Blind Guider: An IT Solution for Visually Impaired People

Mallawarachchi Yashas, Maalik K.S., Ashfaq M.Z.M., De Silva K.P.S.H., Ragulan.S.

Sri Lanka Institute of Information Technology
New Kandy Road Malabe Sri Lanka

Abstract **--** "The Blind Guider" application is a real time application for detecting obstacles such as staircases, potholes, pedestrian crossings and moving vehicle detection and getting current location by GPS and giving directions to travel from place to place using the mobile phone camera with voice instructions. This Android application will warn of the obstacles in the road to the user. The system uses Haar-Cascade, BLOB and SURF algorithms for object detection. In order to identify key matching objects and for voice instructions the system uses a machine learning algorithm. Blind people use a white cane to navigate through roads and even indoors. The developed system is much more accurate than the white cane because the system will identify obstacles for the user within a certain distance. This application will provide assistance to the user when required. The mobile camera input is sufficient to commence the processing, the system will assist the user to navigate through obstacles in indoors and outdoors.

Index Terms — Image Processing, OpenCV, Background subtraction, Object detection, GPS, voice recognitions

INTRODUCTION

According to the estimations of World Health Organization, there are 285 million people who were visually impaired worldwide. Regardless of the disability, these people interact with their day today life against all odds. The application "Blind Guider' will help these people to make their life easy. The system is user-friendly, and a low-cost system which can be installed on a smartphone. Though there are multiple types of equipment exist for blind people, the inability to afford these type of equipment has restricted the usage of them. Rather than using a costly system the "Blind Guider" application can be installed directly on any android mobile phone. They can operate android smartphones as a normal person.

Object detection using image processing is a wide area of development [1]. This application uses image processing for object detection and uses mobile phone GPS system to locate the current location of the user. Using GPS technology system is able to guide the blind user. The visually displayed instructions are given via audio / vocal mode. The application is a real time system which detects and identify obstacles and giving warnings to the user. The application provides information about pothole detection, staircases detection, pedestrian crossing, obstacle detection and moving vehicles detection. Additionally, this system has features of typing messages and reading messages to the users.

METHODOLOGY

A collection of several methods are been used in the research to provide an accurate and effective results. The method analyze the frames of the video to detect the movement of the foreground and to extract the object through the color of the object and subtract the background to enhance the foreground. Through the background subtracted image the feature detection been run to identify the foreground. Parallel the audio analysis is also been commenced to recognize the objects that appears in front of the system.

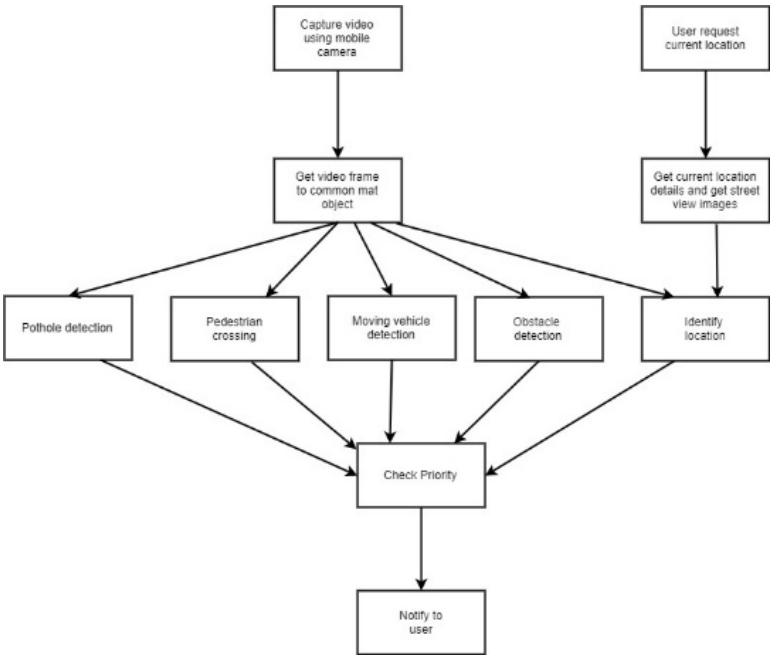

Figure 1.Overview of the Blind Guider System

A. Pothole Detection

Once a pothole is detected in a video frame, the corresponding region is tracked in the subsequent frames until it leaves the viewport [8][9]. To achieve that, the detection algorithm is suspended and the detected pothole region is marked as the region of interest rectangular box. This rectangular area is then

committed to the pothole tracking algorithm that tries to trace this region within the subsequent frames. Once the region of interest leaves the viewport the tracking algorithm is stopped, the detection algorithm is resumed and the number of detected potholes is incremented. The tracking algorithm has to be performed on every frame independently, and in case of multiple potholes, it is executed in parallel.

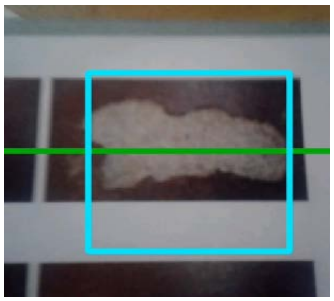

Figure 2: Image of a pothole detection

B. Staircases Detection

There are various kinds of staircases. Stairs consist of a sequence of steps which can be regarded as a group of consecutive curb edges. To extract these features, [17] we start with an edge detection by applying a "Sobel" type operator to obtain the edge map from RGB image of the scene and then perform a Hough transform to extract the lines in the extracted edge map image. These lines are parallel for both stairs. Therefore, a group of concurrent parallel lines represents the structure of stairs. In order to eliminate the noise from unrelated lines, we add constraints including the number of concurrent lines, line length, etc.

C. Pedestrian-Crossing Detection

Pedestrian crossings are dangerous places for a visually impaired person to cross safely [11]. Mostly pedestrian crossing content yellow and white color with a specific shape. Once the pedestrian crossing detected in a video frame, detection algorithm tracks the caption until it remains in the subsequent frames. To detect the pedestrian crossing template matching technique is being used. Once the pedestrian crossing is detected it is alert to the user.

D. Moving vehicle detection

Designed solution mainly focuses on getting information of motion detection and object detection. Because of the noisy background identifying moving vehicle detection was a challenge. All the objects in the video frames are being captured through the camera of the mobile phone. "Haar cascade" algorithm [4] was an efficient method, which was used in order to overcome the challenges. The algorithm uses a collection of trained positive and negative samples via XML files.

Absolute different between frames are used in the system to detect the moving objects. To calculate the absolute frame difference multiple numbers of frames are used. Eight consecutive frames from a video signal are considered and each of them is subtracted from the background image. The difference between images is added together to obtain a resultant image. This image represents the maximum movement of the objects in the given eight frames [5]. Then compute a rectangle box and it is drawn and then moving vehicles are being captured. Then the user identified the direction of the vehicle by drawing a virtual line. Then the frames are touched by the virtual line and then the direction is being identified. The user is being guided by the system. The detected vehicles are notified using voice instructions.

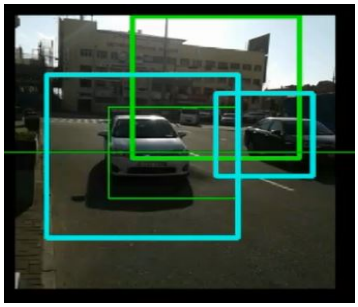

Figure 3: Image from camera frame

E. Obstacle Detection

According to the process of template matching, the technique is being used to detect obstacles in indoor and outdoor environments. For this purpose, the system used the first frame of an object as a template. This template is then used to match with the detected object. This method is the most accurate method for this purpose. For Template matching also OpenCV is being used. After detecting the obstacle within a certain distance using a mobile camera and giving voice instructions to the user.

F. Accurate Determination of Position

This module is capable of guiding the blind by providing voice commands to navigate the user from one place to other by using Google Map API. Here the important part is that the module can differentiate the user's current location with the image provided by the street view API and tell the user as to where to navigate accurately [12][13].

The user will have to specify the destination location through voice commands to the device. Once the user confirms to navigate through voice command, the localization module gets activated. The user will be navigated to the destination through voice command from the device. Once the user reaches the destination, this component should alert the user that he has reached the destination. In order to accurately locate the destination, the localization module will help the user.

Localization module is responsible for accurately identifying or locating the exact place or building that the user intended to reach. To achieve this the module uses Google street view image and image captured from the device. The figure below explains how the localization works.

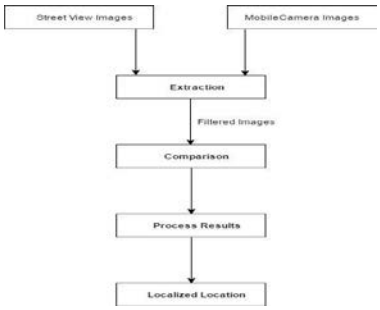

Figure 4: System high level diagram

The extraction component will accept images from Google street view API and image captured from the device. Extraction component will filter the received images and outputs the enhances street view image and the camera image to the comparison module. So the street view image represents the destination place. The camera image taken from the device represents the actual location the user is looking at.

The comparison module should compare the difference between the two-image set and provide data to the process results module. Process results module will obtain the average value from the values received from the comparison component and outputs a percentage value. This value indicated how exactly the two images match. Hence, the system can suggest whether the user has reached his or her exact location.

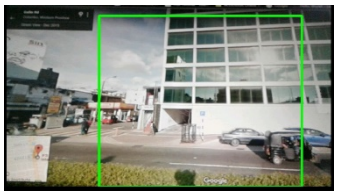

Figure 5: Image from Camera Frame

G. Enable Voice Conversation

Record the sample of each word of the vocabulary is the first task . After speech samples are collected, they are converted an analog signal to digital form by sampling to make them understand the speech system.

Sampling means make the speech signals at regular intervals. The collected data is now quantized to eliminate noise in samples. Then collected samples passed through the feature extraction, feature training, and feature testing stages. Feature extraction transforms the incoming sound into an internal representation. It is possible to reconstruct the original signal from it. Then used MFCC [6] for feature extraction. It is used in every state of speech recognition system. MFCC (Mel-frequency cepstral coefficient) is a standard method for feature extraction in speech recognition task.

Decoding [7] is the most important step in speech recognition. It is performed for finding the best match for the incoming vector matches. A decoder performs the actual decision about recognition of a speech utterance by combining and optimizing the information conveyed by the acoustic and language models. Two types of acoustic models [7]. Word model and phoneme model. It is implemented using HMM (Hidden Markov Model) [&]. It is used in every state of speech recognition system. HMM, state have a set of output symbols known as output probabilities and having a finite numbers of state Q= {q1, q2, ..., qn}. One process is related to the transitions among the states. It is controlled by a set of probabilities called transition probabilities to model the temporal variability of speech. Another process is concerned with the state output observations O= {o1, o2, ..., on} regulated by Gaussian mixture distributions bj(ot) where 1<=j<=N, to simulate the spectral variability of speech. Use of the Viterbi algorithm to generate the given sequence symbols, without having to search all possible sequences. Markov chain has changes to states according to its transition from state i to state j denoted as:

aij= P[Qt+1 = j| Qt = i].

Language models [7] are used to guide the search correct word sequence by predicting the likelihood of nth word using (n-1) preceding words. Language models can be classified into:

- Uniform model: each word has an equal probability of occurrence.
- Stochastic model: a probability of occurrence of a word depends on the word preceding it.
- Finite state languages: languages use a finite state network to define the allowed word sequences.
- Context free grammar: It can be used to encode which kind of sentences is allowed.

In pronunciation modeling [7], during recognition, the sequence of syn del HMM is compared with the se ary to produce sequence of words that is the system's final output contains information about which words are known to the system and how these words are pronounced. A decoder is then used for recognizing words by combining and optimizing the information of acoustic & language models.

RESULTS AND DISCUSSION

The results are conducted to evaluate the performance of the proposed method. Those results are obtained using the detection and tracking algorithms that have been discussed above. Fig.2 shows the pothole detection. Fig.3 shows the results for moving vehicle detection and fig.4 shows the determination of the position. After recognition of features, outputs are given as audio using headphones.

"Blind Guider" system is more accurate for visually impaired people. Though the system is accurate there are some limitations of the system that is this cannot be used in the dim light and the night time. And the other limitation is by using the system user can be see only one side at one time. As future advancement, the

product hope to develop further using LED. Then this system can be used in the night time even.

CONCLUSION

The developed system successfully modeled the object detection using Image processing. This application is developed user friendly and more effectively for blind users. Due to a combination of several methods, the application's accuracy and efficiency have been increased. There using background subtraction, motion detection, edge detection and also feature extraction in open source computer vision (OpenCV) to develop this application. After the obstacles are detected through the camera of the mobile phone, the image processing results are converted into voice instructions to the user. This application is a low-cost application when comparing to other equipment available in the market. Thus this system increases the confidence of blind people by giving them the proper information of the objects around them and enabling them to move independently in indoor and outdoor environment. There are certain limitations and it is important to develop this application for the future.

REFERENCE

[1] "Smart Android Application for Blind People Based on Object Detection," *International Journal of Innovative Research in Computer and Communication Engineering*, Vol. 4,,April.,2016.

[2] "Real-time detection of vehicles using the haar-like features and artificial neuron networks,"[Online] Available : www.sciencedirect.com

[3] Abbas Zohrevand, Alireza Ahmadyfard, Aliakbar Pouyan and Zahra Imani,"A SIFT Based object recognition using contextual Information", *Iranian Conference on Intelligent Systems (ICIS)*,Feb.,pp.1 – 4, 2014.

[4] Priya Samyukthaa M, Siva Bharathi.K, Sivagami.A , "Haar Cascade Algorithm for the Visually Impaired to Detect and Recognize Objects," *International Journal of Engineering Research & Technology (IJERT)*,Vol.5,Apr.,2016. [Online] Available: http://www.ijert.org

[5] Gopal Thapa , Kalpana Sharma , M.K.Ghose ,"Moving Object Detection and Segmentation using Frame differencing and Summing Technique," *Department of Computer Application. International Journal of ComputerApplications*,Vol.102,No.7,Sep., 2014.

[6] Mohit Dua, R.K.Aggarwal, Virender Kadyan and Shelza Dua,"Punjabi Automatic Speech Recognition UsingHTK,",*IJCSI International Journalof Computer Science Issues*,vol.9, no.1,July.,2012.

[7] Rajesh Kumar Aggarwal and M. Dave, "Acoustic modeling problem for automatic speech recognition system,",*Advances and refinements Part(Part II)", IntJSpeechTechnol*,pp.309–320, 2011.

[8] Mednis, G. Strazdins, R. Zviedris, G. Kanonirs, L. Selavo,"Real Time Pothole Detection Using Android Smartphones with Accelerometers,",2011.

[9] A. Kulkarni, N. Mhalgi, S. Gurnani, N. Giri. "Pothole Detection System using Machine Learning on Android," *International Journal of Emerging Technology and Advanced Engineering*,Vol. 4,No.7, pp.360-364, 2014.

[10] Zheng, C. Richard Green,"Vision-based autonomous navigation in indoor environments," ,*IEEE International Conference of Image and Vision Computing New Zealand (IVCNZ), Queenstown, New Zealand*, November., pp.1–7,2010.

[11] McFarlane N.J.B.,Schofield C.P, "Segmentation and tracking of piglets in images,",Vol.8,pp.187–193, 1995.

[12] Jae Sung Cha,Dong Kyun Lim and Yong-Nyuo Shin,"Design and Implementation of a Voice Based Navigation for Visually Impaired Persons".

[13] Mythily.V,Sathish.R, "An Automated Guide Map Application for Blind People Using Android Device".

[14]Jack M. Loomis, Reginald G. Golledge, Roberta L.Klatzky,"Navigation System for the Blind: Auditory Display Modes and Guidance

[15]Mrs. P.M.Daigavane, Dr. P.R.Bajaj, "Real Time Vehicle Detection and Counting Method for Unsupervised Traffic Video on Highways," *International Journal of Computer Science and Network Security*, Vol.10, No.8, August 2010. [Accessed : August 5, 2010,August 20, 2010]

[16]GupteS.,MasoudO.,MartinR.F.K.,PapanikolopoulosN.P,"Detection and Classification of Vehicles," *IEEE Transactions on Intelligent Transportation Systems*, Vol.3 , No.1 , pp.37 - 47, March 2002.

[17] Shuihua Wang, Yingli Tian ,"Detecting Stairs and Pedestrian Crosswalks for the Blind by RGBD Camera",Department of Electrical Engineering,The City College of New York, NY, 10031, USA

AUTHORS

First Author – Mallawarachchi Yasas, yasas@sliit.lk

Second Author – K.S. Maalik, mkamurdeen@gmail.com.

Third Author – Ashfaq M.Z.M, ashfaqzeemath@gmail.com.

Fourth Author – De Silva K.P.S.H, himeshila@gmail.com.

Fifth Author – Ragulan.S, sragulansathya@gmail.com.

Correspondence Author – K.S. Maalik, mkamurdeen@gmail.com, 0754585463