International Conference on Science, Technology, Engineering & Management
[ICON-STEM'15]
Journal of Chemical and Pharmaceutical Sciences ISSN: 0974-2115

Handheld Object Identification and Text Reading For Visually Challenged

S.Pushpalata*, R.InfantAbinaya, E.Esakkiammal
Department of Information Technology, Jeppiaar Engineering College, Chennai, India.
***Corresponding author: Email: latha.spl2004@gmail.com**

Abstract

Search engines are very useful to get the information about things. But, in this faster world, typing a text to search query is consuming more time. Also, small type mistakes will show error information. Fast, accurate visual search is the future, and we've got to admit the future looks bright. In this project work, we propose an effective alternate solution, Cam Find. Cam Find provides fast, accurate results with no typing necessary. All we need is to snap a picture. This technology completely eliminates the hassle of typing in a search query. This project work uses Raspberry Pi processor as an implementation platform.

INTRODUCTION

There are over 314 million visually challenged people worldwide , a major part of this population are still blind even in developed countries like united states, the national health interview survey conducted in 2008 reported that over 85 of the adult Americans lack the ability to see. In recent time's development in computer vision, digital cameras and portable computers help to aid these individuals by developing camera based products that integrate computer vision technology with already existing products such as optical character recognition (OCR). Reading is one of the basic necessity today .everything around us are in the form of reports, receipts, bank statements, product packages, restaurant menus etc...Contain printed text on it although optical aids and video magnifiers and screen readers help blind users and those with lower vision help to facilitate text reading. There are few devices that can render better access to common hand-held objects much as product packages and objects printed with text. Formulating devices which are even more portable and sophisticated can promote independent living and foster economic and social self-dependency. There are already some portable systems in use that cannot handle product labeling for example, bar code readers help identify various products in the extensive product database to enable blind users to access information about products through Braille and speech . But there was difficulty in finding him possession of the barcode so pen scanners might be employed in the cases. These system combine OCR software for the propose of scanning and text recognition and integrated voice output Reader mobile runs on cell phone and reads document which are really flat with the dark surface and mostly contain text. Accurately read black print on a white background but cannot read text with complex background and text on uneven surface.

There are no existing reading assistance that can read from all kinds of challenging background and patterns .Our proposed algorithm that effectively handle likely all the drawbacks of existing models and extract text from hand-held objects and nearby sources which ever captured through the camera. The most challenging part in assistive reading system for blind people in positioning of object of interest within the camera view. Often text from the surrounding areas also included. Thus to extract the hand-held objects from the image we proposed motion based method to isolate the region of interest and text recognition is done only on the area of interest. For text orientation the paper includes text strings in screen images keep approximately horizontal alignment. There are over one million people in this world who are visually challenged. A number of helping aids have been already developed to help them led an independent life as normal human beings. The helping aids come under two categories namely for navigational purpose and for information gaining. Our project comes under the category of information gaining. This prototyping model is developed in such a way that it helps blind people to identify a product by capturing it through a camera and receiving an audio output of what the product is. This captured image is processed, identified and then fed into search engines after which the information regarding the product is displayed. The main objective of our project is to design handheld device for the blind persons. The scope of our project is to aid visually challenged people with an assistive text recognition device that would help them to identity objects in their everyday use without seeking the guidance of our individuals.

EXISTING MODEL

This paper puts forth a prototype system of assistive text reading and image processing .The three fictional components include screen capturing, data processing and audio output. The screen capture components captures the scene by motion- based object-deduction using a camera attached to a pair of sun glasses and 'R' denote the calculated foreground object at the each frame. The object of interest is localized by the mean of foreground .Data processing component deploy our proposed algorithms that include object of interest , text localization to obtain image region containing text and to convert the intensified text into readable codes. Mini laptops are used as the processing device in the current prototype. The audio output components read the recognized text code. The Bluetooth ear piece facilitates speech output. The drawback death with these systems is that it is a complex and heavy device and the blind people found it difficult to carry it along with it. The internal architecture consisted of 8051 microcontroller and separate chips for CPU, GPU, and USB controller, RAM which made this model heavy and complicated. Thus our proposed work deals with designing a device which would be handy and easily portable for visually impaired people to carry it along.

FUTURE WORK

The Raspberry Pi is a credit card sized single computer or Sock uses ARM1176JZF-S core. SoC, or System on a Chip, is a method of placing all necessary electronics for running a computer on a single chip. Instead of having an individual chip for the CPU, GPU, USB controller, RAM everything is compressed down into one tidy package. Raspberry Pi needs an Operating system to start up. In the aim of cost reduction, the Raspberry Pi omits any on-board non-volatile memory used to store the boot loaders, Linux Kernels and file systems as seen in more traditional embedded systems. Rather, a SD/MMC card slot is provided for this purpose. After boot load, as per the application program Raspberry Pi will get execute.

FUNCTIONS OF INTERNAL COMPONENTS:

SOC: Broadcom BCM2835 → High Definition Embedded Multimedia Application Processor.

Power Supply: Micro USB power supply

5v, 700mA=3.5w which is less than a bulb. Standard charging port. Universal charging solution by GSMA.

Monitor: Supports only HDMI. Old analogous TV, modern Digital TV desktop monitor or even your Smartphone as a display.

Sd Memory Card: Min 2GB, expandable up to 32GB. Supports only Linux as Fedora, Archlinux & Desbian. Desbian is the most recommended one; the reason is it supports python programming language. The Raspberry Pi versions of kernels are as follows: Fedora – Pidora, Desbian – Raspbian.

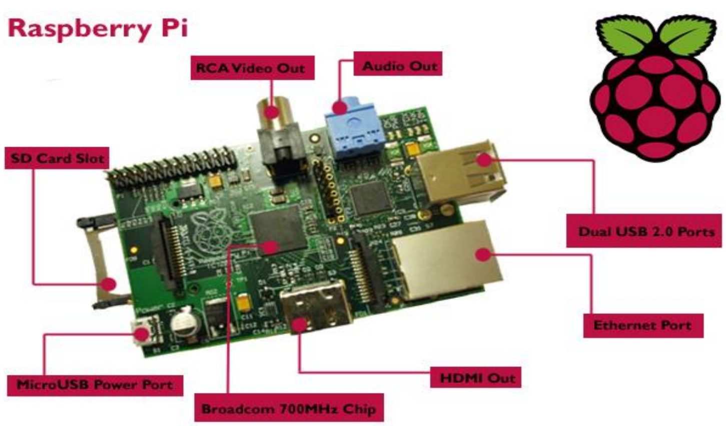


Figure.1. Functional Components of Rasberry pi

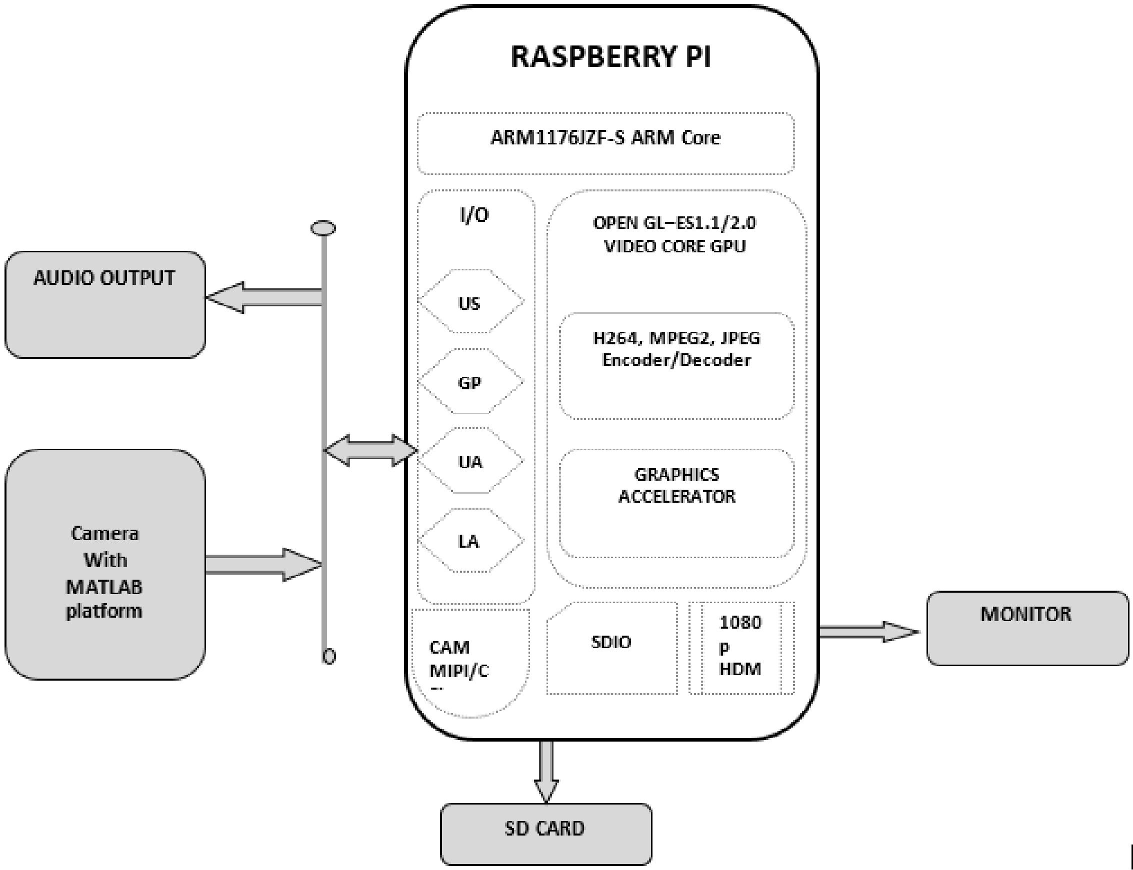


Figure.2.Modules

FUNCTIONAL MODULES

1. Image Capturing: The object from which the text label has to be extracted is captured using a mini-camera which is embedded within our device.

2. Text Extraction: Text from the captured image is extracted using a MATLAB PLATFORM.

2.1: Read Image: This cell of codes read the image to

MATLAB workspace:

2.2: Convert to grayscale image: This cell of codes converts the RGB to gray.

2.3: Convert to binary image: This cell of codes converts the gray to binary image.

2.4: Edge detection: This cell of codes detects the edge of the image.

2.5: Morphology: This cell of codes performs the image dilation and image filling on the image.

2.6: Blobs analysis: This cell of codes finds all the objects on the image, and finds the properties of each object.

2.7: Plot the Object Location:

3. Voice Chip: This cell of codes plots the object locations. We're recovering sounds from objects," he says. "That gives us a lot of information about the sound that's going on around the object, but it also gives us a lot of information about the object itself, because different objects are going to respond to sound in different ways."

Voice module: This is a new and enhanced 8 channel recordable voice module. Each channel can hold up to 1 minute of recorded voice and/or music with a combined total record time of 8 minutes. Recording is quick and easy using the builtin microphone and push to record button.A line-level output jack allows connection to external amplifiers, audio equipment, and paging systems.

Data flow diagram: The data flow diagram clearly explains the flow of the process and also explains how the system works.

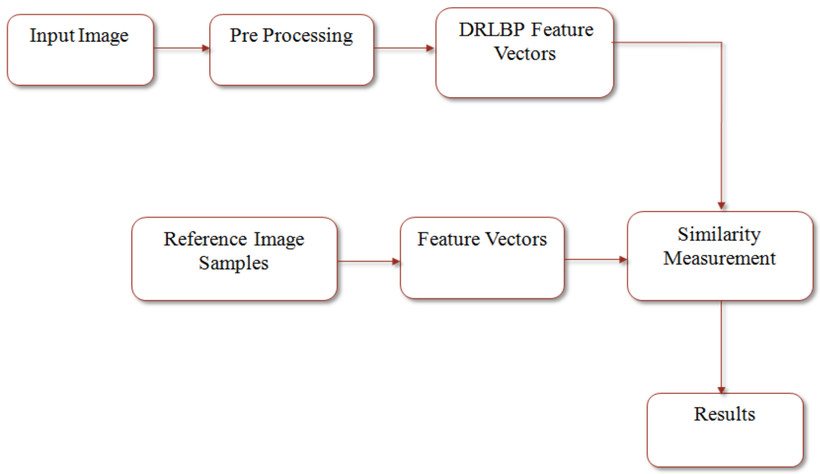


Figure.3. Block diagram

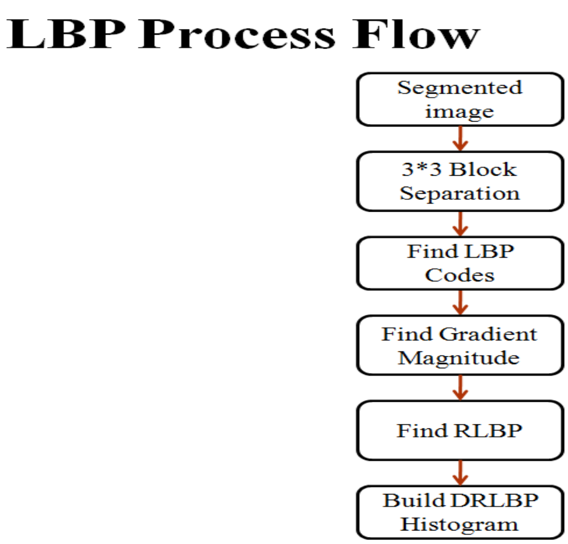


Figure.4. Flow chart

Feature Extraction: The serial communication unit contains MAX 232 IC and two serial ports. One serial port connects the GSM module and another port connects the GPS module**.** Serial communication is basically the transmission or reception of data one bit at a time. Today's computers generally address data in bytes or some multiple thereof. A byte contains 8 bits. A bit is basically either a logical 1 or zero. Every character on this page is actually expressed internally as one byte. The serial port contains an electronic chip called a Universal Asynchronous Receiver/Transmitter (UART) that the conversion. The Video Core MMU maps the ARM physical address space to the bus address space seen by Video Core. The bus addresses for RAM are set up to map onto the uncached1 bus address range on the Video Core starting at 0xC0000000. Physical addresses range from 0x20000000 to 0x20FFFFFF for peripherals. The bus addresses for peripherals are set up to map onto the peripheral bus address range starting at 0x7E000000. Thus a peripheral advertised here at bus address 0x7Ennnnnn is available at physical address 0x20nnnnnn. 1.2.4 Bus addresses the peripheral addresses specified in this document are bus addresses. Software directly accessing peripherals must translate these addresses into

Journal of Chemical and Pharmaceutical Sciences ISSN: 0974-2115
physical or virtual addresses, as described above. Software accessing peripherals using the DMA engines must use bus addresses.

Micro-controller: When you buy a Raspberry Pi, you are essentially buying an assembled printed circuitboard. It does not even include a power supply or operating system. By getting Raspberry Pi set up andready for use.Because the Raspberry Pi just uses standard USB keyboards and mice, most of the setupis pretty straightforward, so we will concentrate on those tasks that are specific to the Raspberry Pi.

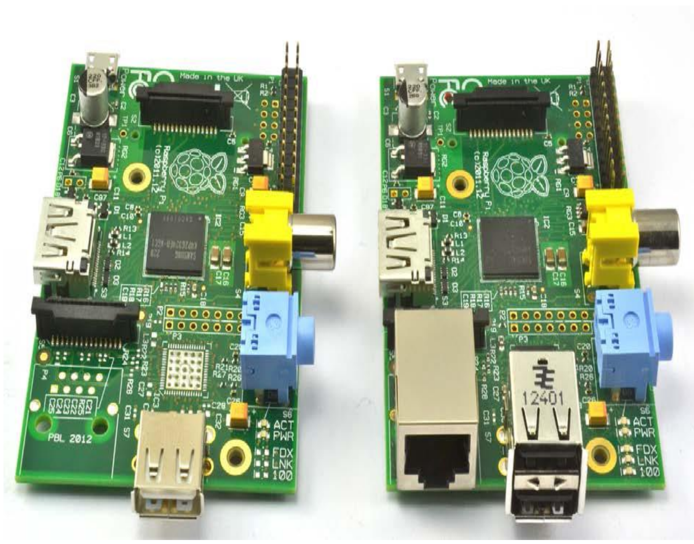


Figure.5.Raspberry pi Microcontroller

MATLAB: MATLAB is for image analysis, image segmentation, image enhancement, noise reduction, geometric transformations, and image registration. Many toolbox functions support multicore processors, GPUs, and C-code generation. Image processing is a method to convert an image into digital form and perform some operations on it, in order to get an enhanced image or to extract some useful information from it. It is a type of signal dispensation in which input is image, like video frame or photograph and output may be image or characteristics associated with that image. Usually Image Processing system includes treating images as two dimensional signals while applying already set signal processing methods to them. It is among rapidly growing technologies today, with its applications in various aspects of a business.

Techniques used:

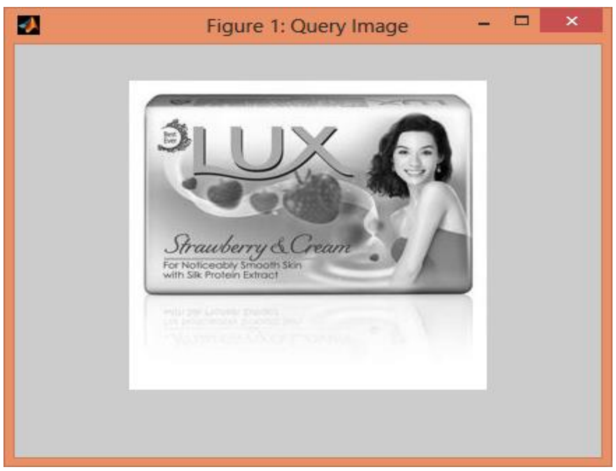
Derivation of generic LBP: The learning vector quantization based approach still has certain unfortunate properties that make its use difficult. First, the differences Mappings of the LBP Labels: Uniform Patterns In many texture analysis applications it is desirable to have features that are invariant or robust to rotations of the input image. As the LBPP,R patterns are obtained by circularly sampling around the center pixel, rotation of the input image has two effects: each local neighborhood is rotated into other pixel location, and within each neighborhood, the sampling points on the circle surrounding the center point are rotated into a different orientation. Another extension to the original operator uses so called uniform patterns. For this, a uniformity measure of a pattern is used: U ("pattern") is the number of bitwise transitions from 0 to 1 or vice versa when the bit pattern is considered circular. A local binary pattern is called uniform if its uniformity measure is at most 2. For example, the patterns 00000000 (0 transitions), 01110000 (2 transitions) and 11001111 (2 transitions) are uniform whereas the patterns 11001001 (4 transitions) and 01010011 (6 transitions) are not. In uniform LBP mapping there is a separate output label for each uniform pattern and all the non-uniform patterns are assigned to a single label. Thus, the number of different output labels for mapping for patterns of P bits is $P(P - 1) + 3$. For instance, the uniform mapping produces 59 output labels for neighborhoods of 8 sampling points, and 243 labels for neighborhoods of 16 sampling points. The reasons for omitting the non-uniform patterns are twofold. First, most of the local binary patterns in natural images are uniform. Ojala et al. noticed that in their experiments with texture images, uniform patterns account for a bit less than 90% of all patterns when using the (8, 1) neighborhood and for around 70% in the (16, 2) neighborhood. In experiments with facial images it was found that 90.6% of the patterns in the (8, 1) neighborhood and 85.2% of the patterns in the (8, 2) neighborhood are uniform. The second reason for considering uniform patterns is the statistical robustness. Using uniform patterns instead of all the possible patterns has produced better recognition results in many applications. On one hand, there are indications that uniform patterns themselves are more stable, i.e. less prone to noise and on the other hand, considering only uniform patterns makes the number of possible LBP labels significantly lower and reliable estimation of their distribution requires fewer samples. The uniform patterns allows to see the LBP method as a unifying approach to the traditionally divergent statistical and structural models of texture.

Eachpixel is labeled with the code of the texture primitive that best matches the local neighborhood. Thus each LBP code can be regarded as a micro-texton. Local primitives detected by the LBP include spots, flat areas, edges, edge ends, curves and so on. The combination of the structural and statistical approaches stems from the fact that the distribution of micro-textons can be seen as statistical placement rules. The LBP distribution therefore has both of the properties of a structural analysis method: texture primitives and placement rules. On the other hand, the distribution is just a statistic of a non-linearly filtered image, clearly making the method a statistical one. For these reasons, the LBP distribution can be successfully used in recognizing a wide variety of different textures, to which statistical and structural methods have normally been applied separately Rotational Invariance Let UP (n,r) denote a specific uniform LBP pattern. The pair (n,r) specifies a uniform pattern so that n is the number of 1-bits in the pattern and r is the rotation of the pattern. Now if the neighborhood has P sampling points, n gets values from 0 to P + 1, where n = P + 1 is the special label marking all the non-uniform patterns. Furthermore, when $1 \leq n \leq P - 1$, the rotation of the pattern is in the range $0 \leq r \leq P - 1$. Let I α◦ (x,y) denote the rotation of image I(x,y) by α degrees. Under this rotation, point (x,y) is rotated to location (x ,y). A circular sampling neighborhood on points I(x,y) and I α◦ (x ,y) also rotates by α If the rotations are limited to integer multiples of the angle between two sampling points, i.e. α = a 360◦ P ,a = 0, 1,...,P − 1, this rotates the sampling neighborhood by exactly a discrete steps. Therefore the uniform pattern UP (n,r) at point (x,y) is replaced by uniform pattern UP (n,r + a mod P) at point (x ,y) of the rotated image. From this observation, the original rotation invariant LBPs introduced in and newer, histogram transformation based rotation invariant features described in can be derived.

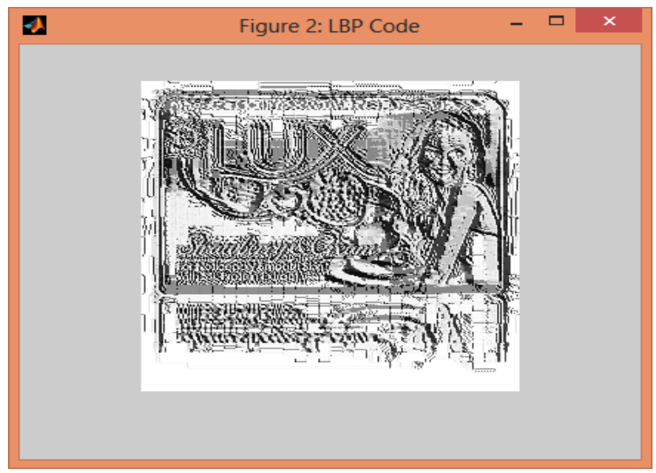
Histogram Transformation: The rotation invariant LBP descriptor discussed above defined a mapping for individual LBP codes so that the histogram of the mapped codes is rotation invariant. In this section, a family of histogram transformations is presented that can be used to compute rotation invariant features from a uniform LBP histogram. Consider the uniform LBP histograms hI (UP (n,r)). The histogram value hI at bin UP (n,r) is the number of occurrences of uniform pattern UP (n,r) in image I . If the image I is rotated by α = a 360◦ P, this rotation of the input image causes a cyclic shift in the histogram along each of the rows, hI α◦ (UP (n,r + a)) = hI (UP (n,r)). (2.14) For example, in the case of 8 neighbor LBP, when the input image is rotated by 45◦, the value from histogram bin U8(1, 0) = 000000001b moves to bin U8(1, 1) = 00000010b, the value from bin U8(1, 1) to bin U8(1, 2), etc. Therefore, to achieve invariance to rotations of input image, features computed along the input histogram rows and are invariant to cyclic shifts can be used. Discrete Fourier Transform is used to construct these features. Let H(n,·) be the DFT of nth row of the histogram hI (UP (n,r)), i.e. H(n,u) = P −1 r=0 hI (UP (n,r))e−i2πur/P . (2.15) In [6] it was shown that the Fourier magnitude spectrum |H(n,u)| = H(n,u)H(n,u) (2.16) of the histogram rows results in features that are invariant to rotations of the input image. Based on this property, an LBP-HF feature vector consisting of three LBP histogram values (all zeros, all ones, non-uniform) and Fourier magnitude spectrum values was defined. The feature vectors have the following form: fvLBP-HF = [|H(1, 0)|,...,|H(1,P/2)|, ..., |H(P − 1, 0)|,...,|H(P − 1,P/2)|, h(UP (0, 0)),h(UP (P, 0)),h(UP (P + 1, 0))]1×((P−1)(P/2+1)+3). It should also be noted that the Fourier magnitude spectrum contains rotation invariant uniform pattern features LBPriu2 as a subset, since |H(n, 0)| = P −1 r=0 hI (UP (n,r)) = hLBPriu2 (n). Complementary Contrast Measure Contrast is a property of texture usually regarded as a very important cue for human vision, but the LBP operator by itself totally ignores the magnitude of gray level differences. In many applications, for example in industrial visual inspection, illumination can be accurately controlled. In such cases, a purely gray-scale invariant texture operator may waste useful information, and adding gray-scale dependent information may enhance the accuracy of the method. In a more general view, texture is distinguished not only by texture patterns but also the strength of the patterns. Texture can thus be regarded as a two-dimensional phenomenon characterized by two orthogonal properties: spatial structure (patterns) and contrast (the strength of the patterns). Pattern information is independent of the gray scale, whereas contrast is not. On the other hand, contrast is not affected by rotation, but patterns are, by default. These two measures supplement each other in a very useful way. The LBP operator was originally designed just for this purpose: to complement a gray-scale dependent measure of the "amount" of texture.

Since contrast is measured locally, the measure can resist even intra-image illumination variation as long as the absolute gray value differences are not much affected. A rotation invariant description of texture in terms of texture patterns and their strength is obtained with the joint distribution of LBP and local variance, denoted as LBPriu2 P1,R1 /VARP2,R2 . Typically, the neighborhood parameters are chosen so that P1 = P2 and R1 = R2, although nothing prevents one from choosing different values. Variance measure has a continuous-valued output; hence, quantization of its feature space is needed. This can be done effectively by adding together feature distributions for every single model image in a total distribution, which is divided into B bins having an equal number of entries. Hence, the cut values of the bins of the histograms correspond to the (100/B) percentile of the combined data.

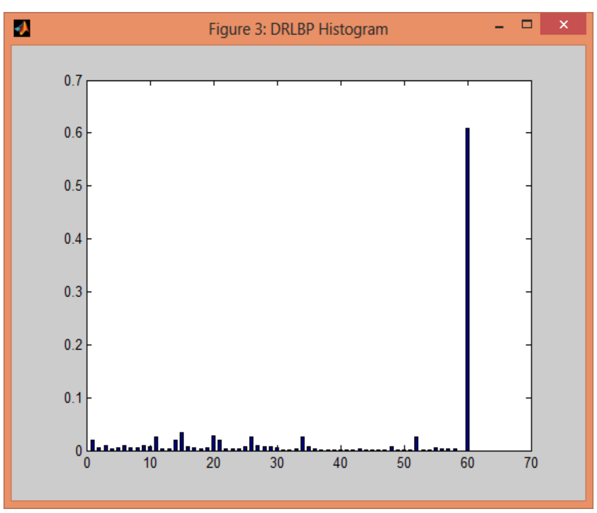
SCREENSHOT



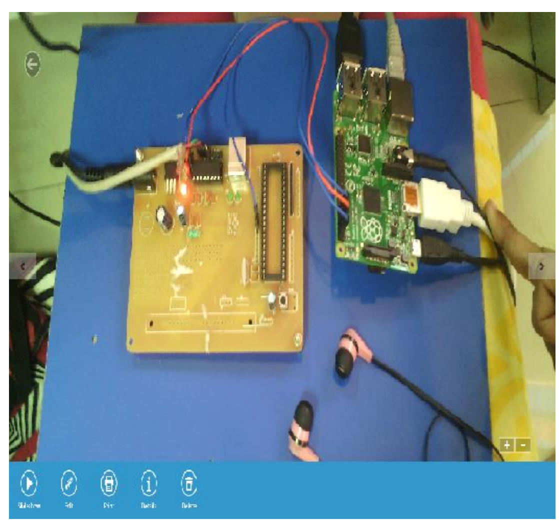
a



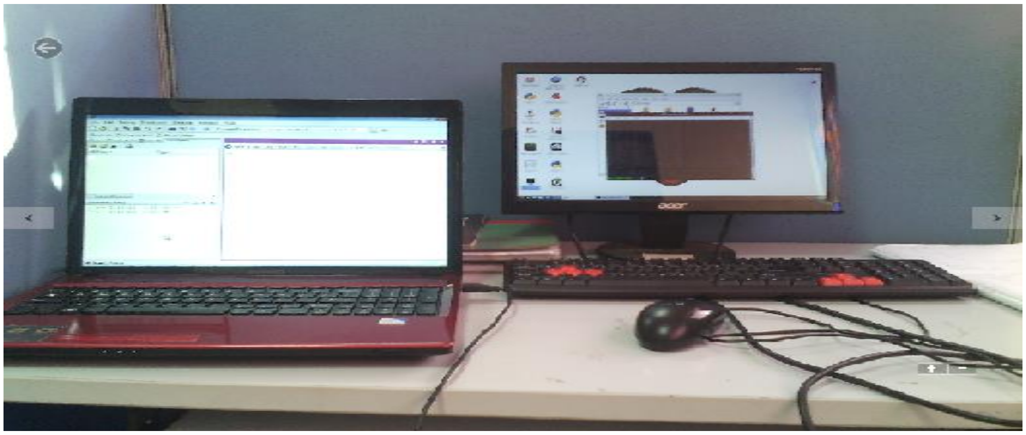
b



c



d



e

Figure.6.Photographs of monitoring system

FApplication and Graphical User Interface: The character recognition application can be used to type every command inside the MATLAB console and workspace and also to use already pre-prepared Graphical User Interface. The GUI consists of files that include all necessary programming code, and also files that include visible interface shapes and forms. The interface employs the workflow of recognition process. After loading the image, we select the character and after that crop it, pre-process, extract feature and finally identify. On every stage, GUI shows us a new image, which is unique for the each step. The images can be viewed in the Main window, RGB, Binary, Crop to Edges and features window.

Comparison: The image processed under mat lab platform in now compared with the images which have been already fed into the database. When the captured image matches with any of the images in the database, then that text is outputted as audio.

Conclusion and future work: This is just a prototyping model. The future work deals with implementing this system in a portable manner. Here only the identified image audio output is taken. As a development to this even the web search output should be outputted as audio.

REFERENCES

A. Shahab, F. Shafait, and A. Dengel, ICDAR 2011 robust reading com- petition: ICDAR Robust Reading Competition Challenge 2: Reading text in scene images, in Proc. Int. Conf. Document Anal, Recognit, 2011, 1491–1496.

B. Epshtein, E. Ofek, and Y. Wexler, Detecting text in natural scenes with stroke width transform, Proc. Compute. Vision Pattern Recognit, 2010, 2963–2970.

D. Dakopoulos and N. G. Bourbakis, Wearable obstacle avoidance elec- tronic travel aids for blind: A survey, IEEE Trans. Syst., Man, Cybern, 40(1), 2010, 25–35.

N. Nikolaou and N. Papamarkos, Color reduction for complex document images, Int. J. Imaging Syst. Technol., 19, 2009, 14–26.

R. Manduchi and J. Coughlan, (Computer) vision without sight, Com- mun. ACM, 55(1), 2012, 96–104.

T.Phan, P.Shivakumara and C.L.Tan, ALaplacianmethodforvideotext detection, Proc. Int. Conf. Document Anal. Recognit, 2009, 66–70.

The Portset Reader, TVI Technologies for the Visually Impaired Inc., Hauppauge, NY, USA. (2012). [Online]. Available: http://www.tvi-web. com/products/porsetreader.html