Intelligent Spectacles for the Visually Impaired

Monisha M R[1] Sneha B[2] Surya Karan R[3]

Rajalakshmi Institute of Technology Rajalakshmi Institute of Technology Rajalakshmi Institute of Technology

Abstract:

The world health organization has proposed that there are 250 million visually impaired people around the world. Among the 250 million, about 15 million are in India. The visually challenged are at a considerable difficulty because they often lack the information for avoiding obstacles and hazards in their path, to overcome this ultrasonic stick were introduced. The navigation systems available were hefty and very difficult to operate. Our project has been aimed at progress of smart spectacles which helps in navigation for the visually challenged people. The proposed idea will detect an obstacle using ultrasonic sensors and gives audio instructions through the ear phones. It will capture the images using web camera by detecting the human face when a person enters into the frame. It also involves text extraction from scanned image using Tesseract Optical Character Recognition (OCR) and converting the text to speech by e-Speak tool, which helps the blind people to recognize the surroundings in real world by extracting the text on image and converting it into speech. This technology helps lots of people in the world who experience a significant loss of vision.

Introduction:

An estimation of about 253 million people live with visual impairment: 36 million people are blind and 217 million have moderate to severe vision impairment. 81% of people who are blind or have moderate or severe vision impairment are aged 50 years. Over the last twenty years, the prevalence of vision impairment worldwide has decreased with overall public health action. Blindness affects a person's ability to self-navigate outside well known environments and even simply walking down a crowded street. Blind people started adapting themselves to the method of touch and feel to guide themselves, but it became extremely difficult following that method with the evolution of time and technology. The visually impaired are at a considerable difficulty because they often lack the information for avoiding obstacles and hazards in their path and also walking sticks have not been as efficient as they are considered to be. Our system will help blind people to navigate themselves without anyone's guidance. We have developed a smart spectacle which helps in navigation for the visually impaired people who experience a significant loss of vision.

One of the consequences of vision loss is being uncomfortable about safety while moving around or traveling independently. Individuals with visual impairment have many difficulties in self-navigation in unfamiliar outdoor environments. Safe navigation on sidewalks is the most important requirement. There are many skills and aids considered by professionals working in the field of orientation and mobility to help visually impaired people go outdoors safely. These include the use of canes, guide dogs, and mobility training. There are many studies that consider outdoors problems as well as techniques conceived for safe navigation.

Reading and writing are critical activities for everyone, but for the person with a visual impairment, use of adaptive aids is an integral part of the activity. Since many visually impaired people cannot read regular size print, alternative ways of access must be found. The choice of which aid is used depends on many factors, including type and severity of vision loss, exposure to the kinds of adaptive equipment available, training on the required equipment, funds available, and application in a given setting. Examples of reading and writing aids include: Slate and stylus, Braille, electronic note takers, various writing guides for letters and check writing, talking books, closed circuit televisions, and talking personal data assistants. These aids can help the person with a visual impairment read and write, and allow them access to the printed word in an alternative format.

Orientation and mobility (O&M) skills are another area in which the person with a visual impairment needs to adapt. O & M refers to the training of the person with a visual impairment to ambulate with the use of a white cane. Safe travel both indoors and outdoors is critical for independent locomotion from place to place. In an adjustment to blindness training program, much emphasis is placed on O&M skills building. The training consists of use of alternative mobility aids to assist the person with a visual impairment to safely cross streets, learn directions, understand and interpret the use of tactile markings in sidewalks and roads, and use public transportation system when they are available. The use of the white cane is the most widely used device; with the use of a dog guide the next most common. Use of electronic canes is used when hearing is limited.

Independent living: The use of aids for everyday living is also an important component of daily living for the person with a visual impairment. These aids include the use of 3 talking timers, bump dots for labeling dials, liquid indicators for safe and efficient pouring, and slicing guides for cutting meat and vegetables in the kitchen. All of these aids are used on an as needed basic. Since each individual is different with different needs, some may choose to use these aids while others will improvise or adapt in another way. These examples are only a sample of the items available for use in daily living activities.

Blind stick was an innovative tool designed for visually disabled people for improved navigation. The advanced blind stick allows visually challenged people to navigate with ease using advanced technology. It is integrated with ultrasonic sensors to detect obstacles ahead using ultrasonic waves. On sensing obstacles the sensor passes this data to the microcontroller. The microcontroller then processes this data and calculates if the obstacle is close enough. If the obstacle is not that close the circuit does nothing. If the obstacle is close the microcontroller sends a signal to sound a buzzer. It also detects and sounds a different buzzer if it detects water and alerts the blind. One more feature is that it allows the blind to detect if there is light or darkness in the room. The system has one more advanced feature integrated to help the blind find their stick if they forget where they kept it. A wireless RF based remote is used for this purpose. Pressing the remote button sounds a buzzer on the stick which helps the blind person to find their stick. Thus this system allows for obstacle detection as well as finding stick if misplaced by visually disabled people

The most advanced techniques include Radio-Frequency Identification (RFID), Global Positioning System (GPS), Infrared Light-Emitting Diode (LEDs), wireless sensors, Navigation (NAVIG) devices, remote sighted guidance, an aid to increase the independent mobility of blind travelers interacting with computers (MoBIC), infrared verbal guidance systems and computer vision modules. It has also been shown that mobility performance improves with mobility training.

Literature Review

Image Segmentation:

In 2015, Revathi.K, et.al[23] proposed the methodology on "Face recognition using image processing for visually challenged", tells about the Fisherface approach is also one of the most widely and commonly used methods for feature extraction in facial images. This approach tries to find the projection direction in which, images belonging to different kinds of classes are separated maximally. According to Shang-Hung Lin, Fisherface algorithm is the refinement of the eigenface algorithm to cater the illumination variation in it. Bulhumeur reported that Fisherface algorithm also performs better than eigenface in a circumstance where the lighting condition is varied. The above mentioned approach requires several training images for each face. Therefore, it cannot be implied to the face recognition applications where only one captured or stored image per person is available for training. The core basis for Haar classifier object detection is the Haar-like features. These features, rather than using the intensity values of a pixel, use the change in contrast values between adjacent rectangular groups of pixels. The contrast variances between the areas. Two or three adjacent groups with a relative contrast variance form a Haar-like feature First we need to load the required XML classifier. Enumerous robust algorithms have been developed and they have claimed to have accurate performance to tackle face detection and recognition problems.

In 2004, Rameswar Debnath and Haruhisa Takahashi "Analyzing the Behavior of Distribution of Data in the Feature Space of SVM with Gaussian Kernel" states, PCA is a method in which is used to simplify the problem of choosing the representation of any eigen values and its corresponding eigen vectors to get a consistent representation. It can be obtained by diminishing the dimensional space of the representation. To obtain fast and robust object recognition, the dimensional space has to be reduced. On the whole, PCA also retains the original information of the data. Eigen face based algorithm applies on the PCA basis.

In 2009, Nurulhuda Ismail, Mas Idayu, Sabri "review of existing algorithms for face detection and recognition" tells aboutEigenface based approach is one of the most widely used method for face detection. According to the proposed Pavanet, eigenface is known due to its simplicity, less sensitive in poses and better performance involving small databases or training sets. This approach uses the presence of eyes, nose and mouth on a face and relative distances among these objects. This characteristic feature is known as Eigenface. This facial feature can be extracted by using a mathematical tool called Principle Component Analysis (PCA). By using PCA, any original image from the training set can be reconstructed by combining the Eigenfaces. Generally, a face is classified as itself by calculating the relative distance of the among the features which form the Eigenfaces. Linear Discriminant Analysis (LDA) is also known as Fisher's Linear Discriminant (FLD). It drastically reduces the dimension space by using the FLD technique. FLD technique uses in-class information, reducing variation within each class and increasing class separation within the features specified.

Text to Speech Conversion:

In 2010, Xiaodong Yang, et.al[21] proposed "Context-based Indoor Object Detection as an Aid to Blind Persons Accessing Unfamiliar Environments" Computer vision based indoor way finding system is implemented for blind people to independently access unfamiliar buildings. An unsighted person can discover vivid rooms and building exit or an escalator. This system includes text recognition. It notices the doors based on extensive symmetric shape, by combining edges and corners. To discriminate between an office doors from a bathroom door, it extracts and discover the text information. The organization benefits OCR and region can be conveyed through voice for unsighted travelers. We presented a system for detecting path borders and the vanishing point, such that blind persons can be instructed to correct their heading direction on paths and in corridors. A biologically inspired algorithm for optical flow based on multi-scale key point annotation and matching is used. Moving obstacles can be detected and tracked, such that the blind user can be alerted and informed about the approximate position on the path and whether the object is approaching or not. Detection of moving obstacles complements detection of static obstacles in front on the path, just beyond the reach of the white cane.

In 2011, Xiaodong Yang, et.al[23] proposed "Recognizing Clothes Patterns for Blind People by Confidence Margin based Feature Combination" clothes pattern can be recognized using this system. There are many clothes patterns. This organization is classifying clothes patterns into 4 subparts: stripe, lattice, special, and pattern less. In this organization consistency scrutiny methods only focused on textures varying with special pattern changes. Due to large intra class variations in each clothes pattern category. It cannot achieve level of accuracy for clothes pattern recognition. Withdraw analytical and architectural feature from picture wavelet sub bands can be a result of this issue.

In 2010, Detecting Boris, et.al[10] proposed the methodology called "Text in Natural Scenes with Stroke Width Transform". A unique picture engineer used to find the cost of blow width for each picture pixel. It is used in text detection in natural images. The proposed driver is info reliant and provincial, which makes it agile and it is active enough to trim the use for examine windows or multi-scale estimation. Evaluation testing shows that the implied design exceeds the updated advertised algorithms. Its modesty allows the algorithm to notice contents in many fonts and vocabularies.

In 2011, Asif Shahabad, et.al[3] proposed the methodology of "Robust Reading Competition Challenge 2: Reading Text in Scene Images" International Conference on Document Analysis and Recognition of Text in natural scene images is becoming a prominent research area because imaging devices like mobile phones are available. The ICDAR 2011 Robust Reading Event objection was formulated to assess the act of novel algorithms in perceive and disclose content from complicated images.

In 2007, Sneha Sharma, et.al[19] developed a methodology in "Extraction of Text Regions in Natural Images". The detection and extraction of text regions in an image is a well-known problem in the computer vision research area. The goal of this project is to compare two basic approaches to text extraction in natural (non-document) images: edge-based and connected-component based. The algorithms are implemented and evaluated using a set of images of natural scenes that vary along the dimensions of lighting, scale and orientation. Accuracy, precision and recall rates for each approach are analyzed to determine the success and limitations of each approach. Recommendations for improvements are given based on the results.

In 2010, Dimitris Dakopoulos, et.al[11] proposed "Wearable Obstacle Avoidance Electronic Travel Aids for Blind: A Survey". The variety of portable or wearable navigation systems have been developed to help blind people during navigation in out indoor or indoor environments. There are three major categories of these organizations: Computerized tour aids, and position locator devices, computerized direction aids. This paper is a provisional review of wearable hurdle disclosure systems to users and instructs the scrutiny center about the abilities of these organizations and about the evolution in dependable technology for visually impaired people. The survey is based on various performance parameters and features of the systems that classify them in categories.

In 2016, Bharat Bhargava, et.al[6] proposed the methodology "A Mobile-Cloud Pedestrian Crossing Guide for the Blind". This system helps blind and visually-impaired persons to detect the status of pedestrian signals at street for safe outdoor navigation. This system introduces a mobile-cloud collaborative access for situation familiar rustic exploration, where it uses the estimating power of Resources are made available by cloud computing provider's organization for real-time image processing. The suggested system architecture has the preferences of being essential framework dependence and extensile, hence granting for ample usability. The suggested way is for real-time crossing navigation for blind strollers.

In ICDAR 2003, Yingli Tian, et.al[7] developed a methodology about "Assistive Text Reading from Complex Background for Blind Persons". This paper presents a system for blind persons to read text from object and signage that are held in the hand. The system read text from complex backgrounds and then communicates this information orally. They design a novel text localization algorithm to localize text regions in images with complex backgrounds, by learning gradient features of distributions of edge pixels in an Ad boost model and stroke orientation. Optical character recognition (OCR) software is used to recognize Text characters in the localized regions and transformed into speech outputs. The performance of the proposed system is evaluated on Robust Reading Dataset.

In 2015, Yasser Genera Sahin, et.al[25] proposed "A Smart Tactile for Visually Impaired People". The impaired persons have many difficulties in society. One of the better important complications is traveling because of not suited city layouts. Latest evolutions in technology have facilitated a few provisions, such as palpable paving surfaces, to enhance their lives, but so far there is no extensive results to the issues they face. This study suggests an advanced, low cost and simple system, which consists: identified paths and nightsticks to make travelling unattended possible. The proposed system is available for Android mobile devices and IOS and consists of two software applications, "Out Guide", and "In Guide" for indoor and outdoor environments respectively. A number of portable reading assistants are designed specifically for the visually impaired. "K-Reader Mobile" is a mobile application which allows the user to read mail, receipts, fliers, and many other documents [19]. But these systems/device fail to give an economic solution of the problem and are available on specific platforms. No smart phones have designed for blind person until now. Thus accessibility of the Mobile application is a different question However, the document to be read must be nearly flat, placed on a clear, dark surface and contain mostly black text printed on white background and it does not reads from complex backgrounds. A Navigation System for blind people to navigate safely and quickly, in the system obstacle detection and recognition is done through ultrasonic sensors and USB camera.

The proposed system detects the obstacles up to 300 cm via ultrasonic sensors and sends feedback in the form of beep sound via earphone to inform the person about the obstacle. USB webcam is connected with Raspberry Pi Embedded board which captures the image of the obstacle, which is used for finding the properties of the obstacle (Human Being). Human presence is identified with the help of human face detection algorithm written in Open CV. The constraints coming while running the algorithm on Embedded System are limited memory and processing time and speed to achieve the real time image processing requirements. The algorithm is implemented in Open CV, which runs on Debian based Linux environment [16] Discusses an intelligent system visually impaired people experience difficulty and inconvenience using computers through a keyboard and mouse. This system provides a way to easily control many functions of a computer via speech. When a blind person speaks, the audio voice input is sent to the speech browser and them the output of the search is send through speakers. Many applications are running on this system but not all the applications able to fulfill the required needs but this system has better aspects in future for normal people as well as blind people.

Obstacle Detection

In 2010, IEEE, Dakopoulos, et.al[9] proposed the methodology, "Wearable obstacle avoidance electronic travel aids for blind: A survey" Transport System. According to the sensor type, the obstacle avoidance method can be categorized as: ultrasonic sensor based method, laser scanner based method, and camera based method. Ultrasonic sensor based method can measure the distance of obstacle and compare it with the given distance threshold for deciding whether to go ahead, but it cannot determine the exact direction of going forward, and may suffer from interference problems with the sensors themselves if ultrasonic radar (ultrasonic sensor array) is used, or other signals in indoor environment. Although laser scanner based method is widely used in mobile robot navigation for their high precision and resolution, the laser scanner is expensive, heavy, and with high power consumption, so it is not suitable for wearable navigation system. As for camera based method, there are many methods based on different cameras, such as mono-camera, stereo-camera, and RGB-D camera.

In 2016, Ayush Wattal, Manoj Kumar, Ashutosh Ojha "Obstacle Detection for Visually Impaired Using Raspberry Pi and Ultrasonic Sensors" [5] is based on the mono-camera, some methods process RGB image to detect obstacles by e.g., floor segmentation, deformable grid based obstacle detection etc. However, these methods cost so much computation that they are not satisfied for real-time applications, and hard to measure the distance of the obstacle. To measure the distance, some stereo-camera based methods are proposed. For example, the method uses local window based matching algorithms for estimating the distance of obstacles, and the method uses genetic algorithm to generate dense disparity maps that can also estimate the distance of obstacles.However, these methods will fail under low-texture or low-light scenarios, which cannot ensure the secure navigation. Recently, RGB-D cameras have been widely used in many applications, for their low cost, good miniaturization and ability of providing wealthy information. The RGB-D cameras provide both dense range information from active sensing and color information from passive sensor such as standard camera. The RGB-D camera based method combines range information with color information to extend the floor segmentation to the entire scene for detecting the obstacles in detail. The one in builds a 3D (3 Dimensional) voxel map of the environment and analyzes 3D traversability for obstacle avoidance. But these methods are constrained to non-transparent objects scenarios due to the imperfection of the depth camera.

In 2010, Xiaodong Yang, et.al[27] proposed "Context-based Indoor Object Detection as an Aid to Blind Persons Accessing Unfamiliar Environments" guiding Information Feedback, there are three main techniques for providing guiding information to visually impaired people, i.e., haptic, audio and visual. Haptic feedback based systems often use vibrators on a belt, helmet or in a backpack. Although they have far less interference with sensing the environment, they are hard to represent complicated information and require more training and concentration. Audio feedback based systems utilize acoustic patterns, semantic speech, different intensities sound or spatially localized auditory cues. The method in directly maps the processed RGB image to acoustic patterns for helping the blind to perceive the surroundings. The method in maps the depth image to semantic speech for telling the blind some information about the obstacles. The method in maps the depth image to different intensities sound for representing obstacles in different distance. The method in maps the depth image to spatially localized auditory cues for expressing the 3D information of the surroundings. However, the user will misunderstand these auditory cues under noisy or complicated environment.

Visual feedback based systems can be used for the partially sighted individuals due to its ability of providing more detailed information than haptic or audio feedback based systems. The method in maps the distance of the obstacle to brightness on LED (Light Emitting Diode) display as a visual enhancement method to help the users more easily to notice the obstacle. But, the LED display only shows the large obstacle due to its low resolution. A novel multi-sensor fusion based obstacle avoiding algorithm is proposed to overcome the above limitations, which utilizes both the depth sensor and ultrasonic sensor to find the optimal traversable direction. The output traversable direction is then converted to three kinds of auditory cues in order to select an optimal one under different scenarios, and integrated in the AR technique based visual enhancement for guiding the visually impaired people.

Proposed Methodology:

In the proposed methodology, it describes the proposed method by using a detailed description on the process used. The block diagram of proposed model is explained and the working principle of various elements is also mentioned in it. The proposed algorithm can effectively handle complex background and multiple patterns, and extract text information from both hand-held objects and nearby surroundings.

Block Diagram:

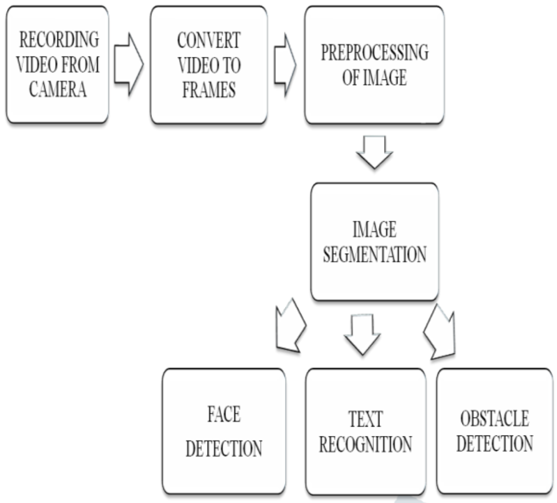

Fig.1 Block Diagram of Image segmentation

The fig.1 describes the block diagram of image segmentation where in, the recording video from the camera is done first and the recorded video is converted into frames which captures 20 frames for 5 seconds. Then pre-processing of the image is done in which the image is filtered and noise is removed. The preprocessed image is then segmented. The results will be in three forms. Firstly the human face is detected and it is stored in the dataset. Then the stored image will mention the person's name and display as unknown if the human face is not stored. Second is the text recognition that will read the printed text in nearby sources which will be captured through web camera. When the image is captured, the texts are segmented using Tesseract OCR will get converted into audio output using e-speak tool. The Motion-based method is used to detect the object using ultrasonic sensor at a distance of about 5m which will alert the person by using a buzzer. For image segmentation, the Haar Cascade algorithm is used where the human faces are detected and stored as captured image. If the detected faces are in the database, it will give their names else it will show them as an unknown person.

Text to speech conversion– Functional Diagram

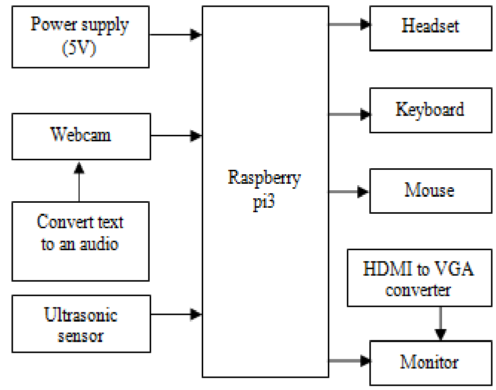

Fig.2 Functional Diagram of Text to Speech conversion

In the fig 2, a prototype system to read printed text for assisting blind persons is given. Our proposed algorithm can effectively extract the text from nearby sources which will be captured using the web camera. When the image is captured, the texts are segmented using Tesseract OCR and the segmented texts are converted into audio output using e-speak tool. Using ultrasonic sensor, the front near view objects will be detected and it will alert the person by using a buzzer when the person approaches the object at a distance of about 50cm. Added to this, by using Haar Cascade algorithm the human faces are detected from the captured image. If the detected faces are in the database, it will give their names else it will mention them as an unknown person and it is implemented using open CV and python software installation. In order to solve the common aiming problem for blind users, we have proposed a motion-based method to detect the object of interest. The automatic ROI detection and text localization algorithms were independently evaluated as unit tests to ensure effectiveness and robustness of the whole system. Since it is difficult for blind users to aim their held objects, we employed a camera with a reasonably wide angle. When our application starts running it first check all the devices and resources which it needs are available or not. After that it checks the connection with the devices and gives control to the user. A label is used for displaying the image taken from the camera. A status box is for representing the detected data from the image. The capture button is to detect the data from the image. The detect button is to detect the human from the video streaming in front of the camera. The audio jack port is the output port here. The Raspberry board comes with integrated peripherals like USB, ADC and Serial etc.

Working Principle:

The camera acts as main vision in detecting the label image of the board, then image is processed internally and separates label from image by using open CV library and finally identifies the text and identified text is pronounced through voice. Now it identifies received image, which is converted to text by using tesseract library. Once the identified name is converted to text and converted text is displayed on display unit connected to controller.

Now converted text should be converted to voice to hear the name as voice through ear phones connected to audio jack port using flite library.

By means of text-to-speech technology, the person comes to know the texts in front of him. Then the camera is focused, which captures the image and written (printed) text is converted into speech which the person can then hear. Since this is mainly for printed texts, it is computationally feasible. Now the person can read directions, books, newspaper, price on grocery items, number plates, and so on, independently. All this is using a Raspberry pi3, camera and audio jack. This proves out to be a cheap and an elegant solution. All this is done by using tesseract software for the conversion of the image to text, can process the image and convert its content to text. Such types of softwares work on the principle of Optical Character Recognition. The text is then passed to any standard text-to-speech software like eSpeak, which renders the text to audio which the person can easily hear. This sounds quite tedious but we have developed a working prototype of the same. Once we had added the camera along with a strong microcomputer like Raspberry pi, we thought of using it to our advantage. So we went ahead and thought of the next major problem faced by visually impaired people, the difficulty that they face in order to freely interact with a person. Hence we integrated facial recognition to our initial prototype. Visually impaired people now have to depend on no one for recognizing people, since the project is capable of saving new faces and also recognize the old ones. It uses Open CV's LBPH Face Recognizer to train on image dataset from Yale. Once trained it takes an input from the Raspberry Pi's camera and add depending whether it is able to recognize the image or not, it places in training images so that it can identify such face in future or correctly predicts the familiar face. Given that in any normal interaction people generally face each other to interact; we used Haar Cascades frontal face features to train our model on. The algorithm is simple but offers a very elegant solution to the ones who actually need it the most.

Optical Character Recognition (OCR)

Optical Character Recognition or OCR is the text recognition system that allows hard copies of written or printed text to be rendered into editable, soft copy versions. It is the translation of optically scanned bitmaps of printed or written text into digitally editable data files. An OCR facilitates the conversion of geometric source object into a digitally representable character in ASCII or Unicode scheme of digital character representation. Many a times we want to have an editable copy of the text which we have in the form of a hard copy like a fax or pages from a book or a magazine. The system employs the use of an optical input device usually a digital camera or a scanner which pass the captured images to a recognition system that after passing it through a number of processes convert it to a soft copy like an MS Word document. When we scan a sheet of paper we reformat it from hard copy to a soft copy, which we save as an image. The image can be handled as a whole but its text cannot be manipulated separately. In order to be able to do so, we need to ask the computer to recognize the text as such and to let us manipulate it as if it was a text in a word document. The OCR application does that; it recognizes the characters and makes the text editable and searchable, which is what we need. The technology has also enabled such materials to be stored using much less storage space than the hard copy materials. OCR technology has made a huge impact on the way information is stored, shared and communicated.

OCR's are of two types,

i. OCR's for recognizing printed characters.
ii. OCR's for recognizing hand written texts.

OCRs meant for printed text recognition are generally more accurate and reliable because the characters belong to standard font files and it is relatively easier to match images with the ones present in the existing library. As far as hand writing recognition is concerned the vast variety of human writing styles and customs make the recognition task more challenging. Today we have OCRs for printed text in Latin script as an everyday tool in offices while an OCR for hand writing is still in the research and development stage to have more result accuracy. Optical Character Recognition (OCR) is one of the most common and useful applications of machine vision, which is a sub-class of artificial intelligence, and has long been a topic of research, recently gaining even more popularity with the development of prototype digital libraries which imply the electronic rendering of paper or film based documents through an imaging process.

3.5 OCR PROCESSES

The OCR process begins with the scanning and subsequent digital reproduction of the text in the image. It involves the following discrete sub-processes, as shown in Fig 3,

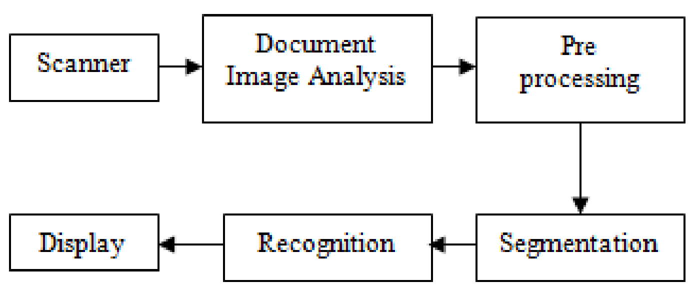

Fig 3 OCR Processes

HAAR- Cascade

Haar-cascade is an object detection algorithm used to locate faces, pedestrians, objects and facial expressions in an image, and mainly used for face detection. In Haar-cascade, the system is provided with several numbers of positive images (like faces of different persons at different backgrounds) and negative images (images that are not faces but can be anything else like chair, table, wall, etc.), and the feature selection is done along with the classifier training using Integral images.

Features used by Haar-cascade

In general, three kinds of features are used in which the value of a two rectangular features is the difference sum of the pixels within two rectangular regions. These regions have same shape and size and are horizontally or vertically adjacent as shown in Fig 4. Where as in the three rectangular features are computed by taking the sum of two outside rectangles and then subtracted with the sum in a center rectangle. Moreover, in the four rectangles feature computes the difference between diagonal pairs of rectangles.

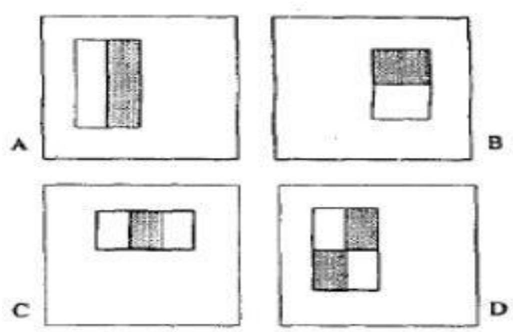

Fig.4 Example rectangle features used in Haar-cascade

The sum of pixels in the white rectangles is subtracted for the sum of the pixels in the grey rectangles. Here A and B are two rectangle feature, and C and D are three and four rectangle feature.

Ultrasonic Sensors

An Ultrasonic sensor (fig 3.5) is a device that can measure the distance to an object by using sound waves. It measures distance by sending out a sound wave at a specific frequency and listening for that sound wave to bounce back. By recording the elapsed time between the sound wave being generated and the sound wave bouncing back, it is possible to calculate the distance between the sonar sensor and the object.

Since it is known that sound travels through air at about 344 m/s (1129 ft/s), you can take the time for the sound wave to return and multiply it by 344 meters (or 1129 feet) to find the total round-trip distance of the sound wave. Round-trip means that the sound wave traveled 2 times the distance to the object before it was detected by the sensor; it includes the 'trip' from the sonar sensor to the object and the 'trip' from the object to the Ultrasonic sensor (after the sound wave bounced off the object). To find the distance to the object, simply divide the round-trip distance in half.

Distance = (Speed of Sound * Time taken) / 2

The Ultrasonic Sensors belongs to a category of sensors that emits ultrasound i.e. sound of frequency more than 20 kHz. A trigger pulse is given as an input to the ultrasonic sensor using Raspberry Pi. The ultrasonic sensor then emits a short 40 kHz ultrasonic burst signal. The burst signal when travelling through the atmosphere at approximately 343ms-1, hits an object and then rebounds back to the sensor resulting in an output pulse. This output pulse is captured by Raspberry Pi. Using the time taken by the pulse to return back we calculate the distance from the obstacle. The sensor consists of four pins:

1. VCC **-** It is used to provide 5V power to the sensor.
2. Trigger (Trig) **-** Takes in Input Pulse to trigger the sensor.
3. Echo **-** It is used to receive the Output Pulse i.e. the echo from the object detected.
4. Ground (GND) **-** It connects sensor to the ground.

The ultrasonic sensor transmits sound waves and receives sound reflected from an object. When ultrasonic waves are incident on an object, diffused reflection of the energy takes place over a wide solid angle which might be as high as 180 degrees. Thus some fraction of the incident energy is reflected back to the transducer in the form of echoes. If the object is very close to the sensor, the sound waves returns quickly, but if the object is far away from the sensor, the sound waves takes longer to return. But if objects are too far away from the sensor, the signal takes so long to come back (or is very weak when it comes back) that the receiver cannot detect it.

Fig.5 Ultrasonic Sensor (HC-SRO4)

Tesseract

Tesseract is meant for OCR engine with open source. Tesseract can be used for OCR tasks, which is very complicated and suitable for the use of a backend which includes layout analysis by using frontend such as OCR opus. When the input images captured/recorded are not preprocessed, the output of Tesseract's will be shown as a poor quality. The text x-height should be at least 20 pixels of a screenshot image, any skew or rotation should be corrected or text will not be recognized, high pass filter must be used for low frequency changes in brightness, or Tesseract's binarization stage will wipe out much of the page, and dark borders must be manually detached or they will be misinterpreted as character.

Thus, this chapter concludes by illustrating detailed explanation of working principle of image segmentation, OCR Process, Ultrasonic sensor, Haar cascade and Tesseract algorithm.

Experimental Results and Discussion

A description about the prototype system is given, to recognize the person by segmenting the image and it mentions him/her as a human through the audio output. It includes that it reads printed text on nearby surroundings for assisting blind persons. Here in the fig 4.1, we have focused the webcam on the text named as "C170 WEBCAM" and the text is displayed in the monitor and the audio of that text is heard through the earphones. The output of the text to speech conversion has efficiency of about 75%. It notifies the user about obstacles by measuring the distance and giving it in the form of audio output. Faces are recognized based on the database stored in the folder. The faces which are not stored are recognized as unknown person. Here, the faces are captured in the real time basis, and they are stored in dataset by capturing 20 frames for 5 seconds. In the fig 4.2, face detection of Dr. Abdul Kalam is done and the name is displayed in the blue box. The image segmentation in the project is having efficiency of about 75%. The system is successful in warning the user about the presence of obstacles and human and the type of texts in their path.

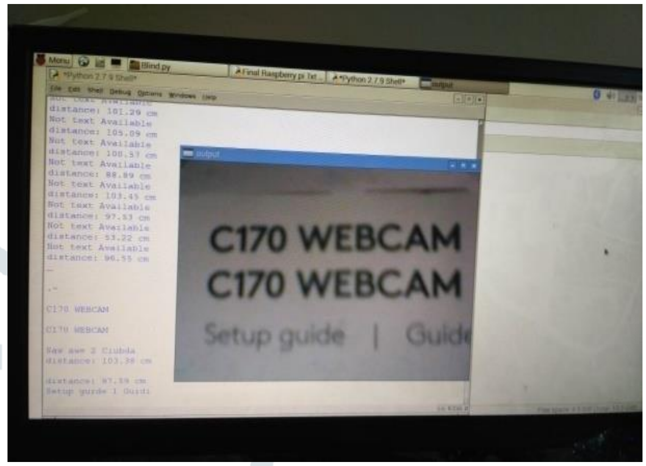

Fig.7 Text to speech Output

AGE GROUP(yrs)	NAME	FACE DETECTED?	
		YES	NO
10-20	Image1	√	
	Image2	√	
21-25	sssImage3	√	
	Image4	√	
30-40	Image5	√	
	Image6	√	
40-50	Image7		×
	Image8	√	
50-70	Image9	√	
	Image10	√	
70 & above	Image11		×
	Image12	√	

Table.1 Experiments conducted for various age groups

Simulated Results

Fig.6 Face Detection Output

FONT FACE NAME	**TEXT DETECTED AND HEARD?**	
	YES	NO
TIMES NEW ROMAN	√	
CALIBRI(BODY)	√	
LUCIDA CALLIGRAPHY		×
ARIAL	√	
ALGERIAN		×

Table.2 Experiments conducted for various font faces

INFERENCE

From the table 4.1, the efficiency of the image segmentation for about 12 images of different age groups has been recorded and the efficiency gained was about 75% in this project. From the table 4.2, the efficiency of the text to speech conversion for about 5 font faces has been observed and the efficiency of this was about 70%.

Conclusion

The prototype system is used to read printed text for assisting blind persons. 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, we have proposed a novel text localization algorithm based on tesseract OCR. The corresponding feature maps estimate the global structural feature of text at every pixel. Furthermore, identifying of human faces is also done by storing their faces in a dataset by real time basis. The faces which are not stored will display as “unknown”. By using ultrasonic sensor, distance measurement for the obstacles is also displayed (in cm) and heard using earphones. As the Raspberry Pi board is powered the camera starts streaming. When the object for text reading is placed in front of the camera then the capture button is clicked to provide image to the board. Using Tesseract library the image will be converted into data and the data detected from the image will be shown on the status bar. The obtained data will be pronounced through the ear phones using file library. Presence of every module has been reasoned out and placed carefully thus contributing to the working of the unit.

References:

[1] Advance Data Reports from the National Health Interview Survey (2008).

[2] An overview of the Tesseract OCR (optical character recognition) engine, and its possible enhancement for use in Wales in a pre-competitive research stage Prepared by the Language Technologies Unit, Bangor University April 2008.

[3] Asif Shahab, Faisal Shaft, Andreas Dingell “Robust Reading Competition Challenge 2: Reading Text in Scene Images” 2011 International Conference on Document Analysis and Recognition.

[4] Athira Panicker, Anupama Pandey, Vrunal Patil “Smart Shopping Assistant Label Reading System With Voice Output For Blind Using Raspberry Pi” International Journal of Advanced Research in Computer Engineering & Technology (IJARCET), 2016.

[5] Ayush Wattal, Manoj Kumar, Ashutosh Ojha “Obstacle Detection for Visually Impaired Using Raspberry Pi and Ultrasonic Sensors” National Conference on Product Design (NCPD 2016), July 2016.

[6] Bharat Bhargava, Paling Angin, Lian Duan “A Mobile-Cloud Pedestrian Crossing Guide for the Blind”2016.

[7] Bingil Tian, Chula Yi “Assistive Text Reading from Complex Background for Blind Persons”2016.

[8] Chen. X and Yuille. A. L, “Detecting and reading text in natural scenes,” in Processing Computer Vision Pattern Recognition, 2004.
[9] Dakopoulos. D, N.G.Bourbakis, “Wearable obstacle avoidance electronic travel aids for blind: A survey” IEEE Transport System, January 2010.

[10] Detecting Boris, Epstein Eyal, Yonatan Wexler “Text in Natural Scenes with Stroke Width Transform” 2010.

[11] Dimitris Dakopoulos, Nikolaos G. Bourbakis, Fellow, “Wearable Obstacle Avoidance Electronic Travel Aids for Blind: A Survey” IEEE 2010.

[12] Epshtein. B, E. Ofek, and Y. Wexler, “Detecting text in natural scenes with stroke width transform” in Processing Computer Vision Pattern Recognition, 2010.

[13] Giudice. N and G. Legge, “Blind navigation and the role of technology” in The Engineering Handbook of Smart Technology for Aging, Disability and Independence, 2008.

[14] http://www.catwig.com/google-glass-teardown/teardown/glass-clearshade-isometric.jpg.

[15] Infant Abinaya. R, E.Esakkiammal, Pushpalatha “Compact Camera Based Assistive Text Product Label Reading and Image Identification for Hand-Held Objects

for Visually Challenged People" International Journal of Computer Science and information technology, 2017.

[16] "Intelligent system based on speech recognition with capability of self-learning" Sneha K. Upadhyay, Vijay N, International journal for research in applied science and engineering technology, 2015.

[17] International Workshop on Camera-Based Document Analysis and Recognition (CBDAR 2005, 2007, 2009, 2011).

[18] Jinqiang Bai, Shiguo Lian, Member, IEEE, Zhaoxiang Liu, Kai Wang, Dijun Liu "Smart Guiding Glasses for Visually Impaired People in Indoor Environment" 2017.

[19] KNFB reader http://www.knfbreader.com/products-kreadermobile.php.

[20] K Reader Mobile User Guide, knfb Reading Technology (2008), http://www.knfbReading.com.

[21] Kumar. S, R. Gupta, N. Khanna, S. Chaudhury, and S. D. Joshi, "Text extraction and document image segmentation using matched wavelets and MRF model," IEEE Image Processing, August 2007.

[22] L. Ma, C. Wang, B. Xiao, "Text detection in natural images based on multi-scale edge detection and classification," In the International Congress on Image and Signal Processing (CISP), 2010.

[23] Revathi.K , Jaya Bharathi.M , Saranya "Face recognition using image processing for visuallychallenged". International Conference on Science, Technology, Engineering & Management [ICON-STEM'15].

[24] Sneha Sharma, Dr.Roxanne Canoas "Extraction of Text Regions in Natural Images" 2007.

[25] Velmurugan. D, M.S.Sonam "A Smart Reader For Visually Impaired People Using Raspberry Pi" International Journal Of Engineering Science And Computing, March 2016.

[26] World Health Organization (2009) 10 facts about blindness and visual impairment[Online].www.who.int/features/factfiles/blindness/blindness_facts/en/index.html.

[27] Xiaodong Yang, Tian Chucai Yi Aries Arditi "Context-based Indoor Object Detection as an Aid to Blind Persons Accessing Unfamiliar Environments" 2010.

[28] Xiaodong Yang, Shuai Yuan, YingLi Tian "Recognizing Clothes Patterns for Blind People Confidence Margin based Feature Combination" 2011.

[29] Yang. J, J.Zhang, and A. Waibel, "Automatic detection and recognition of signs from natural scenes," IEEE Image Processing, January 2004.

[30] Yasser Gunner Sahin, Bari's Aslan, Sinan Talebi, Ayberk Zeray "A Smart Tactile For Visually Impaired People" 2015.