



Obstacle avoidance using stereo vision and depth maps for visual aid devices

Vaibhav Bansal¹  · Krithika Balasubramanian¹ · P. Natarajan¹Received: 26 August 2019 / Accepted: 22 April 2020 / Published online: 28 May 2020
© Springer Nature Switzerland AG 2020

Abstract

Comprehending the environment accurately and proficiently is one of the fundamental undertakings for the visually impaired. Many electronic travel aid devices are utilizing stereo vision techniques. However, there is no existing technology that helps to fulfill the requirements of the visually impaired and be economical at the same time. All the current technologies are implemented only for a specific purpose and have some constraints. In this paper, we propose a new framework that integrates the critical aspects of stereo vision with some additional features for aiding the visually impaired to solve their existing problems efficiently. In this system, both stereo vision and ultrasonic sensors have been integrated to get a precise awareness of the obstacles present in the real-time surroundings. The central concept is to categorize point vertices for the detection of objects in three-dimensional space according to height, width and slope of the adjacent vertices present in the images captured by CCD cameras. The intended assistive system serves to be more cost-effective for the visually impaired with the usage of CCD cameras and low-budget sensors. The assistive system in its recognition phase employs the proposed obstacle detection algorithm using stereo vision to recognize and track impediments that are at a farther distance. While, it employs ultrasonic sensors to detect obstacles present at shorter distances. The system also overcomes the problem of finding holes and descending stairs, faced by devices that use monoscopic vision. Additional features, like the Buzzer, GPS, Voice module, Alert message, are being incorporated. The principle intention of this paper is to contribute our knowledge to assist the visually impaired to become independent and help them to adapt to various daily situations.

Keywords Stereo vision · Image processing · CCD · Buzzer · Glasses · Distance measurement · Electronic travel aids (ETA) · Low-cost devices

1 Introduction

Handicap refers to the inability to perform ordinary tasks, i.e., the disability to do socially structured set of responsibilities. For achieving goals and objectives in life, independence plays a significant role. Visually impaired individuals always find themselves in a challenging situation to travel independently. Many visually impaired individuals face difficulties to steer an independent life without the support of others. For several years, white-cane

technology has been an excellent pillar of support for visually disabled people, which is employed for navigation purposes to avoid any obstacle in their way. Now the white-cane technology is upgraded by the introduction of sensors in it to build its proficiency. To fulfill the requirements of the visually impaired, we need to understand and collect more information about the size and characteristics of this target audience. The gained knowledge and information will help to create more effective visual aids to provide better facilities.

✉ Vaibhav Bansal, vaibhav.bansal2016@vitstudent.ac.in; Krithika Balasubramanian, krithika.bsubramanian2016@vitstudent.ac.in; P. Natarajan, pnatarajan@vit.ac.in | ¹School of Computer Science and Engineering, Vellore Institute of Technology, Vellore, India.



Humans can estimate the object size and how far an obstacle is present with the help of the human “stereo vision” eyes. Using this concept, we introduce a system that utilizes stereo vision to calculate the size and the displacement of the obstacles in its field of view. Stereo vision using CCD cameras acts as an image capturing device, which accounts to be an alternative for the usage of expensive sensors. This work concentrates more on the following features like obstacle detection, blob removal, analyzing the physical features of the obstacle and location of the object. Usage of single cameras for obstacle detection has proven to be ineffective. Therefore, we can obtain accurate results by using stereo vision with the help of vertically oriented CCD cameras. Since CCD cameras are less expensive, easy to deploy and utilize, our system proves that the usage of multiple cameras provides better efficiency and accuracy. The system can overcome all the difficulties that the dog, white cane, CG system, Adaptive thresholding technique, DEEP-SEE framework using monocular vision and other visual aid devices do not respond. The comparison analysis of the mentioned techniques used in assistive devices is mentioned in the literature survey.

Stereo vision gives us a 3-D view of an object, i.e., the height, width and depth of the object. The inclusion of depth dimension makes the stereo vision unique and better. Using this aspect, we have made an effort to integrate stereo vision with the visual aid system, for example, smart stick, wheel-chair etc. The equipment will have a pair of CCD cameras that capture the stereo images of the obstacles ahead, which will lead to the processing of a depth map. Ultrasonic sensors are also used in the system to detect obstacles for a shorter range. Using the information obtained through image processing, the sensors attached to this equipment will send an alert message, either using a buzzer or a notification is sent to their emergency contact person. Some additional features, like the GPS module, voice module, and information about location coordinates, are being incorporated. The system covers the information about the obstacles both at a shorter and broader range. We intend to add a perception framework that utilizes stereo vision to complete the loop.

2 Literature survey

According to surveys taken by WHO on a worldwide scale, it has resulted that almost 1.30 billion people lead a visually impaired life. Out of 7 billion people, approx. 180 million suffer from a clement vision disability, 215 million have average to acute vision disability and 36 million people have distant visual impairment [1]. On the other hand, around 820 million people lead a life with a near

vision disability [2]. According to WHO, the survey, taken for all age groups, shows that around 280 million people have visual impairment, out of which the number of people who have lost their eyesight entirely is approx. 40 million [3].

India is known to have the most significant number of blind people in the world [4]. Blind people have used conventional aids like long cane and guide dogs for a long time. The usage of sonar for the detection of obstacles has proved to be an ineffective method as the information obtained is not precise compared to results obtained from the long cane. Long cane is expensive and does not satisfy the requirements of the visually impaired. Since these conventional methods have been proved ineffective, the introduction of ultrasonic sensors on these devices has been introduced. Ultrasonic sensors are used in the calculation of the distance of any object present in the surrounding. The use of ultrasonic sensors was first started in the early Electronic Travel Aids like Batcane, Columbia talking compass, Guidecane, Lasercane, Mowat sensor, the Sonicguide and the Polaran [5]. The information conveyed is usually through vibration and sound signals.

Among the Electronic Travel Aids, the most conventional ones which primarily used ultrasonic sensors are the Sonic Guide of Leslie Kay, in which echolocation using ultrasonic waves helps the user to understand the environment. The Sonic Guide consists of four Sonic Guide temples which measure the pressure level of sound through the sound level meter. The values of mean, standard deviation and ANOVA obtained using the pressure level value is used to evaluate the regularity of the output. This method has proved to be an effective method for the visually impaired as they were able to regain their partially lost sense of sight and increased their ability to have a better understanding of the surroundings. Unfortunately, the disadvantage of this system is constricted mobility, a large number of ultrasonic reflections, highly overpriced and does not give effective results due to the transfer of feedback signal. Apart from Sonic Guide, other ETA systems have also been used for navigation purposes [6]. The CG system uses the stereoscopy and kinect to compact the cloud of 3D vertices to find the vanishing point to avoid the obstacles. It operates within the range of 40 cm – 4 m. The CG system is not affordable for outdoor use as it has confined perception to recognize the surrounding landmarks. In auto-adaptive thresholding, it has shown a reduction in performance as kinect shows low accuracy in short range. The system faces the challenge of differentiating the objects when there is an increase in the distance between the user and impediments.

Fusion of artificial vision and GPS (FAV&GPS) is an assistive device that enhances the process of identification of the user’s location and its surrounding based on

a map matching approach and artificial vision. The system performs the function of identifying the surrounding objects as well as automatic recognition of visual targets using the SpikNet recognition algorithm. The introduction of the modified Geographic Information System (GIS) manifests a better assessment of the user's current location in comparison with the commercial GIS. However, there is no evidence of the integration of this model in a navigation system. Apart from this, an assistive device using DEEP-SEE framework that uses computer vision algorithms and deep convolution neural network(CNN) are used to recognize and track obstacles in real-time during navigation in outdoor [7]. However, assistive devices using the DEEP-SEE framework cannot overcome the problem of recognizing descending stairs or holes in the ground using a monoscopic vision. Thus, we come to the usage of stereo vision for the calculation of distance. Stereo vision helps in drawing out of 3D representations of obstacles and for depth analysis. The stereo vision technology can be compared to a human eye, which helps us to see the three-dimensional world [8]. This stereo vision technique helps to calculate the relative distance of the same impediments or the identical vertices in world reference (called disparity) as every visual can have a different perspective, and this can be processed further for depth estimation.

The main aim is to reckon the corresponding joint distribution based upon the computations calculated on a separate set of time instances. Each obstacle detection system has been classified into four prime models: (a) probabilistic occupancy map (b) digital elevation map, (c) scene flow segmentation and (d) geometry-based clusters. These models are structured, taking tessellation shapes and clustering strategies into consideration. In probabilistic occupancy maps, the image scene is a rigid grid of cells in which the outcome of the cells can either be free or occupied. It can lead to an undefined outcome if not mapped. Digital elevation maps (DEM) help in identifying the impediment depending on the scenario that they project from a ground surface. The scene flow segmentation helps in the utilization of the temporal correlation linking different instances of an image seized by stereo cameras to differentiate impediments that are in moving state [9], [10], [11]. This technique thus helps in trailing obstacles that are in motion. The geometry-based clustering allows categorization relying on geometric composition of point clouds present in the 3D environment. This method was tested in a real-time environment, as mentioned in [12] with a 10 Hz update rate. The detection of obstacles that is convenient for the situation is based on a search method that groups the point vertices using a double cone model, which plays a significant role in obstacle detection.

Number of people (in thousands) blind, with low vision, visually impaired per million

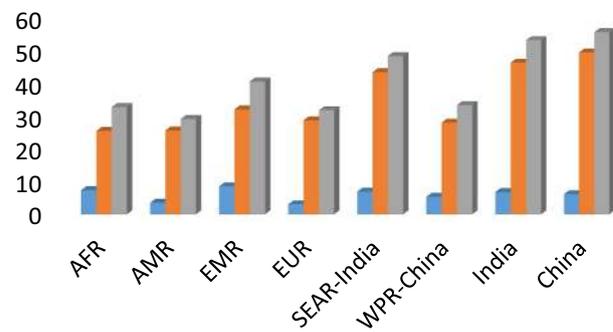


Fig. 1 Population of blind people in the world, with low vision, visually impaired per million

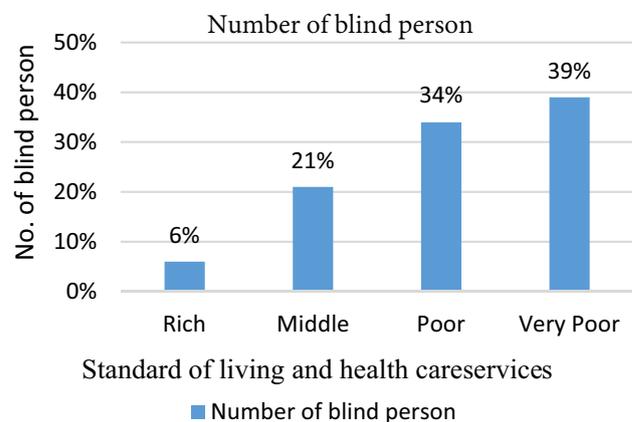


Fig. 2 Population of blind people according to their financial status

3 Problem statement

The main problem faced by many blind people is to perform day-to-day tasks independently without any help. Blind people are always in need of some assistive help from their family members.

A survey conducted by WHO shows statistics of the visually impaired people of all age groups which is estimated to be 284 million, of whom 38 million are blind [3]. Figure 1 represents the data analysis of the number of visually impaired individuals in the world. There are several visual aid devices but are built using expensive sensors and inefficient obstacle detection algorithms for interpretation of the surroundings.

It is interpreted that the population of blind people in the category of poor people is more. As visual aid devices are expensive, these ETA devices are not affordable to poor people. Figure 2 represents the statistics of

the population of blind people in different categories based on financial status. The graph represents that the percentage of blind people in poor and very poor categories is very high, which concludes that healthcare services are inaccessible, and visual aid devices are unaffordable for poor people.

There are many commonly used sensory visual aid devices for visually impaired people [13]. Many ETA devices primarily use expensive sensors for their functioning. Currently, there is no existing technology that aims to bring a solution to satisfy all the needs of visually impaired people. All the current technologies are designed only to perform specific tasks.

Also, Table 1 shows the commonly used sensory aids for the visually impaired. The table also classifies the different sensory aids based on their cost, accuracy and limitations. The main aim of our proposed system is to develop a visual aid device using a stereo vision and geometry-based obstacle detection algorithm that can fulfill the shortcomings of all the existing visual aid systems. This proposed system fulfills all the basic needs of a blind person and is economically viable.

4 Proposed system

In this section, the image processing and microcontroller(Arduino)-based visual aid model for blind people is discussed. The architecture is provided for the proposed system. While proposing the architecture, some basic specifications were taken into consideration like the speed that signifies that the system should be able to operate in real-time and notify the user about the impediments ahead, robustness indicates that the system’s performance should not be dependent on the lighting condition, distance indicates that object detection can be applied on

impediments located at any distance, object type signifies that the system should be able to detect any obstacle even the one’s present underground like potholes, etc., friendliness and cost-effective indicates that the system should be light, easy to use and economically viable.

The system architecture consists of components like Arduino, CCD cameras that use stereo vision technology, ultrasonic sensors, Wi-Fi module, buzzer, speaker, voice module, GPS module, GSM module. The proposed algorithm for object recognition explains the usage of image processing techniques and utilization of output pin state of the Arduino board with the ATmega8 controller by tracing the recognized object. The algorithm is programmed in a MATLAB environment, which involves the amalgamation of various image processing algorithms to observe the physical properties of the object [14]. The combination of CCD cameras and camera adapter connected to Arduino with MATLAB support acts as an image acquisition hardware to develop a stereo vision system in real-time for object recognition and tracking. Arduino programming provides a platform for a hardware system with control signals which are used for object recognition and tracing in a real-time environment. The location of the object can be traced based on the axis coordinates defined in the MATLAB program (Fig. 3).

Here, the system architecture is discussed in this section, which describes the internal and external architecture of the proposed mechanism. The proposed system architecture is depicted in Fig. 4.

4.1 System architecture

The system architecture consists of both the internal and external architecture. The internal architecture consists of the module of distance calculation of the obstacles through stereo vision using image processing

Table 1 Commonly used sensory aids for visually impaired persons

Sensory aids	Cost	Accuracy	Applicability
Guide dog	High-cost	Low	Not found easily, proper care is required
Walking stick	Low-cost	Medium	Physical contact needed to detect the obstacle and is slow
Ultrasonic sensor	High-cost	Low	Inaccurate information can be implied through the result of signal emission or absorption
Camera	Moderate	Good	Proper fitting and installation is required
Infrared sensor	Low-cost	Low	Applicable only for shorter distances
LASER	High-cost	Medium	Sunlight sensitive
NavBelt and The guide cane	High-cost	Good	Detection of all obstacles is not possible through ultrasonic sensors under all situations and is slower than guide cane
CG system	Moderate	Medium	Not affordable for outdoor use
Auto-adaptive thresholding technique	Moderate	Low	Kinect shows less accuracy in short range
Assistive device using DEEP-SEE framework and CNN	Moderate	Medium	The framework cannot overcome the problem of recognizing descending stairs or holes in the ground using monoscopic vision

Fig. 3 System design flow

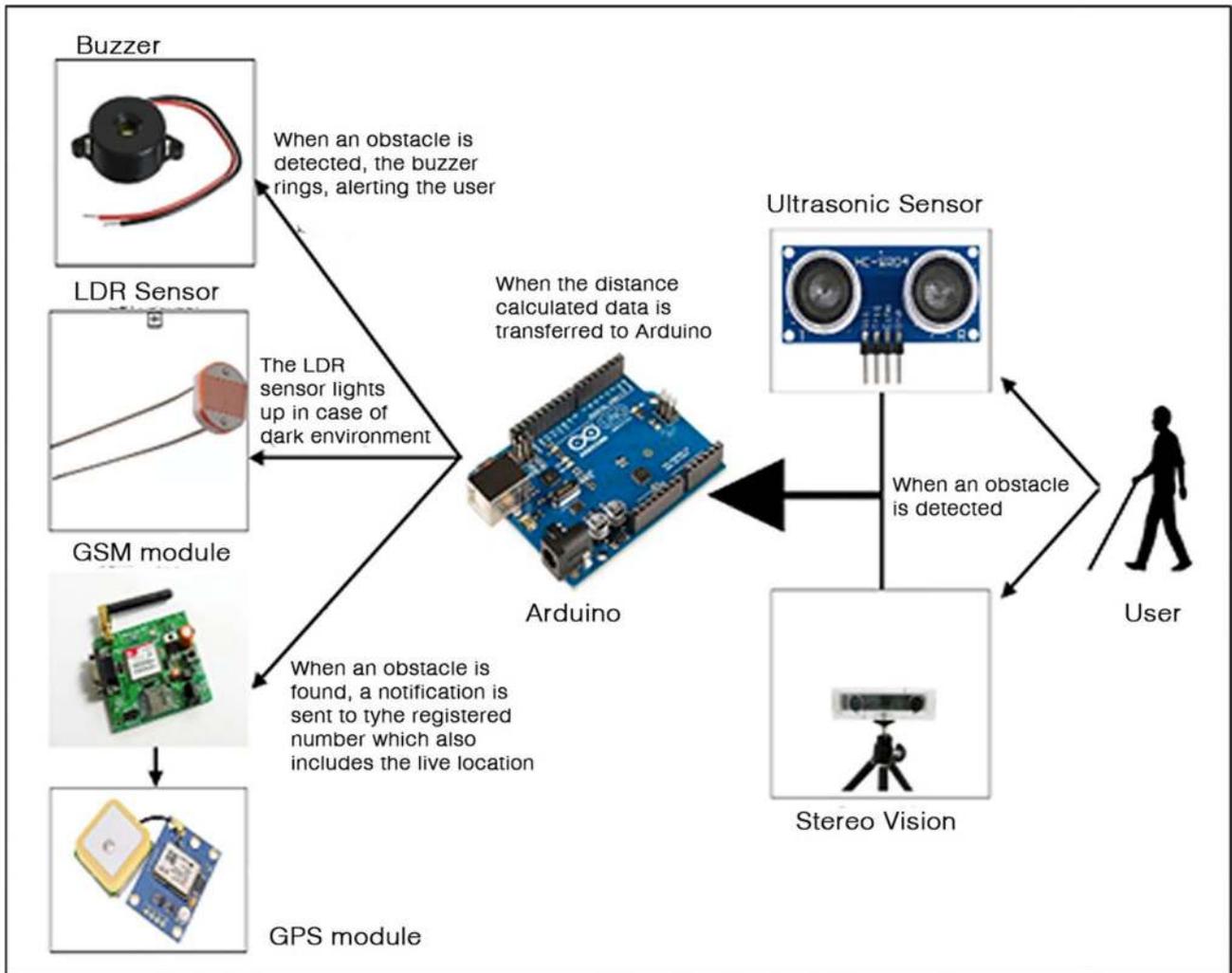
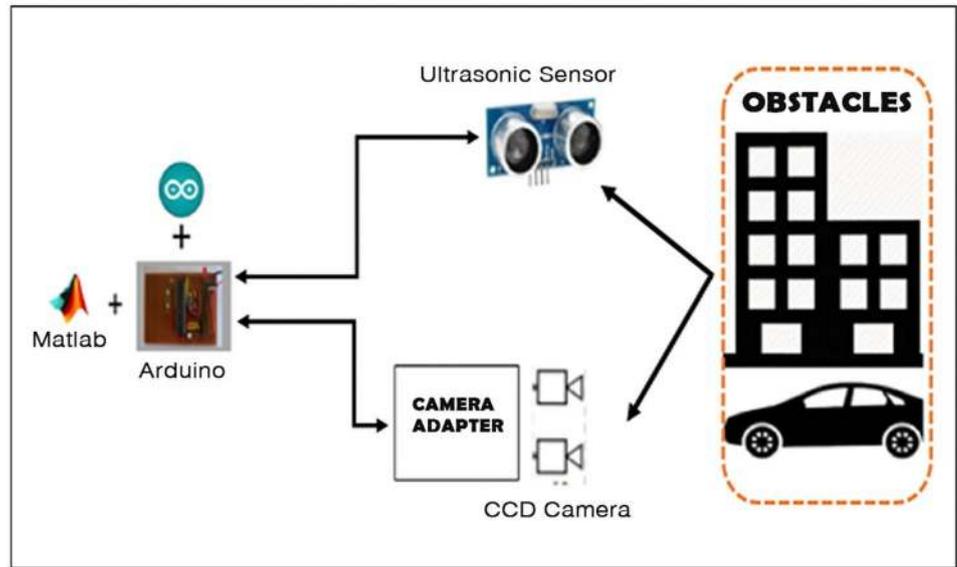
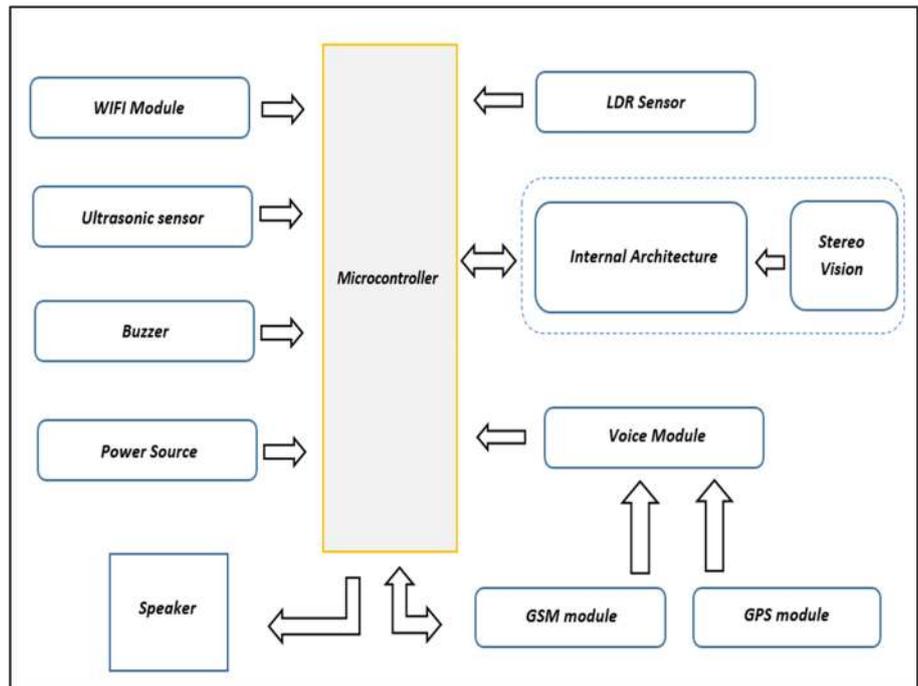


Fig. 4 System architecture

Fig. 5 External architecture

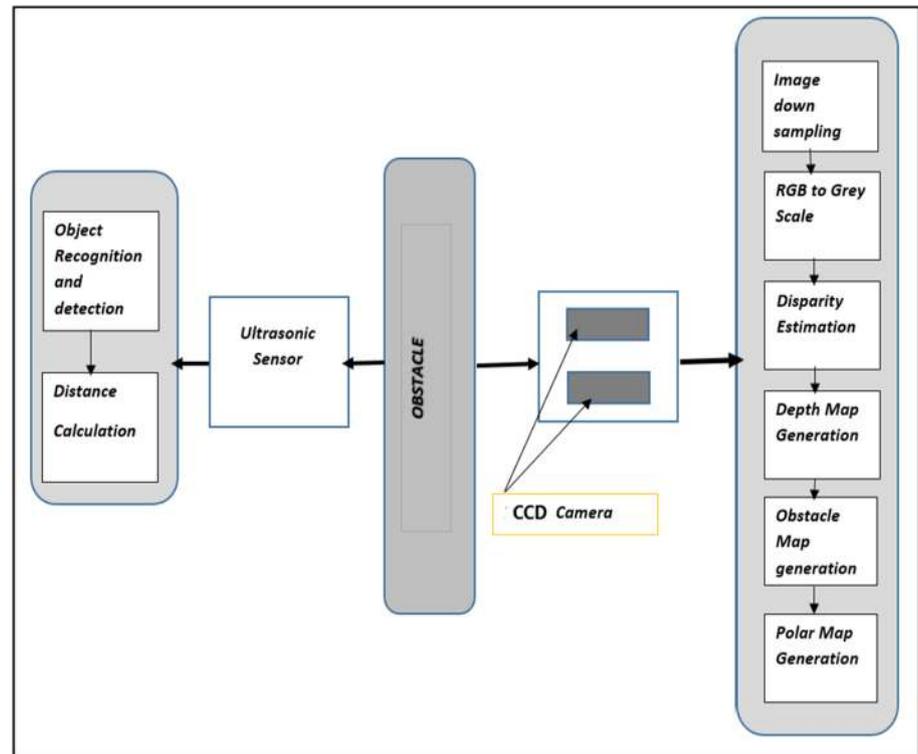


techniques like image down-sampling, red to greyscale conversion, disparity estimation, depth map generation, obstacle map generation and polar map generation. The internal architecture module is connected to the Arduino(ATmega8) module which is also connected with other components like ultrasonic sensors(Ultrasonic-HC-HR04), Wi-Fi module(ESP8266), GSM module(SIM9004), LDR sensor(LDR07), voice module, buzzer and GPS module(GPSNEO1) which together constitutes for the external architecture of the system.

4.2 External architecture (arduino and sensors)

In this section, the working methodology is depicted in Fig. 5. The system consists of a combination of both stereo vision and sensors. The calculation of the long distanced obstacles through stereo vision and short distanced using the ultrasonic sensor, a precise awareness of the obstacles, is obtained for real-time surroundings. The internal architecture is connected to the Arduino module, which is connected to other sensors like Wi-Fi module, GSM module, LDR sensor, voice module, buzzer and GPS module. When the system is turned on, farther distanced obstacles are detected using the stereo vision (which can be installed

on a helmet, connected to the Arduino module present on the walking stick through a wire) and ultrasonic sensor will detect the shorter distanced obstacles, to obtain a precise calculation of all the obstacles present in the surrounding. In case an obstacle is found, the buzzer will start to buzz, alerting the user about the obstacle to take necessary precautions. The user can also send an alert message/ notification with the help of the Wi-Fi module and GSM module. The Wi-Fi module sends the location of the user to the connected cloud. GSM module helps in sending a notification to the registered mobile number (which could be your emergency contact person) using a press of a button. The notification also provides the current location of the person using the GPS module. The buzzer and notification alert works simultaneously, making this an immediate response system. The GSM module also allows in receiving messages from the registered number. The message will be communicated to the user using a voice module. At times, the user might be traveling in a dark location where it would be difficult for people in the surroundings to identify in case of a disabled person walking by. Thus, the LDR sensor is activated that produces a source of light for better indication for the people around.

Fig. 6 Internal architecture

4.3 Internal architecture (stereo vision and image processing)

In this section, the internal architecture of the system is depicted by Fig. 6. In the internal architecture, we have used the image processing techniques to process the images captured by two vertically oriented CCD cameras. The captured image is down-sampled and resized for further processing. Then the RGB image is changed to a greyscale image, which is chopped again and resized for speeding up the process of disparity estimation. Then using the disparity map, the depth map is generated.

Obstacles are detected using the obstacle detection algorithm, and noise is removed from the obstacle map using median filtering and morphological closing operations. Then the polar map is generated for navigation planning. The ultrasonic sensors are utilized for obstacle detection and distance calculation at a short range to alert the blind person. The system proposed is more effective than any other system as stereo vision gives a more accurate distance calculation of the obstacles, especially for the farther ones. Also, CCD cameras are more cost-effective than a standard sensor.

5 Algorithm

Input: Video streams of snapshots from CCD camera

1. Initialization
2. **while** ultrasonic sensor is working **do**
3. **if** obstacle is at shorter distance
4. **detection** of obstacles by ultrasonic sensor
5. **Distance calculation**, $L = \frac{1}{2} \times (T \times C)$
6. **Get** alert through buzzer
7. **end if**
8. **end while**
9. **while** getting the video stream **do**
10. Image capture using CCD cameras
11. **Get** top/bottom image scene frame captured by CCD cameras
12. **Chopping** to the required field of view for Obstacle detection
13. **Scaling** down the images
14. **Transforming** RGB to grey scale images
15. **Disparity estimation** using optical flow software¹
16. **Generation** of Disparity map takes place
17. **Estimate** the depth
18.
$$\text{depth} = \sqrt{PA^2 + \left(\frac{l}{2}\right)^2 - 2 \cdot PA \cdot \frac{l}{2} \cdot \cos\left(\frac{\pi}{2} + \theta_1\right)}$$
19. **Applying** geometry based obstacle detection algorithm,
20. **Loop** through each pixel
21. **Categorize** all point vertices as non-obstacles.
22. **Review** the image pixels and inspect pixel P, in the scene.
23. **Notice** the cluster of picture pixels, TP, which is available on the protruded trapezium of P on the 2D picture surface
24. **Validate** all the vertices of TP and find set OP of vertices Pi in TP which is accordant with P.
25. **If** there are points in OP,
26. **then** categorize all points of OP as obstacle points
27. **formation** of Obstacle map takes place
28. **Using** median filter and morphological closing operations to remove noise
29. **Generation** of Polar map for navigation planning for system and obstacles at farther distances are detected.
30. **Planning** of the path for visually impaired person with respect to the polar map is done.
31. **end if**
32. **If** distance is less than a critical limit
33. **Get** location of the visual impaired person
34. notification sent to emergency contact person and visually impaired person is made alert
35. **end if**
36. **end while**

Output: object position, location of user, alert, path planning w.r.t polar map.

1

6 Software implementation

6.1 Flowchart

In this section, we have shown the work flow and methodology of the proposed system. The flowchart of the system is depicted in Fig. 7.

The flowchart pictures the sequence of activities and functionalities that take place in the proposed system. Once an obstacle is detected, using an ultrasonic sensor and CCD cameras. The system will calculate the distance of the obstacles from the visually impaired. Using disparity map generation and depth estimation in stereo vision, i.e., if the depth is 40 m or above, the buzzer will ring. A notification to the registered number will be delivered, alerting them regarding an emergency.

¹ <http://cs.brown.edu/people/mjblack/code.html>

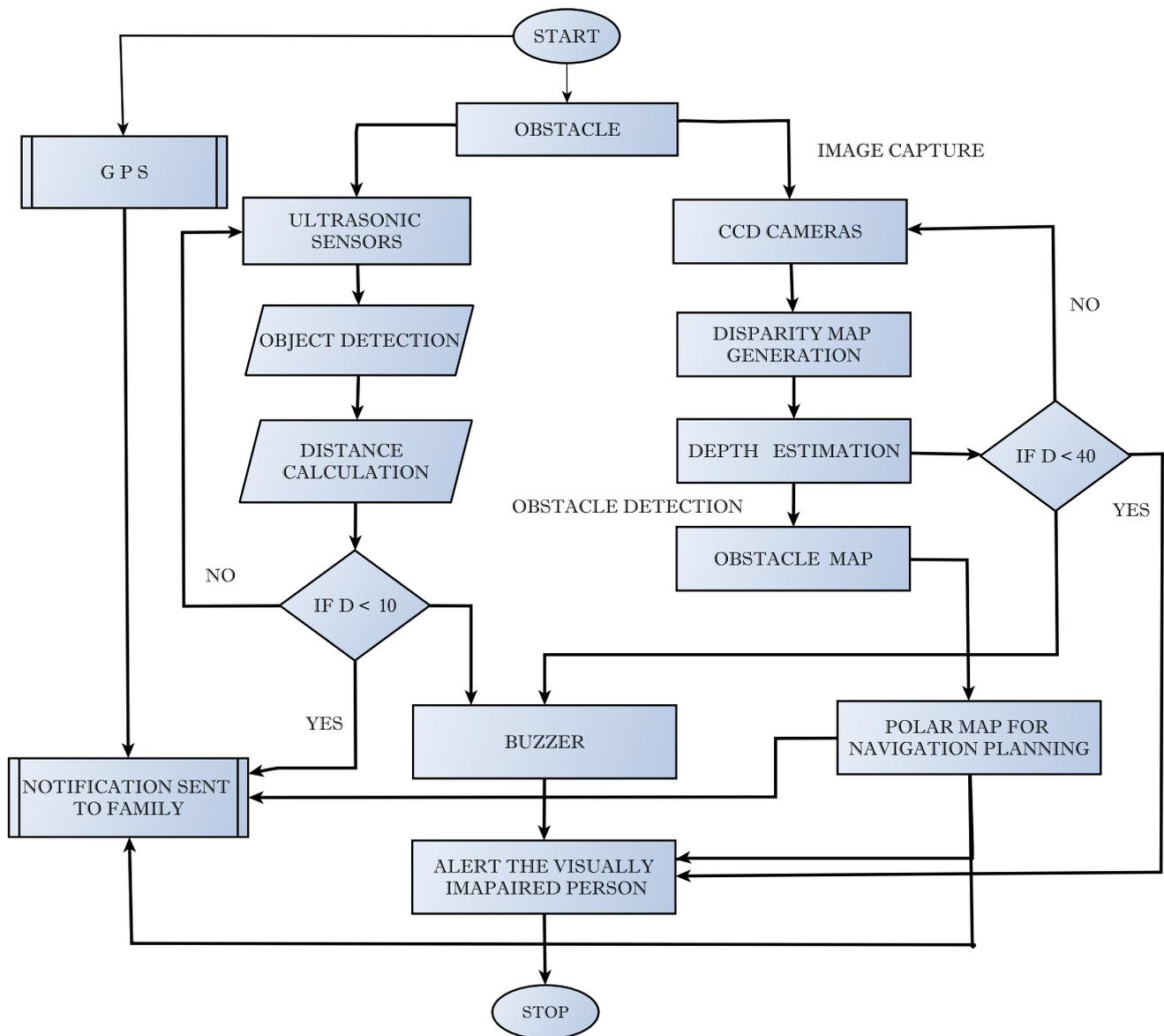


Fig. 7 Flowchart

Similarly, we can detect objects using ultrasonic sensors, i.e., if the distance is less than 10 m, the buzzer will ring to alert the user. As the user travels through its course, the live location coordinates of the blind person is updated on the cloud. A polar map helps in planning a safe navigation path for the visually impaired.

6.2 Obstacle detection using ultrasonic sensors

The ultrasonic sensor determines the distance of any obstacle using ultrasonic waves. The sensors emit ultrasonic waves toward obstacles ahead and are reflected to the sensor head. This emission and reception help in measuring the distance of the impediments from the sensor. A

single ultrasonic element is used for both emission and reception

6.2.1 Distance calculation

Ultrasonic sensors calculate the distance to the object by calculating the time between the emission and reception. The distance is calculated with the following formula:

$$\text{Distance } L = \frac{1}{2} \times (T \times C)$$

where L is the distance, T is the time between the emission and reception, and C is the sonic speed.



Fig. 8 Original image captured

6.3 Obstacle detection using stereo vision and depth maps

6.3.1 Image capture

The two vertically oriented CCD cameras perceive the images. The captured images are consistently sampled in azimuth and altitude angles. The altitude range is around 180° along the y-axis, and the Azimuth range is around 360° along the x-axis of the scene. The image captured by the camera embedded in the visual aid system is depicted in Fig. 8.

6.3.2 Disparity estimation

Disparity refers to the space between two corresponding pixels in the top and bottom image of a stereo image pair. The variation in the altitude angles $\Delta\theta$ between the two image scenes implies the disparity for the spherical image pair as the cameras are positioned vertically. Disparity values are evaluated in various ways [15]. Through these methods, the optical flow of all the point vertices in the image scene can be calculated. The optical flow is determined through the optical flow software² developed by [16]. The vertical direction of the optical flow is due to the vertically positioned CCD cameras. Disparity value in terms of pixels can be calculated by optical flow at each pixel in the image scene. The disparity map shows the visible pixel difference between a pair of stereo images. The disparity map is generated using optical flow that explains the narrow or thick vector field, where a displacement vector is allocated to a particular pixel point that focuses on where that pixel can be observed in another image. It helps in estimating the separation between the pixel points in the stereo image pair. By coordinating every pixel in the image with its corresponding pixel in another image, we can calculate the separation between the pixel points results in the generation of the disparity map.

Figure 9 shows the stages of the algorithm for the generation of disparity maps. The process flow represented

in the above image contains four stages. The matching cost computation is the stage determines whether the estimations of two-pixel points correspond to the same point in the image scene. So, this method is defined to determine the parallax estimations of each pixel point between the top and bottom images. The matching cost is calculated at each pixel for every pixel point under inspection. The variation in pixel intensity between a pair of the pixel point in both the top and bottom image of a stereo image pair is called the disparity and can be linked with depth estimations through three-dimensional projection. Cost aggregation is the phase of the algorithm that determines the performance of the disparity map algorithm, which serves the purpose of minimizing the dissimilarities. It is required as the information gathered for a single pixel point upon calculating the matching cost is not adequate for precise matching. Disparity computation and optimization is the phase for estimation of final disparity at each pixel of the image. Disparity map refinement is the stage that works on decreasing noise and improving the quality of disparity maps. This step consists of regularization and interpolation.

Down-sampling the original image will increase the disparity estimation step. The process requires an altitude which is ranging from $-\pi/6$ to $\pi/6$. The original image is further reduced to 767×129 pixels. The resulting grey-scaled image of the obstacle map has an increased resolution than that of the previous down-sampled version of the original image obtained from the obstacle detection algorithm. Figure 10 depicts the chopped and the down-sampled version of the original image (Fig. 11).

The disparity map is generated from optical flow estimation. More significant disparity and smaller depth will give brighter pixels. The uniform gradient of disparity along the ground is observed.

6.3.3 Depth estimation

The depth map is attained by implementing basic geometry and trigonometry to the disparity map.

$$\frac{l}{\sin \Delta\theta} = \frac{PA}{\sin \left(\frac{\pi}{2} + \theta_1 \right)} \quad (1)$$

The depth is calculated as the parameter in the cone diagram referenced to the above camera center, which constitutes l as surface distance and $\Delta\theta$ as disparity value. θ_1 represents the altitude angle of the 3D-vertex point, according to the above camera. PA is the measured value from the upper camera to the point in the 3-D space.

² <http://cs.brown.edu/people/mjblack/code.html>

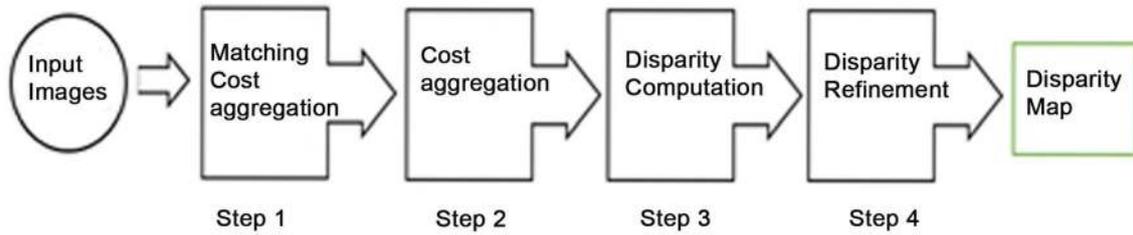


Fig. 9 Stages of algorithm for generation of disparity map



Fig. 10 Image chopped and down-sampled



Fig. 11 Disparity map

$$\text{Depth} = \sqrt{PA^2 + \left(\frac{l}{2}\right)^2 - 2.PA.\frac{l}{2}.\cos\left(\frac{\pi}{2} + \theta_1\right)} \quad (2)$$

Using the formula of Eq. 2, we can obtain the depth value for each pixel from the disparity value corresponding to that pixel. The depth map is generated from the disparity map. Pixels having a darker shade represent a smaller depth. The disparity value of every pixel present in the image scene is used to attain the depth value of the corresponding pixels (Fig. 12).

6.3.4 Object detection

To detect obstacles, we have proposed the geometry-based clustering technique as an appropriate technique. Before we see the obstacle detection algorithm, it is necessary to envision obstacles in the 3D space. To work on this technique, we first inspect essential obstacle points. Thus, impediments are vertices in the image scene, which are at an altitude from the baseline.

Obstacles can be represented mathematically in terms of two separate point vertices in the subsequent way: Point vertices P1 and P2 which lie on the same impediment are said to be consistent if



Fig. 12 Depth map

$$(1) H_T < |P2Z - P1Z| < H_{max}$$

The deviation between the two elevated points can be determined by the scope of H_T and H_{max}

$$(2) \left((P2 - P1) \cdot \frac{P3 - P1}{||P2 - P1|| ||P3 - P1||} \right) > \cos\theta_T$$

By displacing P1 through H_{max} in the z-direction, point P3 is obtained. This state implies that the angle formed by the point P1, P2 wrt the z-direction is lower compared to the threshold value.

H_T and H_{max} are the minimum and maximum height, respectively, of an object which is observed as obstacle. θ_T can be estimated according to the acquired traversable slope which helps in the categorization of points as obstacles or non-obstacles.

The definition is illustrated in Fig. 13. The upward cone is structured based on H_T, H_{max} and θ_T values in the 3-D space. The points lying in the frustum are determined as obstacle points.

The generated depth map conveys the information regarding the depth of all the pixels. All the pixels are currently three-dimensional point vertices in Azimuth altitude-depth space, which is converted to the X-Y-Z space. The naïve algorithm (which analyses the point pairs having $O(N^2)$ complexity) is more effective and differentiates a point within a group of points. The group of points restrained within the structure of the trapezium created by the formation of the cone above the point vertex present in three-dimensional space. The trapezium is formed based on the depth from the base point. The set of parameters is as follows

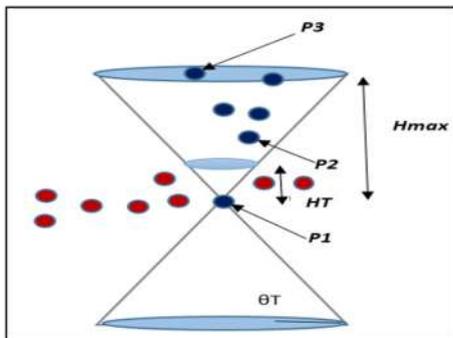


Fig. 13 Cone formed in 3D space

$$h_t = \frac{H}{\pi} \cdot \frac{H_T}{\text{depth}}$$

where H denotes the height of the base. The height to the altitude angle is transformed by a measuring factor of H/π ; therefore, it indicates the number of pixels. Likewise,

$$h_{\max} = \frac{H}{\pi} \cdot \frac{H_{\max}}{\text{depth}}$$

H_T and H_{\max} represent the altitude of the nearer parallel edge of the trapezium from the base vertex and altitude of the distant parallel edge from the base vertex, which is the threshold angle that represents the top left angle in the trapezium.

By looping through the vertices within the structure of trapezium, it forms a key pair for every pixel with the base vertex. Key pair points formed can be congruent if they satisfy the grounds of being an obstacle, then we can imply that the point vertex present in the trapezium is an obstacle. This algorithm is more efficient than the Naive algorithm in terms of time complexity. Let the number of points taken at an average in the trapezium represented as M. The complexity is $O(MN)$.

Algorithm:

- Categorize the point vertices as non-obstacles.
- Review the image pixels and inspect pixel P in the scene.
 - Notice the cluster of picture pixels, PT, which is available on the protruded trapezium of P on the 2D picture surface
 - Validate all the vertices of TP and find set OP of vertices Pi in TP, which is accordant with P.
 - If there are points in OP, then categorize all points of OP as obstacle points.



Fig. 14 Grey-scaled images of the obstacle map

6.3.5 Post-processing

In certain regions, we obtained fewer feature points, which resulted in the absence of texture on the horizontal surface, providing less accurate disparity values. The output results in noise as the disparity gradient is varied and not continuous in those regions. We have applied the obstacle detection algorithm on the output, which results in the closing of the small holes using median filtering and morphological closing operation. Figure 15 depicts the obstacle map after median filtering to remove noise and fill in the holes for the grey-scaled image, as shown in Fig. 14.

6.3.6 Polar map

The objective of the system is to detect impediments making it easier for the movement of the visually impaired people. Obstacle points should be represented in such a way that they might be useful for motion planning. A polar map depiction shows the projection of the barrier points, present in the three-dimensional visual space, onto a horizontal surface plane. In Fig. 16, the blind person is at the center of the polar map. The region represented in blue color is the obstacle points with the radius defining the depth for the obstacle. The tiny spot present near the center of the polar map, i.e., at 0° , depicts the white mark board, and the spot near 30° depicts the black area, shown in Fig. 10. The buildings are shown at a far of stretch in the map. With the help of the polar map, we can decide the directions through which any impediments will not block the motion of the visually impaired people. A polar map can provide a safe path of navigation path for the visually impaired.

7 Hardware implementation

Camera plays a significant role in the image acquisition phase of the hardware system. The CCD camera, along with the camera adapter, captures the dynamics of the image scene. While working on this system, the right hardware choice and execution of the object recognition algorithm was a zealous task. The system uses ATmega 8 microcontroller, which is programmed using the Arduino



Fig. 15 Grey-scaled images of the obstacle after post-processing

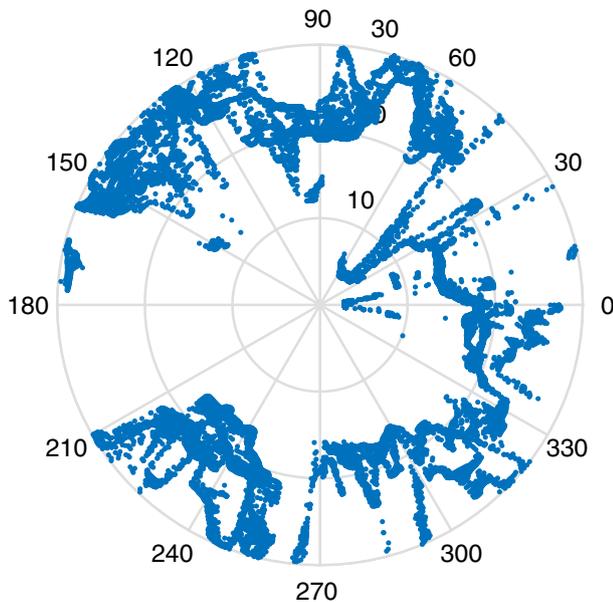


Fig. 16 Polar map representation of obstacles

IDE. Details regarding the hardware components and their purposes are described below.

7.1 Image acquisition device

The use of high-resolution CCD cameras ensures the effectiveness of the system. Generally, CCD cameras are used as an imaging device. These cameras are oriented to take multiple snapshots of a particular scene with a predefined triggering rate. Since these snapshots are taken in a fraction of time, they appear to look like a video.

7.2 Arduino board AT mega 8

Arduino is a microcontroller board based on the ATmega8. It allows structuring the programs in segments of code to perform individual tasks. Arduino microcontrollers are pre-programmed with a boot loader that simplifies uploading of programs to the on-chip flash memory. Its functionality can be enhanced by adding output drivers, often for use in school-level education, to simplify making buggies and small robots (Figs. 17, 18).

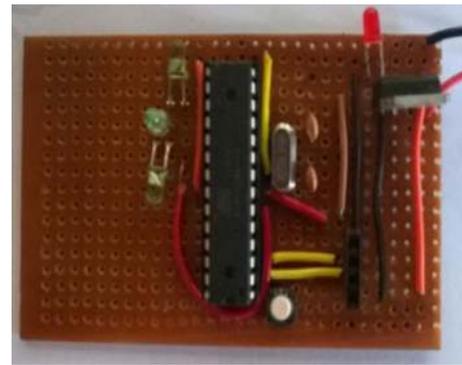


Fig. 17 Arduino ATmega8

7.3 Ultrasonic sensor (ultrasonic-HC-HR04)

The ultrasonic sensors consist of three categories: transmitters, receivers and transceivers. Transmitters convert electrical signals into ultrasound, and receivers convert ultrasound into electrical signals. The transceivers can both transmit and receive ultrasound. The sensor is connected to the Arduino for detecting the obstacles which are within a radius of 8 m (Fig. 19).

7.4 GSM module (SIM9004)

A GSM module is a device used for primary communication purposes using a GSM and GPRS network. A GSM module works on a communication network that can be activated using a SIM card. This module is connected to the Arduino to send and receive SMS and for the use of making calls. The GSM module has the following components - GSM modem connected to a power supply, communication interfaces for a computer (Fig. 20).

7.5 GPS Module (GPSNE01)

The GPS module helps in finding the geographical location of a person using the information received from the GPS satellite. GPS was initially established for the usage of the US military, but later, in the 1980s, the government permitted the usage of GPS to the ordinary people. The GPS module is connected to the Arduino, which sends the location of the user to the registered mobile number to keep track of the user's mobility (Fig. 21).

7.6 Wi-Fi module (ESP8266)

The ESP8266 Arduino compatible module is a low-cost Wi-Fi chip with full TCP/IP capability, and the fantastic thing is that this little board has an MCU (Micro Controller

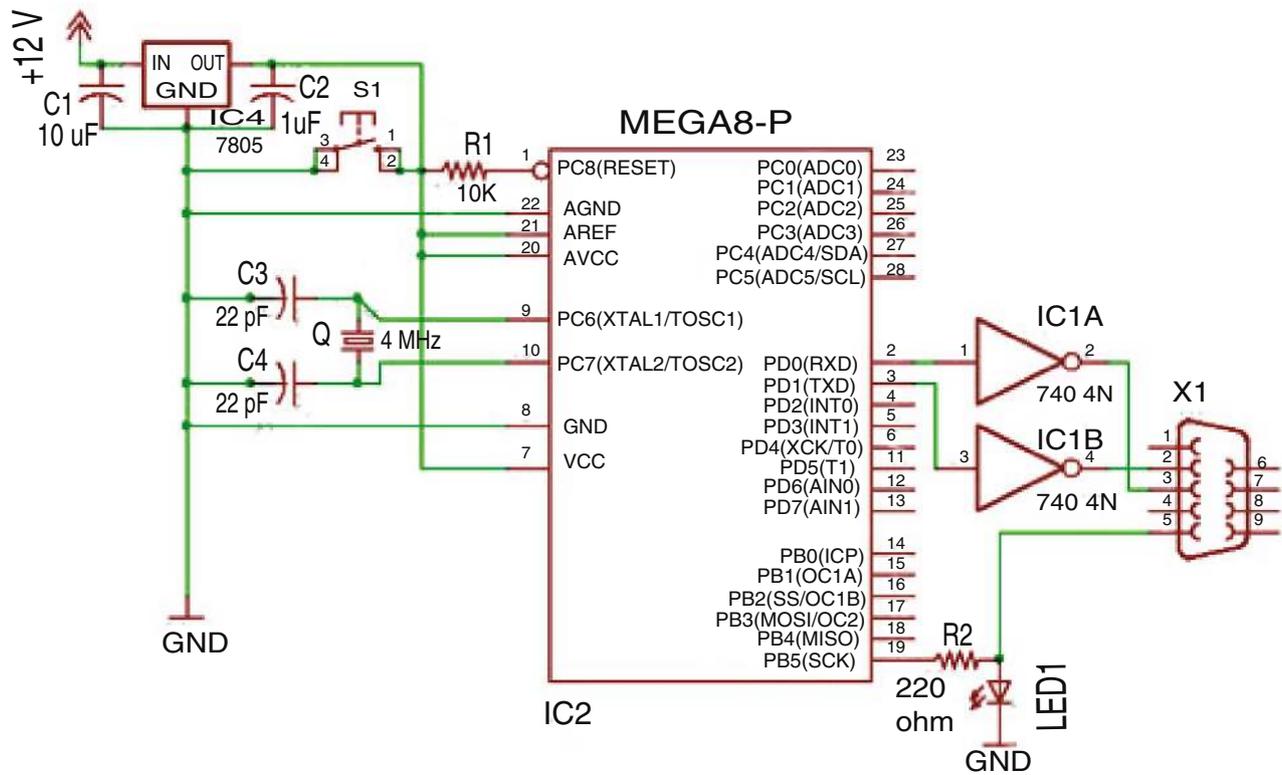


Fig. 18 Pin diagram of Arduino Atmega8



Fig. 19 Ultrasonic sensor

Unit) integrated, which gives the possibility to control I/O digital pins via simple and almost pseudo-code like programming language. The Wi-Fi module sends the data to the AWS cloud (Fig. 22).

7.7 LDR sensor (LDR07)

A photo-resistor (LDR) is a light-controlled variable resistor. It exhibits photoconductivity. In the dark, a photo-resistor



Fig. 20 GSM module

can have a resistance as high as several mega ohms (MΩ), while in the light, it can have a resistance as low as a few hundred ohms. The sensor contributes to the robustness factor of the system. The sensor emits light to alert the people in the course of the navigation (Fig. 23).



Fig. 21 GPS module



Fig. 23 LDR sensor

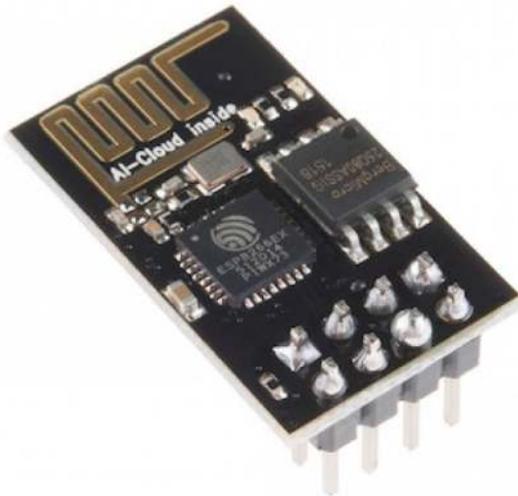


Fig. 22 Wi-Fi module



Fig. 24 Buzzer

7.8 Buzzer

A buzzer is an audio producing device for alerting the people in danger, for example, during the crisis of a fire. The buzzer connected to the Arduino alerts the visually impaired people in case of approaching an obstacle (Fig. 24).

7.9 Camera adapter

A camera adapter provides a high-resolution camera image. A camera adapter can support multiple cameras to work simultaneously. In our system, we have used the Arducam 4 Cameras Adapter Board and Robocraze OV7670 300KP VGA Camera. The adapter acts as a connector between MATLAB and Arduino board for communication using COM port (Fig. 25).

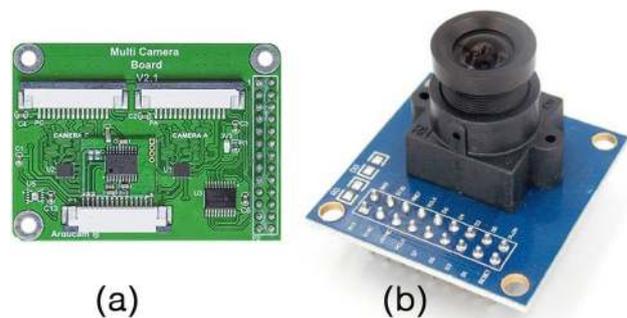
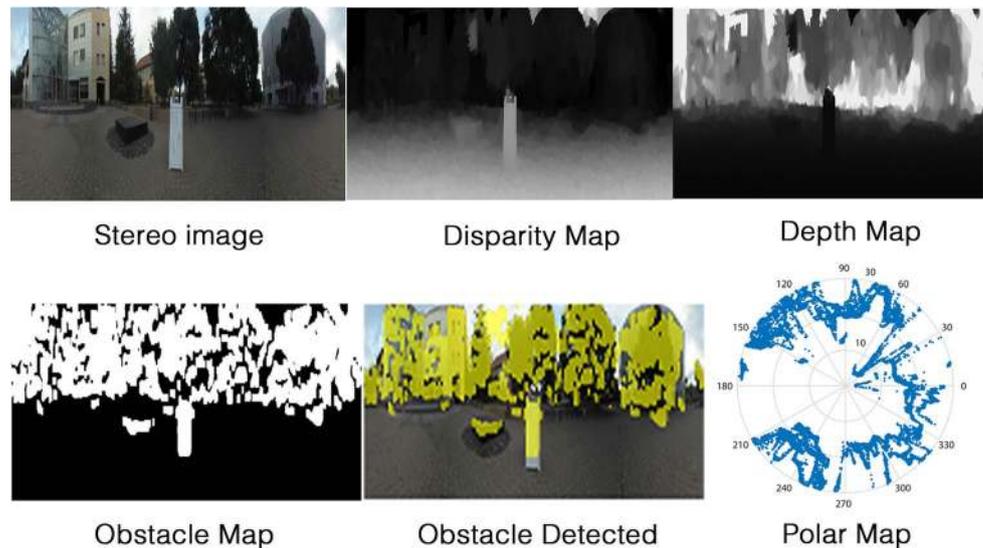


Fig. 25 a Camera adapter board b CCD camera

8 Results and conclusion

Figure 26 represents an image scene having impediments at both shorter and longer range in the field of view. As the user is traveling through the course, the ultrasonic sensor continuously detects the obstacles within the scope of

Fig. 26 Test scene results for obstacle detection algorithm



8 m. Simultaneously the series of snapshots captured by the CCD cameras are processed using image processing techniques. The obstacle detection algorithm is applied to the processed image for the generation of the disparity map and depth estimation. This process leads to the generation of a polar map that helps to plan the navigation path. The ultrasonic sensor detects all the obstacles at a shorter range while the obstacle detection algorithm discovers all the impediments in the field of view of the image scene.

After testing the process outdoors, we noticed that the obstacles in the image scene are precisely pointed in the polar map. The accuracy of the results of the algorithm is based on the efficiency of disparity estimation. The internal system is programmed in MATLAB. The technique also works fine with the potholes that are neglected as the point vertices around the holes present in three-dimensional space are considered as impediments. These impediments are depicted in the polar map for deciding on the direction to navigate through the space. Figure 26 shows the object detection framework using stereo vision.

In this paper, we are trying to incorporate stereo vision and image processing techniques in real-time application—aiding the visually impaired with a more reliable navigation device. Image processing has always played a significant role in making the lives of disabled people more comfortable and independent. In this paper, we have tried to incorporate not only standard sensors like ultrasonic sensors into a blind man's walking stick but also combine it with stereo vision. The geometry-based obstacle detection algorithm has been able to fulfill the shortcoming of assistive devices using an adaptive thresholding technique and DEEP-SEE framework. Accurate results are obtained for every object present around

the blind person's surroundings using these conventional methods. Alert notification and Buzzer alert has made this more applicable for the real-time world, as an immediate response is received. The blind person gets immediate help in case of any emergency.

9 Further scope

The work is continued to beat the issues faced because of obstruction and irregular matching of stereo images. We are planning to extend our work in making the system suitable to tackle any situation for the benefits of the user. Apart from this, we plan to integrate the google maps and polar map for efficient motion planning.

Acknowledgements We want to thank the Vellore Institute of Technology, Vellore, for giving us this opportunity and support throughout this work. We wish to extend our sincere gratitude to our professor Prof. Dr. Natarajan P, and all other faculties of CSE4001 for the knowledge and information that was taught and shared in class.

Compliance with ethical standards

Conflict of interest On behalf of all authors, the corresponding author states that there is no conflict of interest.

References

1. Bourne RRA, Flaxman SR, Braithwaite T, Cicinelli MV, Das A, Jonas JB et al (2017) Vision loss expert group. Magnitude, temporal trends, and projections of the global prevalence of blindness and distance and near vision impairment: a systematic review and meta-analysis. *Lancet Glob Health* 5(9):e888–e897
2. Fricke TR, Tahhan N, Resnikoff S, Papas E, Burnett A, Ho SM, Naduvilath T, Naidoo KS (2018) Global prevalence of presbyopia

- and vision impairment from uncorrected presbyopia: systematic review, meta-analysis, and modelling. *Ophthalmology* 125:1492–1499
3. Pascolini D, Mariotti SPM (2011) Global estimates of visual impairment: 2010. *British J Ophthalmol Online* First published December 1, 2011, as <https://doi.org/10.1136/bjophthalmol-2011-300539>
 4. Koley S, Mishra R (2012) "Voice operated outdoor navigation system for visually impaired persons". *Int J Eng Trends Technol (IJETT)*. V3(2):153–157 Mar–Apr 2012. ISSN:2231-5381. www.ijettjournal.org. published by seventh sense research group
 5. Shoval Shraga, Ulrich Iwan, Borenstein Johann (2001) Computerized obstacle avoidance systems for the blind and visually impaired. *J Logic Comput LOGCOM*. <https://doi.org/10.1201/9781420042122.ch14>
 6. Demby's J, Gao Y, Shafiekhani A, Desouza GN (2019) Object detection and pose estimation using CNN in embedded hardware for assistive technology. <https://doi.org/10.1109/ssci44817.2019.9002767>
 7. Mocanu, Bogdan & Tapu, Ruxandra & Zaharia, Titus. (2018). DEEP-SEE FACE: a mobile face recognition system dedicated to visually impaired people. *IEEE Access*. PP. 1-1. <https://doi.org/10.1109/access.2018.2870334>
 8. Brown MZ, Burschka Darius, Hager Gregory (2003) Advances in computational stereo. *IEEE Trans Pattern Anal Mach Intell* 25:993–1008. <https://doi.org/10.1109/TPAMI.2003.1217603>
 9. Wedel A, Meißner A, Rabe C, Franke U, Cremers D (2009). Detection and segmentation of independently moving objects from dense scene flow. In: *Lecture notes in computer science*. vol 14–27. 10.1007/978-3-642-03641-5_2.
 10. Franke Uwe, Rabe Clemens, Badino Hernán, Gehrig Stefan (2005) 6D-Vision: Fusion of Stereo and Motion for Robust Environment Perception. *Lect Notes Comput Sci* 3663:216–223. https://doi.org/10.1007/11550518_27
 11. Lenz P, Ziegler J Geiger A, Roser M (2011) Sparse scene flow segmentation for moving object detection in urban environments. In: *IEEE intelligent vehicles symposium, proceedings*. 926–932. 10.1109/IVS.2011.5940558.
 12. Talukder A, Manduchi R, Rankin A, Matthies L (2002) Fast and reliable obstacle detection and segmentation for cross-country navigation. 2:610–618. <https://doi.org/10.1109/ivs.2002.1188019>
 13. Kanwal Nadia (2015) A navigation system for visually impaired: a fusion of vision and depth sensor. *Appl Bionics Biomech*. <https://doi.org/10.1155/2015/479857>
 14. Ayi M, Ganti AK, Adimulam M, Karthik B, Banam M, Kumari GV (2017) Interfacing of MATLAB with Arduino for face detection and tracking algorithm using serial communication. 944–948 <https://doi.org/10.1109/icici.2017.8365276>
 15. Hamzah Rostam, Ibrahim Haidi (2016) Literature survey on stereo vision disparity map algorithms. *J Sens* 2016:1–23. <https://doi.org/10.1155/2016/8742920>
 16. Sun D, Roth S, Black M (2010) Secrets of optical flow estimation and their principles. In: *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Francisco, CA*, pp 2432–2439. <https://doi.org/10.1109/CVPR.2010.5539939>
 17. Oniga Florin, Nedeveschi Sergiu (2010) Processing dense stereo data using elevation maps: road surface, traffic isle, and obstacle detection. *IEEE Trans Veh Technol* 59:1172–1182. <https://doi.org/10.1109/TVT.2009.2039718>
 18. Elmannai Wafa, Elleithy Khaled (2017) Sensor-based assistive devices for visually-impaired people: current status, challenges, and future directions. *Sensors* 17:565. <https://doi.org/10.3390/s17030565>
- Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.