An Intelligent Virtual Assistant using Raspberry Pi

Narendar Singh.D, Asst.Prof, Dept. of ECE,Anurag University,Hyderabad

Pavitra.B Asst.Prof, Dept. of ECE,Anurag University,Hyderabad

Nagaswetha.R, Asst.Prof, Dept. of ECE,Anurag University,Hyderabad

Anil Kumar G, Asst.Prof, Dept. of ECE,Anurag University,Hyderabad

Ashwini.G Asst.Prof, Dept. of ECE,Anurag University,Hyderabad

Article Info
Page Number: 1261-1270
Publication Issue:
Vol. 71 No. 3 (2022)

Article History
Article Received: 12 January 2022
Revised: 25 Febuary 2022
Accepted: 20 April 2022
Publication: 09 June 2022

Abstract
Recent breakthroughs in speech recognition have gained attention. It was once science fantasy to conceive talking to a computer. Things change quickly. More people are utilising smart phones to send e-mail, text messages, and seek up information online. Our project involves one such application. This project's purpose is to create an Intelligent Virtual Assistant (IVA) that can do several tasks for a user. All of these functionalities need human input, location awareness, and internet data access (such as weather forecasts and news articles; stock prices and quotes; user schedules; telling the time; travel assistants; calendar events; notifications from social media applications; and so on). This device uses Raspberry Pi and a user's voice as input. Use speech-to-text engines to write spoken words. This procedure's text was used to process queries and obtains data. After obtaining the information, it was converted to voice using text-to-speech and sent to the user. Several modules used keyword matching. It gives visually impaired people speech access to Wikipedia, Calculator, etc. This model may be used as a surveillance system that compares a visitor's speech to a pre-recorded one. Also, blind/visually impaired people may enjoy it. This paradigm provides an automated system that interacts with IOT (Internet of Things). The results of much effort were well-documented.

A. Introduction

For people who seem to have physical limitations on the outside, driving in crowded areas is a challenge. Because they can't access the logical and geographical info that's going on around them, those with visual impairments are at a distinct disadvantage. Approximately 39 million of the 285 million persons worldwide who showed outward symptoms of disability in 2010 had visual impairments, according to a research. Visibly impaired people should have access to a practical application framework that allows them to roam freely, regardless of the environment they are in, known or unfamiliar. People in today's fast-paced society have benefited greatly from the aid and support that may be supplied through technology. There are now a broad range of technologies that can aid people with a wide range of abilities in the activities they do in every day. The great majority of devices made in this business are either not in use or need a large amount of setup before they can be utilized, which is a problem. As an example, consider what follows as an explanation for some recent work on the subject: People who are physically or mentally disadvantaged may use DRISHTI's remote passerby

route structure. In order to help people who are limited by external conditions, it emphasizes increasing logical awareness. As a consequence of all of the labor that went into incorporating this idea, it was unable to develop the components to their full potential. A remote PC, cameras, common language processor, amplifier, go sensors, GPS sensors, content to-discourse gadget, and so on were all included into the TYFLOSframework to provide for more flexibility when travelling or reading. To its detriment, since no one with vision impairments was ever subjected to this methodology for testing, no input could have been obtained to assist improve the equipment and programming reconciliation.

Using a robot obstacle shirking device, the NAVBELT navigational assistance is used. Headphones, a personal computer, and ultrasonic range sensors were used in the prototype[8]. One of the system's major drawbacks is that it only accepts sound input. This makes it difficult for consumers to use the system. Customers also have to put in a lot of effort to be able to make advantage of this framework.

For many of the newly formed firms, it is necessary to keep a continuous network running, which is not only challenging in certain places but also expensive due to the additional costs of information consumption. Proximity sensors, infrared sensors, laser diodes, and the like are used by these companies. Some of these sensors may not work effectively in the open air due to their sensitivity to the elements such as sunshine, rain and dust[9].

These goods are also a pain to wear, lack adaptability and are prohibitively pricey, making them impossible to purchase for the average consumer. An innovative technology was created to help people with visual impairments interact with others because of all of these qualities. The device is simple to use, practical, and affordable, and it doesn't need any other resources like the internet to operate. Being aware of the obstacles in a way that isn't too demanding. Section II of this piece of literature explains the framework's construction in detail. The process of calculating the deterrent and measuring the distance of separation takes place one point at a time in Area III. Detailed instructions on how to use the electronic gadget may be found in Section IV of this manual. Next, we'll discuss what we've learned from our experiments, what those findings mean, and what further needs to be done moving forward.

B. Literature review

Virtual assistants are application programmes that interpret natural language voice instructions and accomplish tasks for the user.[1]

Virtual assistants execute basic tasks for end users, such as adding tasks to a calendar, searching the web, or managing smart home devices like lights, cameras, and thermostats.[2]

Virtual assistants may make and receive phone calls, compose text messages, acquire directions, hear news and weather reports, locate hotels or restaurants, verify travel bookings, play music, and more.[3] Such activities include taking dictation, reading text or email messages aloud, searching up phone numbers, scheduling, placing calls, and reminding the end user about appointments[6]. Any automation system or artificial system aims to minimise human work, effort, time, and carelessness. Virtual assistants need large volumes of data to

support AI systems like machine learning, NLP, and voice recognition[5]. As a user interacts with a virtual assistant, AI programming learns from data input and predicts human demands[4].Now, we want to combine all these features into a single, user-friendly system.

Existing systems

Existing systems lack customizability, thus users can't access complete information. These systems don't employ text-to-speech. [10] The output is text, which may be difficult for visually handicapped users, and display devices may increase system cost. Existing systems don't use open source platforms. They lack future potential.

C. **Proposed System**

The suggested system overcomes current system flaws. The project uses text-to-speech. After the command, the system's input will be output as voice. The project's open source platforms allow significant customizability. The system is very adaptable, so new modules may be added without affecting existing ones.

Block Diagram:

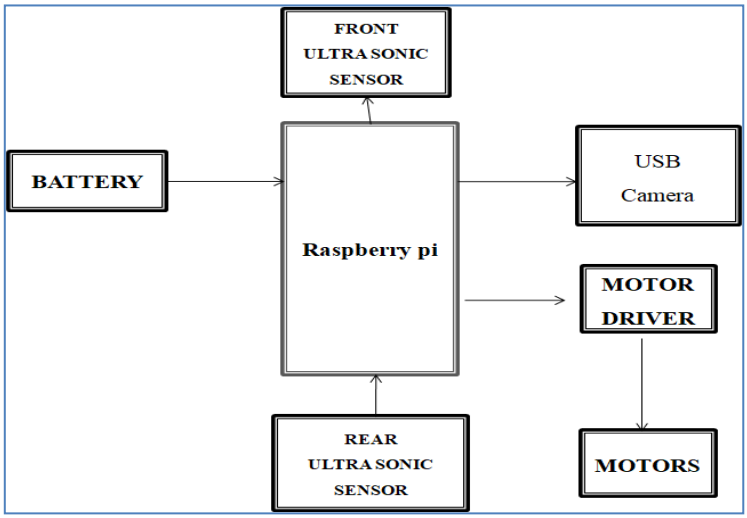

Figure 1.1: Design Structure

Description Block diagram has five sections.

1.Microphone

2. Pi-Board

3. Power Supply

4. Internet

5. Speakers

Microphones record sound. This audio input is searched for keywords. These keywords are vital for the voice command system since our modules operate by matching keywords.

Pi-Board

Raspberry Pi processes data and connects components for the voice command system. Raspbian OS is installed onto an SD card, which is then inserted into the card slot.

Power Supply

Raspberry Pi requires 5V, 2.1 to 2.5 mA. This may be done using an AC adapter or a power bank.

Voice command system uses the internet. Since the system uses online text-to-speech, query processing, and speech-to-text conversion, we require a persistent connection.

Speakers

After a user's inquiry is processed, the text result is translated to voice online. This speech is conveyed to the user utilising audio out speakers.

Schematic diagram

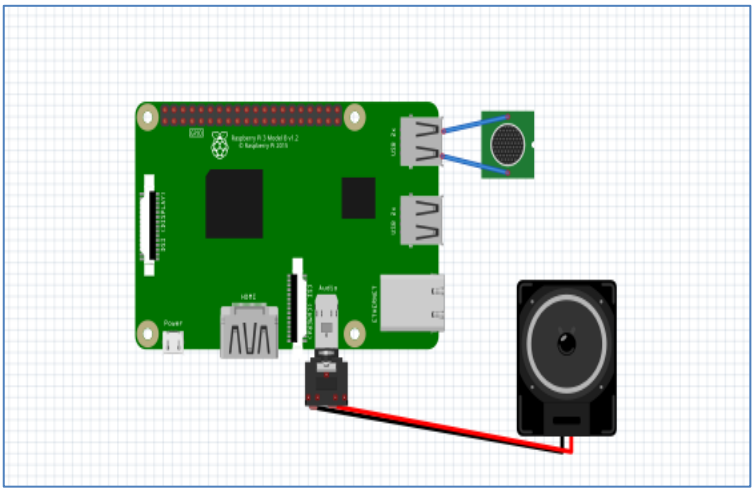

Figure1.2: System circuit diagram

Schematic

In this design, we attached a USB microphone to one of the raspberry pi's four USB ports as the virtual assistant's input. After, process microphone data using raspberry pi and python. The raspberry pi's speakers talk the output. The portable tiny rechargeable speaker is attached to raspberry pi's 3.5mm audio jack output. It is self-powered since raspberry pi cannot deliver adequate electricity. The raspberry pi may be linked to the internet by Ethernet or Wi-Fi, depending on practicality, however Wi-Fi is preferable since it's more efficient.

Work

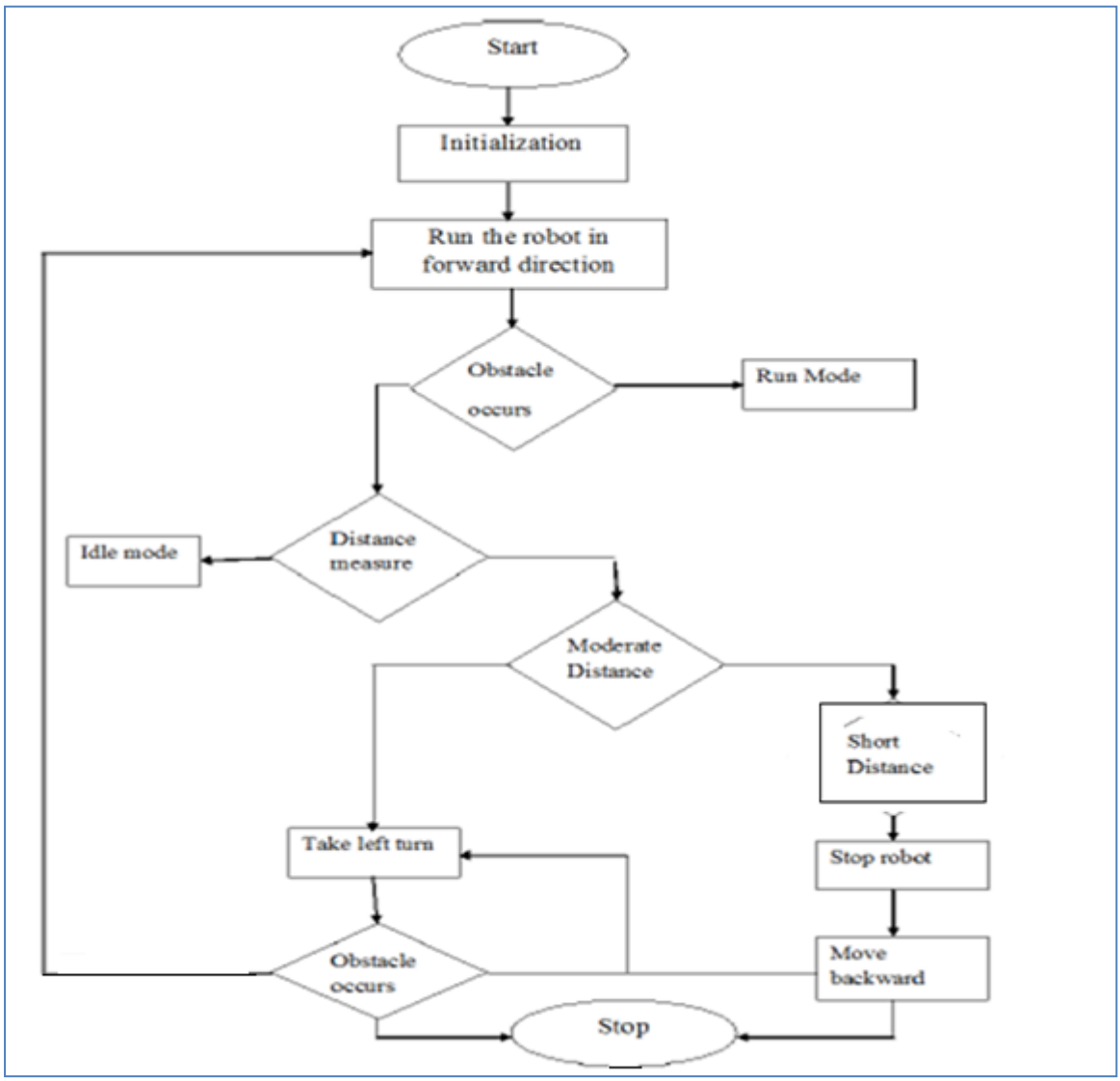

Figure1.3: NLP implementation diagram

The project's workings may be described in three components. They are:

- STT Engine
- Logic Engine 38
- TTS Engine

STTSpeech Text Engine turns user speech into text that the logic engine can process. This requires recording the user's voice, capturing the words (cancelling noise and repairing a virtual assistant. This text is supplied to an NLP module, and depending on the context, APIs are called for Banking or Weather services.

We employed the logic engine to distortion), then using NLP to transform the audio to a text string. Regional accents and speech impairments may hinder word identification, and background noise can be difficult to overcome.

Linguistics (NLP)

Natural-language processing (NLP) focuses on how to train computers to handle vast volumes of natural-language input. Natural-language processing (NLP) converts voice to text when using convert speech-to-text data (query processing system). This is how we apply NLP, which is detailed later.

Logic Engine (Query Processing System)

The logic engine processes the STT engine's text string and sends the result to the TTS engine.The logic engine processes user queries using Python if-then-else clauses. It determines outputs based on inputs. Taking user input, looking for suitable results, and displaying them to the user.When a user speaks into a microphone, the STT engine converts it to text. The logic engine processes the text depending on the user query and triggers the TTS engine.

Logic Engine Implementation

The logic engine's fundamental architecture has two issues. It's:

The brain's recognition system is inadequate because it compares strings.

Having an array of messages and using random.choice() to answer user's inquiry solves issue 1. Importing random on line 1 handles static answers. Then we provide random.choice(array of appropriate messages) to tts. This makes the virtual assistant answer questions differently each time.

Problem 2's solution is difficult. Even if the user asks, "Who are you?" the logical engine will call undefined(). "Who are you?" and "Hey, who are you?" are interchangeable. Checking speech text for keywords solves this issue.

First, separate speech text into words of message. This creates speech words. Check and words of message delete duplicate words. Check if check is a subset of words of message. If a subset, then True; otherwise, False.

Logic Engine keyword matching diagram

The following image explains how logic engine scans for keywords in STT text data and gives the user the relevant result.

TTS Text to Speech Engines translate written text to phonemic representations and subsequently to sound waveforms. Third-party publishers provide TTS engines in multiple languages, dialects, and vocabularies.

Text to Voice Engine translates Virtual Assistant's logic engine output to speech for user interaction. Text-to-Speech Engine makes Virtual Assistant more human than text confirmation.

First, the user begins the system with a microphone. It takes user-supplied sound and processes it on a computer. Hotword detection wakes the system. Only when Hotword is

identified will speech recognition commence and collect input. That sound input given to the speech to text converter is the command, which transforms audio command input to text output identifiable by the computer.

The text is processed and keyword-searched. Our voice command system examines text for matching keywords. After matching key words, it displays the results.

Text output. This is converted to voice using an OCR-based text-to-speech converter. OCR categorises and recognises text, and then text-to-speech transforms it. This output is sent over the raspberry pi's speakers.

F. TESTING

Creating a New Group of Users (Normal and Physically Challenged). An authorised person must be enrolled by following the procedures given below in the correct order.

Step 1: Once you've gotten root access on your Raspberry Pi, navigate to the directory where the code is located and copy it (in this case Desktop).

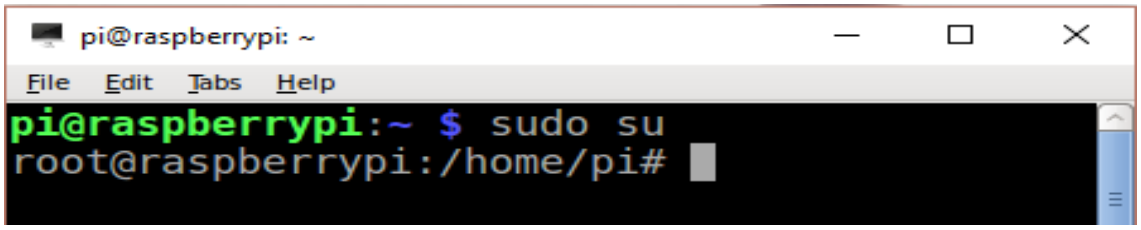

Figure1.4: Getting Root Account Access

Step2: Run the instance of OpenCV by using below commands.

$ source ~/.profile

$ workon cv

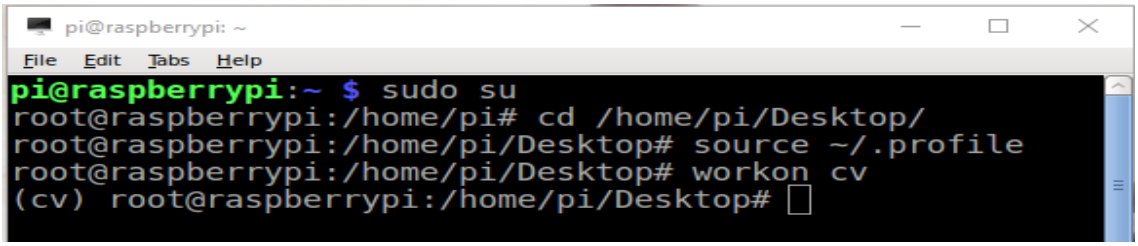

Figure1.5: OpenCV Software Instance Creation

Step3: The third part of the procedure is to run the Python code that can see the environment.

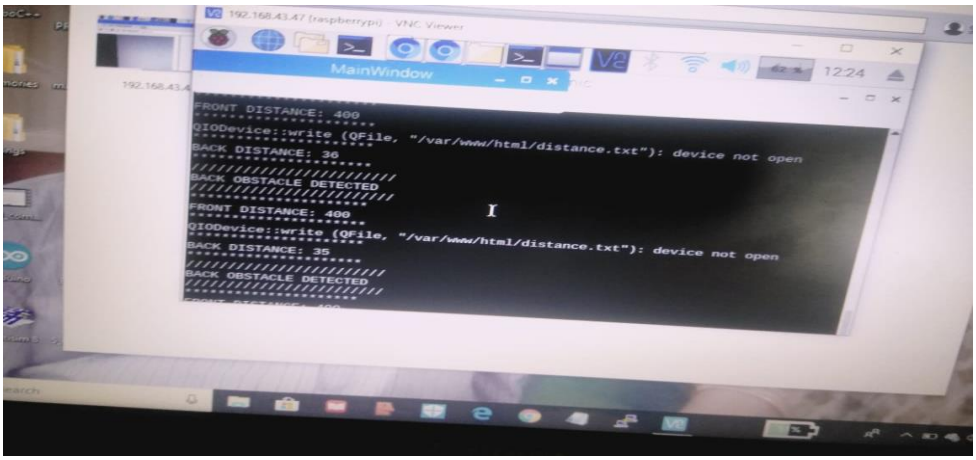

Figure1.6: How Far Things Are by Looking at Pictures

Step4: A video capture window will appear on your desktop after the programme has completed its run without issues.

We may say that the system has been effectively set up with the trained images of authorized persons if all of the aforementioned processes have been completed.

Step5: To begin, open a terminal and go to the code directory.

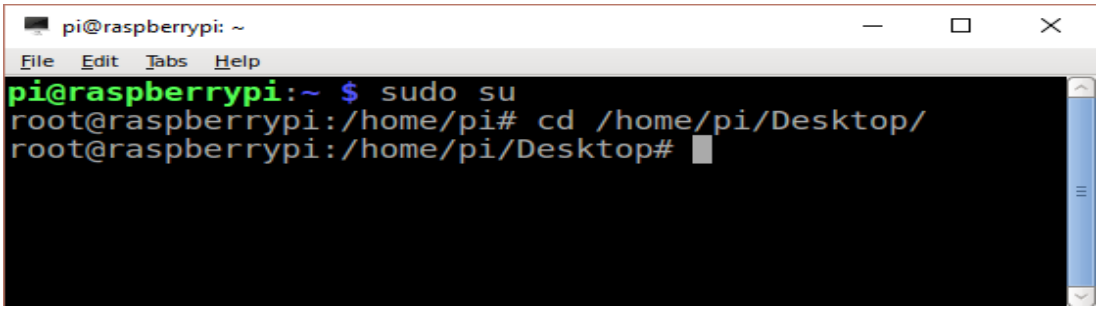

Figure1.7: Making a Reach-Out to the Coding Directory

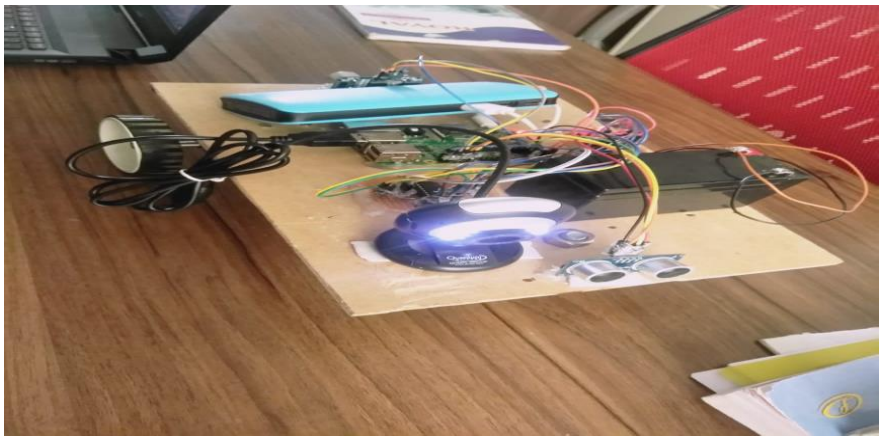

Figure1.8: Capturing Video

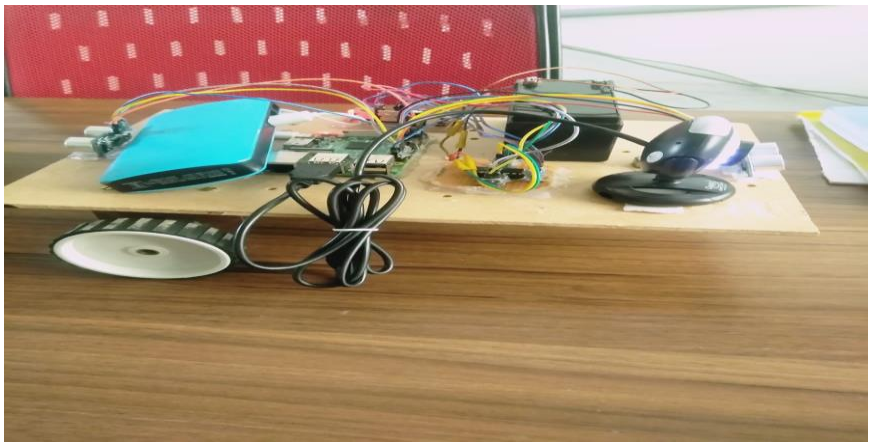

Figure1.9: Processing data to authorized persons

G.CONCLUSION&FURTHER ENHANCEMENTS

A decision was reached to go with the HCSR-04 ultrasonic sensor after conducting testing on it that provided good results for its application in the automobile prototype system now under development. In order to get an accurate reading of how far away the prototype was from any potential obstacles, this tool was employed. The obstacle identification method was written in Python once the distance measurement technique had been successfully completed with the fewest errors possible. Automobiles with the capacity to detect barriers are very valuable since they may prevent many accidents and save many lives.

Both human and unmanned systems will play a vital role in future battlefields, and each will have a distinct position. Using unmanned systems may be more costly, but it is well worth the investment in order to keep people safe. Unmanned systems provide a number of benefits, but they should not be seen as a substitute for present human systems, but rather as an addition to them. Unmanned ground vehicles are used for a variety of purposes, including military operations, surveillance, security, riot control, hostage rescue, law enforcement, border patrol, and more (UGVs). They are more able to withstand adverse conditions including extreme heat or cold, as well as contamination from radioactive, chemical, or biological sources. As a result, the military's operational capabilities may be enhanced through the use of UGVs.

H.REFERENCES

1. Pascolini D, MariottiSPM. Global estimates of visual impairment: 2010. British Journal Ophthalmology Online First published December 1, 2011 as 10.1136/bjophthalmol-2011-300539.
2. A. Helal, S. E. Moore and B. Ramachandran, "Drishti: an integrated navigation system for visually impaired and disabled," Wearable Computers, 2001. Proceedings. Fifth International Symposium on, Zurich, 2001, pp. 149-156.
3. D. Narendhar Singh , B. Pavitra “Web Based Remote Navigational Robot For Multiple Applications”, International Journal of Innovations in Engineering and Technology, Volume 7 Issue 2 August 2016, ISSN: 2319 – 1058

4. B. Pavitra, D. Narendar Singh and MohamamdFarukhHashmi, Title: Voice-Controlled Biped Walking Robot for Industrial Applications Copyright: © 2021 |Pages: 79-92 DOI: 10.4018/978-1-7998-3375-8.ch006
5. Dakopoulos and N. G. Bourbakis, "Wearable Obstacle Avoidance Electronic Travel Aids for Blind: A Survey," in IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), vol. 40, no. 1, pp. 25-35, Jan. 2010.
6. Shraga Shovel, Iwan Ulrich, and Johann Borenstien, “NavBelt and the Guide Cane”, IEEE Transactions on Robotics & Automation, Vol.10, No.1, pp. 9-20, March 2003.
7. Z. Hunaiti, V. Garaj, W. Balachandran and F. Cecelja, "Mobile Link Assessment for Visually Impaired Navigation System," 2005 IEEE Instrumentationand Measurement Technology Conference Proceedings, Ottawa, Ont., 2005, pp. 883-887.
8. Naga Swetha R , D.Narendhar Singh , B. Pavitra “Smart Security System For Oil And Fuel Tankers” -- Palarch’s Journal Of Archaeology Of Egypt/Egyptology 17(7). ISSN 1567-214x, Dec 2020,page No:10047-10054.
9. A. Carullo and M. Parvis, "An ultrasonic sensor for distance measurement in automotive applications," in IEEE Sensors Journal, vol. 1, no. 2, pp. 143-, Aug 2001. [10] V. Magori, "Ultrasonic sensors in air," Ultrasonics Symposium, 1994.Proceedings., 1994 IEEE, Cannes, France, 1994, pp. 471-481 vol.1.
10. A. A. Nada, M. A. Fakhr and A. F. Seddik, "Assistive infrared sensor based smart stick for blind people," Science and Information Conference (SAI), 2015, London, 2015, pp. 1149-1154.