International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Index Copernicus Value (2013): 6.14 | Impact Factor (2013): 4.438

Non Homogenous Assistive Text Reading From Objects For Blind Persons

Kumar P[1], Sivaraman P R[2]

[1]PG Scholar, Department of Electrical And Electronics Engg., Rajalakshmi Engineering College.

[2]Assistant Professor, Department of Electrical And Electronics Engg., Rajalakshmi Engineering College

Abstract: ***Text information provides important clues for many image based application such as assistive navigation. However locating and retrieving non homogeneous text is challenging task. The goal is to provide a framework for non homogeneous assistive text reading from objects for blind person based on text orientation. Non horizontal text is recognized using key-points generated by Scale Invariant Feature Transform and individual text characters are recognized based on line detection on text characters along with Optical Character Recognition. Development of the framework is performed using the Image Processing Toolbox of MATLAB. The extracted text characters are given in speech through audio for blind persons.***

Keywords: assistive text,scalar invariant feature transform, optical character recognization

1. Introduction

Assistive technology means a product or piece of equipment which is developed or modified in a manner which helps the physically challenged people to gather more details as that of normal peoples. In today's world there are numerous assistive technology devices of which vibrating watch and talker device are common. Pictures with text act as important communication medium for expressing information. Digital image processing makes computational capabilities on image which makes a platform for performing image processing.

One of the basic things to consider for assistive text recognition is region of interest and character recognition. Two basic methods for obtaining region of interest is rule based method and learning based methodologies. In rule based methodology, certain threshold values are chosen as factor of evaluation. Opposite to this is learning methods where neural networks are trained for gathering the information about picture.

Printed text information is widely spread in today's world for example in consumer product labels, bank forms, receipts etc. There are certain devices like magnifying glasses, certain forms of optical aids which help blind users for obtaining text information. Today's existing environment which helps blind user in product text reading is bar code reader with little difficulty in positioning the barcode. When it comes for text extraction it is not true that the blind users will hold the exact manner of product for recognition.

There is possibility for holding the product box in upside down or with some orientation of box. Hence there is needed for a method to detect text labels even with some orientation.

In assistive reading especially for blind users it is difficult to various image resolution factors. However the classifier fails to classify very small text and low intensity images [4].

The image intensity as a factor for retrieving text from predict that they will hold the product which is suitable for text extraction or simply region of interest. Hence there is need to adapt a hybrid method which provides both region of interest and also the able to detect the oriented product label too.

2. Related Works

Navbelt, an electronic travel aid is employed for their work. This Navbelt comprises of a belt of ultrasonic sensors, a data processing element and auditory system. The work of multiple ultrasonic sensors placed around the belt is to detect the obstacle through the principle of sonar's. To reduce the acoustic noise, Error Eliminating Rapid Ultrasonic Firing (EERUF) algorithm is employed along with low pass filter. They lack their performance when uneven walking patterns are performed by user and unable to detect overhanging objects [8].

The method of employing Gaussian distribution as a key factor in segmenting image sequence captured using camera is proposed. The pixels are modeled as mixture of Gaussian functions. Based on persistence and variance of Gaussian mixture the interested region is obtained. This method found to be less efficient if background contains overlapped objects. Adding prediction to each Gaussian (prediction by employing Kalman filters) approach may lead to more robust tracking of images in movable camera [9].

An approach of employing Support Vector Machines (SVM) and CAMSHIFT (Continuously Adaptive Mean Shift Algorithm) for text detection is handled in this paper. The SVM algorithm leads to a linear classifier of regions. The CAMSHIFT ultimately helps in determining the search window of candidate region. The major approach of this method is it does not need any external feature extraction module .It also gives fast text detection irrespective of

complex backgrounds is considered. Here the images are applied with Gabor transform to obtain local features and later to Linear Discriminant Analysis (LDA) for feature selection. They suggest that distortions are due to non frontal

viewing angle of camera. It is overcome by applying affine rectification hence the detection rate is increased [1].

The method of employing mixture of Gaussian function and multiple cues to detect the background of fixed camera is analyzed here. Here the background is modeled by three Gaussian mixtures. Based on color intensity and texture information the shadow region is removed. The gradient value of Gaussian mixture plays a role in determining the background feature for extraction. Learning rate determines the rate of period for which the object should be static such that the analyses are carried since the camera is fixed position. If large homogenous objects with less text are analyzed, produces patch holes in background which reduces the precision of extracting text feature [10].

An approach for text detection in videos is located based on Laplacian approach in frequency domain. Here the text is subjected to Fourier Laplacian for filtering then k- mean clustering to obtain candidate text region. The performances are evaluated based on detection rate and false positive rate. The proposed method consumes longer time for detection because of segmentation of complex component classification to simple component classification [6].

The idea of text detection and text caption in video frames based on corners is analyzed. In order to obtain the text from video they utilized the properties of corner of a text as key feature. Initially the captured video frame containing text is subjected to Harris Corner Detector. This detector performs autocorrelation on image signal to obtain corner points of text. After extracting corner points then comes the shape description. Features like area, saturation, orientation, aspect ratio and position are utilized for text shape extraction. Secondly Lucas-Kanada algorithm (Optimal Flow estimation) is used for detecting text from moving images [12].

To obtain the desired text from complex background they employed an text detector (Wald boost classifier).the output of this classifier is given to boosted classifier calibration method to detect the probabilities of text position and it also provides scale information of text. To perform segmentation they used Niblack's local binarization algorithm. To filter non text region conditional random field model is used. Unary component properties and binary context of images are key factors for eliminating these non-text backgrounds. Later the text lines are grouped by learning based energy minimization method. This method performs quick recognition of images [5].

The method of Scalar Invariant Feature Transform as a key feature image registration is considered here. SIFT matching and global optimization on the whole matched key points based on iterative reweighted least squares provides the method of image registration. Here they employ block level SIFT operation on larger images. If the image is too large then it is divided into smaller blocks which help them to apply SIFT operations instead of performing as single which causes data to be loosed [2].

Ground truth generation involves text line segmentation, word segmentation, bounding box drawing, scene text separation and script identification with some manual tuning if required. Line and word segmentation is carried out by scanning the image from left to right. It is followed by placing bounding box which encloses the text. If boundary box doesn't fix with text then manual adjustment can be done. Hence this method provides semiautomatic design for obtaining text from video frames [7].

3. Frame Work

The system architecture mainly involves three basic component .One is to acquire image from which text has to be retrieved. Secondly and the most important part is process of retrieving the text. And finally the processed text has to be given in voice for blind persons.

A. Camera Unit

The camera captures the image to be processed and it act as entry point for whole process. Hence it is necessary to select a camera which provides high resolution apart from normal scenario because here assistive text has to be recognized. Considering the other metrics of camera selection here Microsoft camera is employed.

B. Processing Unit

It is necessary to carry the entire setup, so portability plays an important role in selection of processor. In this project raspberry pi b plus module is employed it is also known as minicomputer. The processing unit has to undergo several things of which non homogenous text recognition and individual text character extraction are interesting one. The processing unit is furnished with various image processing features such as camera connector, ARM 11 based Broadcom SOC and certain other peripherals of audio jag makes it a suitable processing unit for text recognition for proposed system. The overall architecture of proposed system is given in figure 1 which shows three major sections i.e., camera acquisition, processing unit and audio unit.

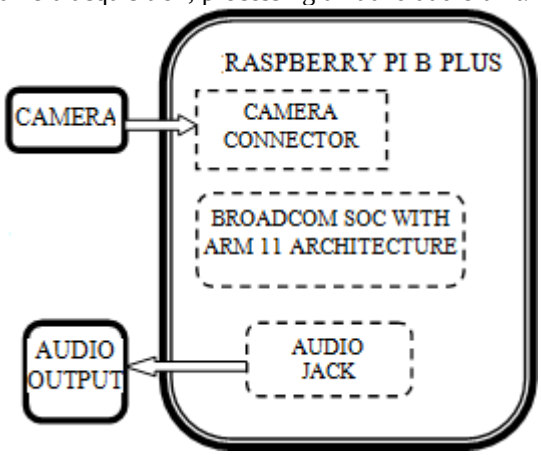

Figure 1: Overall Block Diagram of Proposed Assistive Text Reading

4. Methodologies

This section gives an idea of methods involved in oriented assistive text extraction and recognition methods .Initially the region of interest gives the method of text region extraction followed by individual text recognition is given.

A. Camera acquisition

The product label to be read is captured through camera. Since image captured is associated with numerous backgrounds elimination of background noise is important. One of basic produce to reduce the background noise or in simple to acquire the exact product to be read, the users are asked to shake the product.

Background subtraction [6][9] is methodology for removing background noise in which the current frame is subtracted with last frame hence the residue of interested region remains. This region is processed for text extraction the followed by text recognition. This methodology is also widely known as foreground analysis.

B. Region of Interest

The region of interest specifies the text region to be extracted. Here scalar invariant feature transform [2] is employed for obtaining the ROI. In addition to that this transform provides a way to extract the oriented image also. This transform involves four major phases.

The first phase involves the operation of scale space extrema detection; this phase is interested in collection of scales and image locations of captured image. Potential interested points are gathered using difference of guassian function. Literally it is represented by,

$$L(x, y, \sigma) = G(x, y, \sigma) * I(x, y) \quad \text{----- (1)}$$

Where,

$L(x, y, \sigma)$	Scale Space of an image
$G(x, y, \sigma)$	Variable scale guassian
$I(x, y)$	Input image in x and y coordinates.
$*$	Convolution operator

The next phase is key point localization, from above phase too many candidates are generated hence in order to compute the candidate list is to be reduced. Based on the location, scale and low contrast certain candidates are rejected. Two important factors considered in elimination of candidates are principle of curvature and low extrema detection. The principle of curvature is determined by hessian matrix and it is presented as H[X],

$$H[X] = \begin{bmatrix} D_{XX} & D_{XY} \\ D_{YX} & D_{YY} \end{bmatrix} \quad \text{------ (2)}$$

The third phase of operation involves orientation and angular assignment .this phase is quite helpful in attaining the key points even if the picture is held in orientated position.

Where its magnitude is represented by m(x,y) and angular orientation is given as Θ(x,y),

$$\Theta(x,y) = \sqrt{\tan^{-1}\left[\frac{(L(x,y+1) - L(x,y-1))}{(L(x+1,y) - L(x-1,y))}\right]} \quad \text{------ (3)}$$

The operations performed before have assigned an image location, scale, and orientation to each key- point. These parameters impose a repeatable local 2D coordinate system in which to describe the local image region, and therefore provide invariance to these parameters. The ultimate step is obtaining key point descriptor. The figure 2 diagram describes the 2x2 array of key point descriptor. This method widely reduces the noise also.

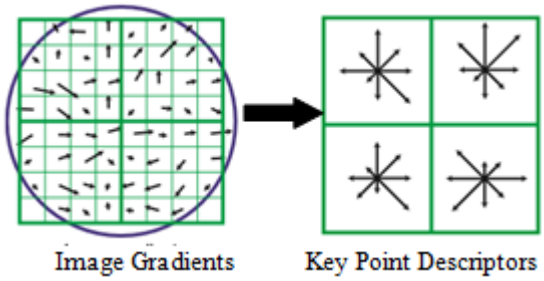

Figure 2: Key Point Descriptor

C. Optical Character Recognition

Optical character recognition is one of the basic forms in extraction of individual text characters [5]. The foremost step is to pre process the image and store that image in computational format i.e., in matrix format. Later the text characters are obtained by employing connected component analysis. The overall flow of operation is given by

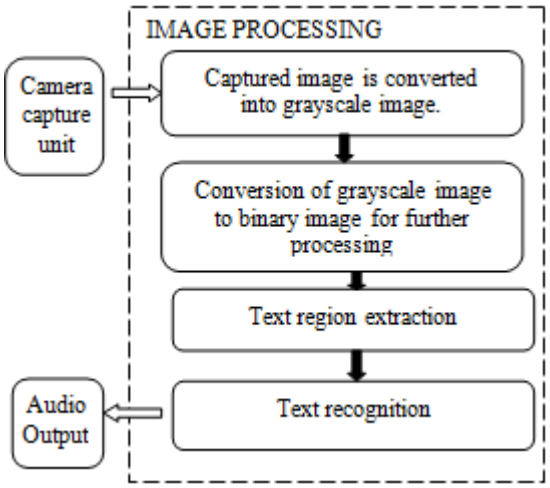

Figure 3: Flow of Operations

Connected component analysis is a procedure in which the text file containing text characters are examined to determine whether there is any overlapping regions. In that case it marks the letter as single letter. Hence this method of extraction provides a way to extract the individual letter of characters.

5. Simulation Results

The various operations on text extraction from image and their images are shown as follows. Initially the image are exposed to some pre-processing works later it was followed by SIFT feature key point extraction and matching of key points to region of interest. Finally optical character recognition is performed text characters. These text characters are given to text to speech function to retrieve the text characters. The image to be processed is initially captured and it is shown in figure 4

Figure 4: Input Images

Know the primary work is to extract the key points generated from scalar invariant feature transform is given in figure 5 and the matched key points are displayed in command window as in figure 6.

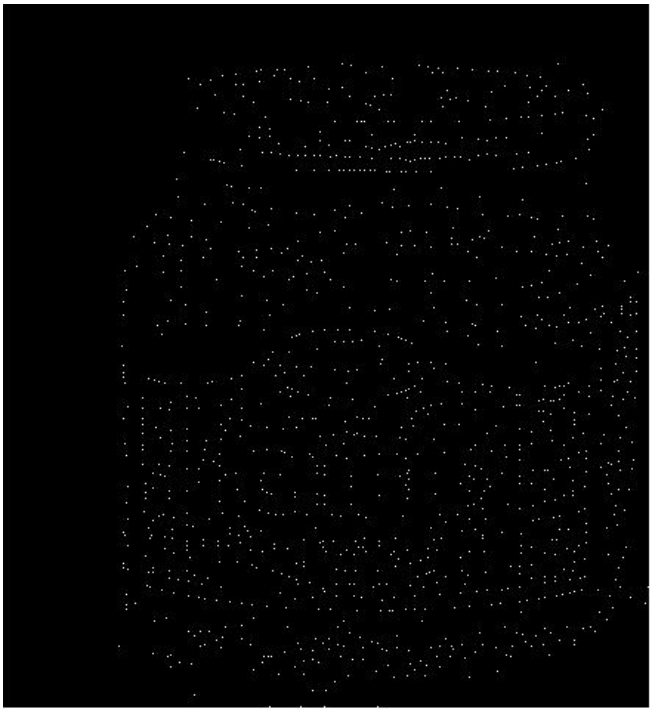

Figure 5: Key Points from SIFT

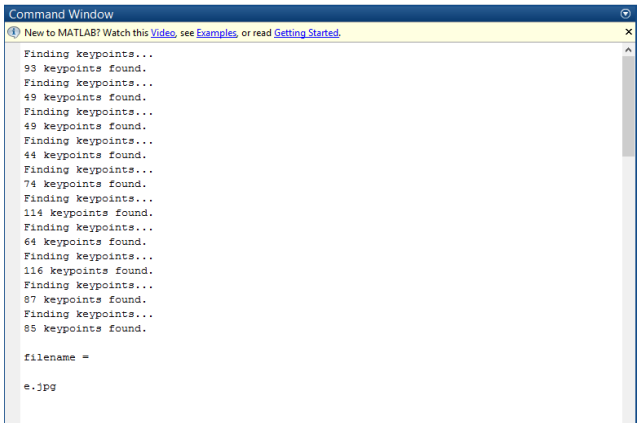

Figure 6: Command Window Expressing Matching Points

After the process of locating the region of interest as shown in figure 7, know it is time to apply optical character recognition to obtain the individual text characters. The individual characters extracted using optical character recognition is given in figure 8.

Figure 7: Extraction of Region of Interest.

Once the region of interest is obtained then inorder to perform certain mathematical calculations the extracted text region must be converted in to grayscale followed by binary scaled image this process of conversion in general termed as pre processing.hence before applying OCR ,the preprocessing work is done.

The recognized text codes are recorded in script files. Then, we employ the Microsoft Speech Software Development Kit to load these file and display the audio output of text information. Blind users can adjust speech rate, volume, and tone according to their preferences.

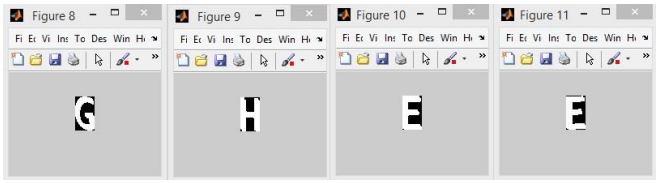

Figure 8: Extracted characters

6. Hardware Results

The web camera is connected to one of the USB ports of raspberry pi. To obtain the voice the data card is connected to another port of board. The raspberry pi board is a standalone system such that it fulfills the portability constraint. This raspberry pi board can also operated using remote control which is additional feature. During the execution the raspberry pi is remotely viewed through the Ethernet cable. To achieve this Ethernet cable is connect to remote monitor at one end and other end is connected to raspberry pi board is depicted in Fig 9.

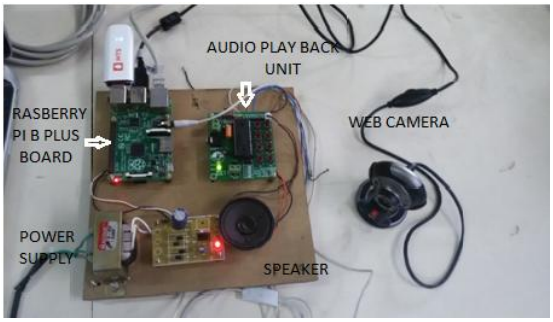

Figure 9: Hardware Setup

After removing the noise through background subtraction technique, the individual characters are obtained by applying optical character recognition method this method extracts individual text. Later the voice process is achieved through espeak with help of internet. For reference the offline voice is obtained through audio play back unit. The voice is heard through headphone. The various products taken in to consideration is given in Table 1.

Table 1: Sample of Inputs Taken And Their Results

Input taken (product)	*Recognized Text*	*No. Of Mistakenly Read Text*
CAPSOYL	CAPSOYL	NONE
ABEXIN	ABEXIN	NONE
FLUR0D0NE	FLUR0D0NE	2 -o
CINTHOL	C0NTHOL	1-i
CLEAR	CLEAR	NONE
GHEE	GHEE	NONE
PEANUTBUTTER	PEANUTBUTTER	NONE
CUPCAKES	CUPCAKES	NONE
TRACT0R	TRACT0R	1-o
PEPSl	PEPSl	1-i
Rexona	kdona	3-Rex
complan	compe0n	2 -la
Fortune	Fortune	NONE
RedBull	RedBull	NONE

7. Conclusions

Non homogenous texts are found in most of places which are found to be difficult in extraction as they are filled with most of inter connected components. Here Scale Invariant Feature Transform is proposed for retrieval of non homogenous texts. The SIFT operations extracts the key-points which was widely evident characteristic in matching region of interest. The individual characters are recognized using Optical Character Recognition. In addition to that the recognized text is provided in voice for easy access for blind persons. The algorithm is generated in the MATLAB environment and the results of simulation are displayed as images.

References

[1] Chen X., Yang J., Zhang J. and Waibel A., Automatic detection and recognition of signs from natural scenes," IEEE Trans. Image Process., vol. 13, no. 1, pp. 87–99 in 2004.
[2] Huo C., Pan C., Huo L., and Zhou Z. "Multilevel SIFT Matching for Large Size VHR Image Registration" in IEEE Geoscience And Remote Sensing Letters, VOL. 9.no 2.pp . 171-175. March 2012.
[3] Kim K.,Jung K.,andKim J."Texture- based approach for text detection in images using support vector machines and continuously adaptive mean shift algorithm" IEEETrans.Pattern Anal.Intell.,vol.25,no.12,pp. 1631–1639.Dec 2003.
[4] Kumar S., Gupta R., Khanna N., Chaudhury S., and Joshi S. D. "Text extraction and document image segmentation using matched wavelets and MRF model" IEEE Trans Image Process., vol. 16, no. 8, pp.2117–2128. Aug 2007.
[5] Pan Y., Hou X., and Liu C., "A Hybrid Approach to Detect and localize Texts in Natural Scene Images" IEEE Trans.,On image processing, VOL. 20, NO. 3. pp. 800-813. March 2011.
[6] Phan T., Shivakumara P., and Tan C. L. "A Laplacian method for video text detection" in Proc. Int. Conf. Document Anal. Recognit.,pp.66–70. 2009.
[7] Phan T., Shivakumara P., Bhowmick S., Li S., Tan C. And pal U. "Semi-Automatic Ground Truth Generation for Truth Detection and Recognition in Video Images" IEEE Trans., on circuits and system for video technology. 2013.
[8] Shoval S., Borenstein J., and Koren Y., "Auditory guidance with the Navbelt: A computerized travel for the blind" IEEE Trans. Syst., Man, Cybern. C. Appl. Rev., vol. 28, no. 3, pp. 459–467.Aug 1998.
[9] Stauffer C. And Grimson W. E. L., "Adaptive background mixture models for real-time tracking", IEEEComput. Soc. ConferenComputer Vision Pattern Recognition.1999.
[10] Tian Y., Lu M., and Hampapur A., "Robust and efficient foreground analysis for real-time video surveillance" IEEE Comput. Soc. Conf. Computer Vision Pattern Recognition. 2005.
[11] Yi C. and Tian Y., "Text string detection from natural scenes by structure based partition and grouping" IEEE Trans. Image Process, vol.20 no.9,pp.2594-2605. Sep 2011.
[12] C.Yi and Y.Tian, "Text detection in natural scene images by stroke gabor words" in Proc. Int. Conf. Document Anal. Recognit. , pp.177–181. 2011.
[13] Zhao X., Lin K., Yun Fu, Yuxiao Hu, Yuncai Liu and Huang Thomas "Text From Corners: A Novel Approach to Detect Text andCaption in Videos" IEEE Trans., on image processing , VOL. 20, NO. 3. 2011.