[3]

Drawbacks:

- Requires RFID tags for each product
- Cost inefficient on large scale.

C. Computer Vision

Computer vision is a field which includes methods for gathering , processing, analyzing, and understanding images and high-dimensional data from the real world so as to produce numerical or symbolic information, e.g., in the forms of decisions. A theme in the development of this field has been to duplicate the abilities of human vision by perceiving electronically and understanding an image. Understanding in this context means the transformation of visual images into descriptions that can interface with other thought processes and evoke appropriate action.

Calit2 researchers are developing GroZi system, a device which combines a mobile vision system and ZigZag, where complete, the GroZi system will allow a user to navigate the supermarket, find a specific aisle, read aisle labelsand then check key words will undergo spellcheck against grocery categories. The box then scans the aisle for objects that look like specific products on the shoppers list, which is compiled online and downloaded onto the handheld device before going to the store. The technology which involves sign reading, object recognition and text- to speech notification. One component of GroZi now under development is called Texty. It's a real-time text detection and optical character recognition system. It uses text-to-speech to give voice output of the contents of a sign it sees. The system will be adapted and optimized for being embedded inside the device, but has already proved successful in tests.[4]Since edge detection is in the forefront of image processing for object detection, it is necessary to have a good understanding of edge detection algorithms. [17] Two sets of images resulting from the application of those algorithms are then presented. Is shown that under noisy conditions, ISEF, Canny, MarrHildreth, Kirsch, Sobel, Lapla2, Lapla1 exhibit better performance, respectively. The Classification in this paper is based on the behavioural study of these edges with respect to the following differentiation operators:- Gradient edge Detectors , Zero crossing, 11 Laplacian of Gaussian (LoG) Gaussian edge detectors Colored edge detectors From every category in the classification, at least one algorithm has been chosen for implementation.[17]

Video text provides important and meaningful information about video con- tent which is helpful for many applications like automatic annotation of video and images, indexing etc The problem of automatic extraction of text from the video is a difficult task because of large variety in text style,

font, size, line orientation and because of complex background also, Text extraction problem has been addressed by various methods. The aim of this paper is to propose a framework for extracting text from a video using Java platform. This framework performs preprocessing, detection and recognition. Preprocessing involves converting frames into gray scale, blurring and binarization. Detection includes localization in image and segmentation, and recognition is done by template matching. This framework can be applied for multiple languages.[5] This paper introduces the rule-based layout algorithm for analysis and learning based text classifier for localization of text. This feature defines maps based on stroke orientations and edge distributions. Representative and discriminative text features are generated to distinguish text characters. This paper uses a Region of Interest (ROI) to detect the object on its motion.

Automatic text recognition from natural images receives a increase in attention because of potential in applications in image retrieval, intelligent and robotics transport system. Camera- based document analysis becomes a good solution with the increasing resolution and availability of digital cameras. This paper uses four text extraction methods which are dependent on the varying sizes of text in natural scenes.[12]

Drawbacks:

- Can't cope with rapid change in lighting conditions.
- Image captured rapidly can disrupt progress.

3. PROPOSED SYSTEM

Our idea is to design new android shopping Assistant app which will be helpful for physically impaired and elderly people .In the App object's image will be captured through camera holding the object in hand or just focusing the object of interest. Here the captured image will undergo various image processing steps. Text will be extracted through the captured image Finally the extracted text will be converted to voice output using mobile hardware .

3. SYSTEM ARCHITECTURE

The flow of our entire system is illustrated in figure1.User will take the product in front of camera. Camera will capture image on that product. Then it localize the text region. Then text region cropped by image and text code recognized by OCR

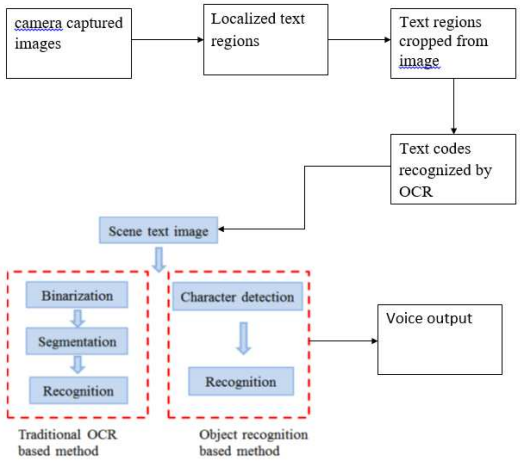

Fig1 System Architecture

.Optical character recognition (OCR): As you read these words on your computer screen, without you even noticing your eyes and brain are carrying out Optical character recognition. Your eyes are recognizing the patterns that light and dark that make character (letters, numbers, and things like punctuation marks) printed on the screen and your brain is using those to figure out what I'm trying to say (sometimes by reading individual characters but mostly by scanning the entire words and whole groups of words at once). Computers can do this too, but it's really hard work for them. The first problem is computer has no eyes, so if you want it to read something like the page of an old book, you have to represent it with an image of that page generated with an optical scanner or a digital camera. The page you create is a graphic file (often in the form of a JPG) and, as far as a computer's concerned, there's no difference between it and a photo of the Taj Mahal or any other graphic: it's a completely meaningless pattern of pixels (the colored dots that make up any computer graphic image). In other words, the computer has a picture of the page and not the text itself it can't read the words on the page like we can, just like that. OCR is the process of turning a picture of text into text itself that is, producing something like a TXT or DOC file from a scanned JPG of a printed or handwritten page.

OCR sense the text image and recognizing in two different ways that is:

A. Traditional OCR based method

In this method text image will binarize. Binarization means to select the text image from product. Then from text image select the actual text is known as recognition.

B. Object Recognition based method

In this method we will used a character detection. In this step it will detect the text character by character and then recognize actual text. Voice output: And finally it will give the voice output to the user

Stimulus or Response Sequences:

A. Image Capturing and Processing:

- Description and Priority: Image Processing is the key feature that will run this system. This system feature will be the highest prioritized feature and will provide the basic input to the system.[9][14][15]
- Stimulus or Response Sequences:

The customer will now have to hold the device and and product in front of camera. Here, the high resolution camera on the mobile device will be facing the product and will detect the edges of the text. In order to ensure that the edge detection takes place depending on the size. This will be dependent and will vary according to the size of each product, which in turn will be calculated by the edge detection algorithm. This information will also be relayed in the Android Application.

B. Image Capturing:[5][14]

- Description and Priority:

In order to generate output system requires a video or image feed to process. This feed is provided by the camera present in the system. The processing mainly depends on the quality of the image captured. It is not necessary to have very high resolution image but a very low quality of image will also be difficult to process. So an optimum quality of image is required. Since the process will be to capture natural scenes and extract scenic text, the methods are dependent on character size. The risk lies in this feature where if the image capture fails the system might fail to generate correct output . Real-time Operation: It should operate in real-time so it should be quick. Delay in processing output may render system useless to the customer.

- Text Extracting [15]:

Description and Priority: The extraction of text from Labels and logo must be robust. As the performance of the system is dependent on this feature it is of high priority. The extraction of text will define the overall performance of the product. Text extraction can be divided in following

Sub-problems:

a) Detection, b) Localization, c) Tracking, e) Extraction and enhancement, and f) Recognition (OCR) Real-time Operation: Real-time operation requires instant response and extraction of text will play a vital role to improve the response and delay of the product Text localization determines the location of text in frame. Text tracking plays a vital role to reduce the processing time for text localization and maintains integrity across various frames. Functional Requirements: Real-time text extraction is difficult since the natural scenes consist of many complex textures and background. Distorted Image: The scenic images are not easy to process for their complex nature and a distorted image will add to the difficulty in text extraction.

Voice Output:

Description and Priority: The product is designed with the perspective of elderly or physically disabled people. So the voice output is quite necessary. For instance a blind person might not be able to see the text on the products so the voice output will provide a solution to the problem of interaction in this situation.

5. CONCLUSION AND FUTURE SCOPE

This System will provide the people with visual or some sort of limitations that obstruct from understanding the products and their labels easily and conveniently shop with other people. Creating such a system which will be easily available to any user and will be used in common shopping places.

In future we can overcome the present challenges of the system and can have full -proof system which will capture images and can process images captured on its own on its own processor which will be wearable.

7. REFERENCES

[1] Xiaodong Yang, Student Member, IEEE, Shuai Yuan, and YingLi Tian, Senior Member,"Assistive Clothing Pattern Recognition for Visually Impaired People.", IEEE, VOL. 44, NO. 2, APRIL 2014

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

[3] Galande Jayshree, Rutuja Gholap, Preeti Yadav, IJETAE, "RFID Based Automatic Billing Trolley.", ISSN 2250-2459, ISO 9001:2008 Certified Journal, Volume 4, Issue 3, March 2014 P.R.E.C. Loni, Ahmednagar

[4] Calit2 at UCSD/TechBrief, "Helping the visually impaired SHOP."

[5] CiiT International Journal of Digital Image Processing, "Development of Framework for Text Extraction from Videos.", Vol 6, No 07, August 2014

[6] Patrick E. Lanigan, Aaron M. Paulos, Andrew W. Williams, Dan Rossi and Priya Narasimhan Carnegie Mellon University, Pittsburgh, PA 15213, USA,"Trinetra: Assistive Technologies for Grocery Shopping for the Blind.".

[7] Adla.Vamshika, P.V.Vara Prasad Rao , "CAMERA-BASED ASSISTIVE SYSTEM FOR LABEL DETECTION WITH VOICE OUTPUT FOR BLIND PERSONS.", Volume 4, Issue 5 AUG 2015

[8] Richard J.M. den Hollander and Alan Hanjalic, "LOGO RECOGNITION

IN VIDEO STILLS BY STRING MATCHING", 2003 IEEE.
[9] Cun-Zhao Shi, Student Member, IEEE, Chun-HengWang, Member, IEEE, Bai-Hua Xiao, Member, IEEE, and Song Gao, Jin-Long Hu, "Scene Text Recognition Using Structure-Guided Character Detection and Linguistic Knowledge.", VOL. 24, NO. 7, JULY 2014
[10] Antonio Marin-Hernandez, Guillermo de Jesus Hoyos-Rivera, Marlon Garca-Arroyo, Luis Felipe Marin-Urias Department of Artificial Intelligence, Universidad Veracruzana, Xalapa, Mexico , "Conception and Implementation of a Supermarket Shopping Assistant System.", 2012 IEEE.
[11] Rajkumar N, Anand M.G, Barathiraja N, "Portable Camera-Based Product Label Reading For Blind People", (IJETT) Volume 10 Number 11-Apr 2014
[12] Nobuo Ezaki, Marius Bulacu, Lambert Schomaker Information and Control Engineering Department Toba National College of Maritime Technology, Japan
[13] Boris Epshtein, Eyal Ofek, YonatanWexler Microsoft Corporation," Detecting Text in Natural Scenes with Stroke Width Transform"
[14] C.P. Sumathi, T. Santhanam and G.Gayathri Devi."A SURVEY ON VARIOUS APPROACHES OF TEXT EXTRACTION IN IMAGES", (IJC-SES) Vol.3, No.4, August 2012.
[15] Weilin Huang, , , Zhe Lin, Jianchao Yang, and Jue Wang,"Text Localization in Natural Images using Stroke Feature Transform and Text Covariance Descriptors",(iccv)2013
[16] Debosmit Ray," Edge Detection in Digital Image Processing",Thursday, June 06, 2013.