

IoT platform based Smart Assistant for Surveillance

DEBALINA BANERJEE, AKASHJYOTI BANIK, SANJIB KUMAR SINGH, KANDARPA KUMAR SARMA
Department of Electronics and Communication Engineering, Gauhati University Guwahati, Assam, INDIA.

Abstract—Surveillance operations designed to be carried out by a robotic vehicle for entry into an area of higher risks and perform hazardous tasks form the core of this work. The system is integrated with a robotic vehicle that is controlled through a virtual interface and well supported by live video streaming. Here, the motion detection sensor is used as a simple but powerful human presence detector and alarm trigger. Also, the design has a metal detector and gas detecting sensor that can provide precaution against potential landmines present in the operations area and presence of chemicals, high energy materials or poisonous gases on regular and event-based occurrence. The real-time data of the gas sensor is stored in the local machine and also uses a speech recognition system developed using Raspberry Pi microcomputer to detect audio signals. It generates routine alarms on special/unknown/ first time patterns of audio threats. The system is designed using low-cost components.

Keywords—Internet of Things, Surveillance, Sensor-fusion, Face Recognition.

Received: January 29, 2021. Revised: July 13, 2021. Accepted: August 1, 2021. Published: August 10, 2021.

1. Introduction

A robot is an electromechanical and controlled system that can be tasked to perform a range of operations. Lately, with the proliferation of connected devices, the internet of things (IoT) has become a platform of huge application potential. Robotics mixed with IoT and coupled together with sensor fusion, voice-threat recognition, face recognition, real-time data monitoring, live location tracking, and video surveillance provides a diverse range of capabilities. Such combinations are useful for application in hazardous areas and operations as well.

Surveillance operations designed to be carried out by a robotic vehicle for entry into an area of higher risk is vital to security operations. Military and security personnel are required to perform hazardous tasks. Many such tasks can be done by IoT aided robotic systems equipped to carry out surveillance and initiate certain actions. These issues form the core of this work. The system works together with a robotic vehicle. This vehicle is controlled through a virtual interface and has the provision of using live video streaming. Here, the motion detection sensor is used for detecting human presence. Further, object detection and alarm triggers are added as additional components. Also, the design has a metal detector and gas detecting sensor that can provide precaution against potential landmines present in the operations area and presence of chemicals, high energy materials or poisonous gases on regular and event-based occurrence. The real-time data of the gas sensor is stored in the local machine and also uses a speech recognition system developed using Raspberry Pi microcomputer to detect audio signals. It generates routine alarms on special/unknown/ first time patterns of audio threats. The system is designed using low-cost components.

As robots evolve, they will become more versatile, emulating the human capacity and ability to switch job tasks easily [1]. Due to the use of the internet, there is no limitation on the range or distance between the user and the robotic vehicle [2].

Identification of enemy through machine learning algorithms is another important aspect that has grown in recent times. Facial recognition system was used to help

confirm the identity of Osama bin Laden after he was killed in a U.S. raid [3]. A lot of progress has been made in recent years in the field of speech recognition technology. The paper [4] represents a speech recognition system developed using Raspberry Pi microcomputer and USB webcam's microphone to detect and record speech. Many countries around the world are impacted by landmines and other unexploded items remain a danger to people and the environment. Landmine detector for military or humanitarian use is highly dependable and easy to set-up, however, it requires trained personnel to perform these tasks which over time it cost us many lives and money. Upon detecting a landmine, the robot will mark the location where the landmine is detected [5]. Work on real-time video devoid of noticeable frame lag has been reported in [6].

Work to develop and implement an affordable low-cost web-camera based surveillance system for remote security monitoring is discussed in [7] [8].

2. Problem Formulation

The literature survey reveals that even though there are numerous task-specific robots in military systems, an all in one solution is of utmost need in the present Indian Defense scenario. Therefore, to develop such a fusion of sensors, and devices we need to solve the following problems for the proposed system:

- Interfacing several sensors and devices together on the same platform to implement sensor fusion.
- To control the integrated system and monitor the system on a real-time basis.
- Storing relevant data in local machine and taking a calculative decision based upon the outcome of sensor fusion.

3. Proposed Design Considerations

3.1 Sensor Fusion-

Multisensor integration and fusion refer to the synergistic combination of sensory data from multiple sensors to provide more reliable and accurate information. The potential advantages of multisensor integration and fusion are redundancy, complementarity, timeliness, and cost of the information [7].

3.2 IoT

The term IoT represents a general construct for the capability of network devices to sense and collect knowledge from around the world, then share that knowledge across the net wherever it may be processed and used for numerous attention-grabbing functions. The IoT is comprised of smart machines interacting and communicating with other machines, objects, environments, and infrastructures. IOT is a sort of “universal global neural network” in the cloud, it connects various things.

3.3 Speech Processing

Voice or speech recognition is the ability of a machine or program to receive and interpret dictation or to understand and carry out spoken commands [4]. In the case of speech signals, it is a special case of digital signal processing. It is so because the signals are processed in the digital representation. Acquisition, manipulation, storage, transfer, and output of speech signals are included in the aspects of speech processing.

3.4 Face Recognition

Face detection and recognition is the technology which is used to identify a person from a video or photo source. Ever since then the recognition system is being improved and optimized constantly, the technology becomes gradually mature and is more and more widely used in human daily life. It has been used increasingly for forensics by law enforcement and military professionals [3]. The difference between face detection and identification is, face detection is to identify a face from an image and locate the face. Face recognition is deciding "whose face is it?" using an image database [6].

The proposed system contains three modules which are detection, training, and recognition. This method consists of four steps-

1. HAAR FEATURE EXTRACTION

Collecting Haar Features is the first step. Adjacent rectangular regions at a selected location in a detection window, rundown the pixel intensities in every region and the calculative distinction between these sums, all together it is taken into account as a Haar feature.

2. CREATING INTEGRAL IMAGES

Calculating the average intensity of the given image is what mainly an integral image is used for. It is used to make the process of face recognition much faster and easier. Most of the calculated features are irrelevant.

3. ADABOOST TRAINING

AdaBoost is a training method for the classifiers used by selecting the best features. Any machine learning algorithm can be boosted by using AdaBoost. The linear combination of weighted simple weak classifiers results in a strong classifier constructed by this algorithm.

4. CASCADING CLASSIFIERS

The cascade classifier consists of several simple classifiers(stages) known as decision stumps. These stages are considered to be weak learners. The method boosting is used to train each stage. A highly accurate classifier can be trained by taking the weighted average of decisions made by the weak learners. This ability of training is provided by boosting.

Face Detection using HAAR-cascades: Haar wavelet is a mathematical fiction that produces square-shaped waves with a beginning and an end and used to create box-shaped patterns to recognize signals with sudden transformations. An example is shown in Fig. 1. By combining several wavelets, a cascade can be created that can identify edges, lines, and circles with different color intensities.

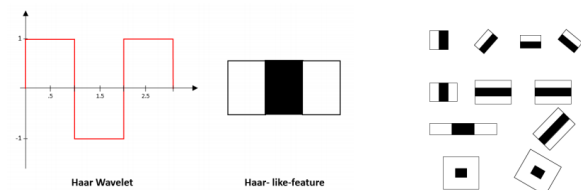


Fig. 1. A Haar wavelet and resulting haar-like features.



Fig. 2. Several Haar-like-features matched to the features of authors face

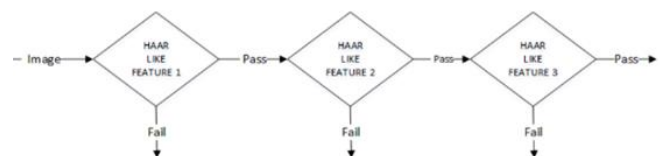


Fig. 3. A Haar cascade flow chart

To analyze an image using Haar cascades, a scale is selected smaller than the target image. It is then placed on the image, and the average of the values of pixels in each section is taken. If the difference between two values passes a given threshold, it is considered a match. Face detection on a human face is performed by matching a combination of different Haar-like-features. For example, the forehead, eyebrows, and eyes contrast as well as the nose with eyes as shown below in Fig. 2. A single classifier is not accurate

enough. Several classifiers are combined to provide an accurate face detection system as shown in the block diagram below in Fig. 3 [6].

3.5 Gas Detection

Gas detection in this context refers to the identification of harmful gases present in the environment. Flammable, combustible and toxic gases and oxygen depletion can be detected using gas detectors. Simultaneous detection of four to five gases and alerting the user if the exposure level of the gases was a matter of concern, is possible in most cases.

3.6 Motion Detection

Motion detection can be determined by using motion detection sensors or by using surveillance systems. Motion detection is one of the important tasks in video processing and scene understanding systems. Visual surveillance is a key technology for the fight against terrorism, crime, public safety (e.g., in transport networks, town centers, schools, and hospitals), efficient management of transport networks and public facilities (e.g., traffic lights, railroad crossings) [9,10].

3.7 Mine Detection

Approximately there are around 100 million mines buried all over the world, which is a big number. Now to remove that huge number of threats a big amount of money is required and not to forget the life-threatening risk taken by the human detectors which shows the lack of technology in this area of research. To relieve the human detectors from this dangerous work, efforts should be made to make the process automated.

3.8 Hardwares and Softwares

1. Hardware used

a) Raspberry Pi

The Raspberry Pi could be a low value, credit-card sized laptop that plugs into a laptop monitor or TV and uses a regular keyboard and mouse.

b) DHT11 sensor

The DHT11 is a very basic, low-cost, real-time digital temperature and humidity sensor. To measure the surrounding air, it uses a thermistor and a capacitive humidity sensor and provides out a digital signal on the data pin (analog input pins not needed).

c) PIR sensor

The term PIR is the short form of the Passive Infra-Red. The term "passive" indicates that the detector doesn't actively participate within the task, which implies, it doesn't emit the referred IR signals itself, rather passively detects the infrared radiation coming from the human body in the surrounding area.

d) MQ-135

The MQ-135 gas sensor senses gases like ammonia, nitrogen, oxygen, alcohols, aromatic compounds, sulfides, and smoke. The boost converter of the chip MQ-3 gas sensing element is PT1301. The operating voltage of this gas sensor is from 2.5V to 5.0V. The MQ-3 gas sensing element encompasses a lower physical phenomenon to scrub the air as a gas sensing material.

e) Metal detector

A metal detector is an electronic instrument that detects the presence of metal. They are useful for finding metal inclusions hidden within objects, or metal objects buried underground. They usually contain a handheld unit with a detector probe that can be swept over the ground or alternative objects.

f) MCP 3008

The MCP3008 is a low cost 8-channel 10-bit ADC. The precision of this ADC is similar to that of an Arduino Uno, and with 8 channels you can read quite a few analog signals from the Pi. This chip is a great option if you just need to read simple analog signals, like from a temperature or light sensor.

g) Microphone

Most of our music collections are probably in digital format, either on CDs or as files on your computer. To be able to listen to our music, a sound card converts digital data to analog sound waves which we can hear. The output is then connected to an earphone or set of speakers. A sound card can also be used to record audio with a microphone.

h) NEO-6M

The NEO-6M GPS module is a well-executable complete GPS receiver that comes with a built-in 25 x 25 x 4 mm ceramic antenna, that provides a powerful satellite search capability. With the power and signal indicators, the status of the module can be monitored.

i) Pi Camera

The 5 MP Raspberry Pi 3 model B camera module is a plug-n-play camera for Raspberry Pi 3. It Supports Video Formats such as: 1080p @ 30fps, 720p @ 60fps and 640x480p 60/90 video. Also, it works fine python programming language.

2. Software used

a) PyCharm

PyCharm is an integrated development environment (IDE) utilized in computer programming, specifically for the Python language. PyCharm is cross-platform, with Windows, macOS and Unix Operating system versions.

b) VNC viewer

In computing, Virtual Network Computing (VNC) is a graphical desktop sharing system that uses the Remote Framebuffer protocol (RFB) to remotely manage a different personal computer.

c) OpenCV

OpenCV is an open source library for image and video analysis, originally introduced more than a decade ago by Intel. Since then, a variety of programmers have contributed to the foremost recent library developments.

4. Design of an Integrated Platform for Surveillance Application

4.1 Proposed Work

This work is intended for the design of a surveillance system integrated with a robotic vehicle for military applications.

A combination of elements is required by robots to be effective: the sophistication of mobility, movement, intelligence, navigation, and purpose. The system that we have proposed is controlled using a free and open source

terminal emulator. The robotic assistance is developed on a Raspberry pi3 microcontroller-based platform. It consists of many features that include monitoring several parameters of the environment like face recognition, harmful gas detection, motion detection, metal detection, and speech recognition. A generalized block diagram of the proposed work is shown in Fig. 4.

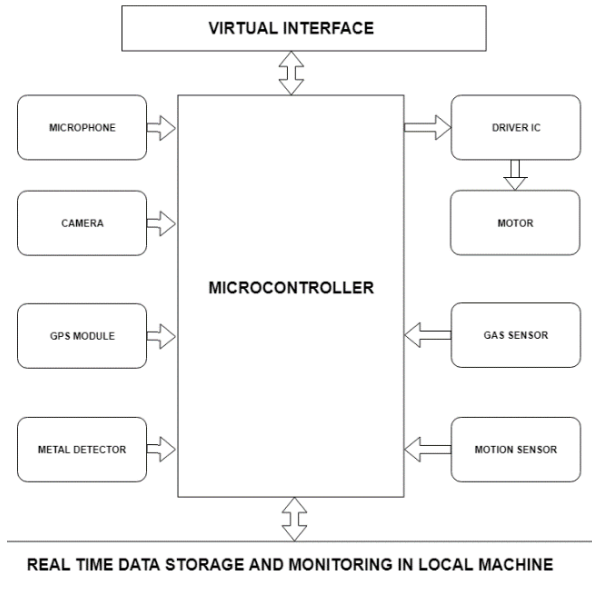


Fig. 4. Generalized Block Diagram of the Proposed System

4.2 Methodology

In this section, we will discuss the working methodology of our system. Fig. 5 shows a detailed block diagram of the integrated system. The whole process is divided into two modes namely The Commanding mode and The Surveillance mode. Both the modes are controlled with the help of a microcontroller i.e. Raspberry Pi 3 through a virtual interface.

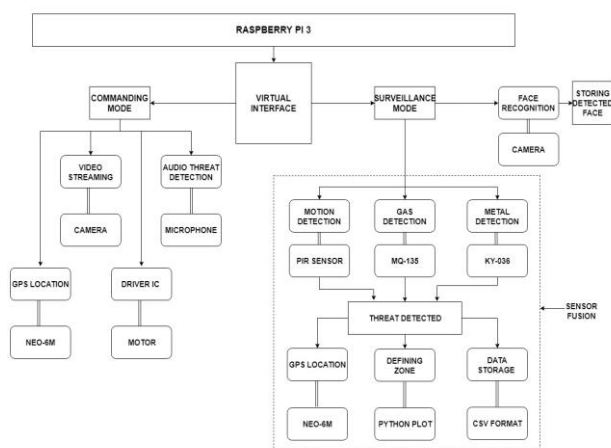


Fig. 5. Detailed Block Diagram of the integrated system

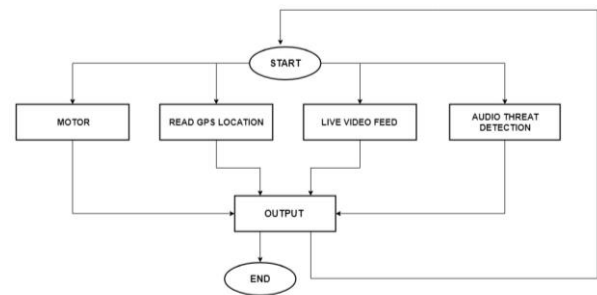


Fig. 6. Flow Chart of Commanding Mode

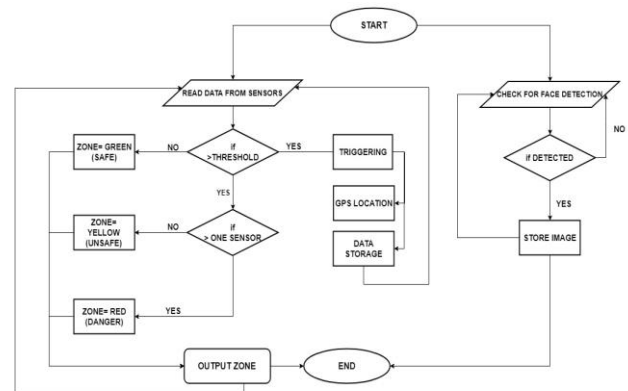


Fig. 7. Flow Chart of Surveillance Mode

Now in the commanding mode, four tasks will be accomplished in parallel. They are real-time location tracking, live video streaming, direction control, and audio threat detection. The sensors and devices used for these tasks are NEO-6M, Pi camera, DC motor and wheels, and microphone respectively as seen in Fig 5. Now in the surveillance mode Sensor Fusion and Human Face Recognition are performed simultaneously. Further, the modes are explained in detail in Fig. 6 and Fig. 7 respectively through flowcharts. We can control the robot car with the help of navigation keys on a laptop or computer. For sensor fusion, we have used three sensors namely PIR motion sensor, MQ-135 gas sensor, and KY-036 metal touch sensor. The algorithm checks for any threat detected by these sensors and on detection triggers the GPS location and simultaneously stores relevant data in a CSV file. Sensor fusion combines the data of several sensors to improve the application of the system and its performance. The data combining from each sensor corrects the deficiencies of them to come to an accurate conclusion. Here also the sensor fusion helps us in determining the range of the zones by combining the information received from the individual sensors. Three defined zones are using real-time python plot, they are Safe (Green), Unsafe (Yellow) and Danger (Red) zones respectively.

Along with this a Face Recognition system also works. It checks for human faces and if detected tries to recognize from a known set of data and stores the image with timestamp and id of the person. If it does not recognize then it stores the image with timestamp and id as "Unknown". In this way, the complete integrated system functions for better supervision of the military personnel.

5. Experimental Results and Discussions

First of all, we have collected the dataset of four known faces (two males and two females). This dataset consists of a total of 200 images (50 images each) captured at different light intensities, angular positions, and backgrounds. Using this dataset, the machine is trained. After training is completed, we proceed for recognition of the trained faces along with an unknown face for determining the efficiency of the used algorithm. Then, while checking the recognition algorithm we attempted to plot different confusion matrices by changing the background, lighting conditions and angular positions. We got a total of 8 confusion matrices where table 4.1 to 4.4 represents the matrix containing the number of times the system recognized particularly trained faces with a high-resolution camera. The same set of experiments is repeated with a low-resolution camera, the result of which is seen in Tables 4.5 to 4.8.

TABLE I. DAYLIGHT SIMPLE BACKGROUND (720 P)

□	A	B	C	D	E	F
1	n=250	Akash	Debalina	Sanjib	Akanksha	Unknown
2	Akash	50	0	0	0	0
3	Debalina	0	47	0	3	0
4	Sanjib	0	0	47	3	0
5	Akanksha	0	0	0	50	0
6	Unknown	3	1	0	0	46

In Table I. we carry out the experiment in daylight with a simple background and high-resolution camera (720 p). Here the total TP elements are 240 and total FP elements are 10. Hence, accuracy and error rates are as follows. Accuracy = 0.96, Error rate = 0.04

TABLE II. DAYLIGHT COMPLEX BACKGROUND (720 P)

□	A	B	C	D	E	F
1	n=250	Akash	Debalina	Sanjib	Akanksha	Unknown
2	Akash	44	2	0	0	4
3	Debalina	0	49	0	1	0
4	Sanjib	4	0	45	0	1
5	Akanksha	0	0	3	47	0
6	Unknown	1	0	2	0	47

In Table II. we carry out the experiment in daylight with a complex background and high-resolution camera (720 p). Here the total TP elements are 233 and total FP elements are 17. Hence, accuracy and error rates are as follows. Accuracy = 0.932, Error rate = 0.072. In Table III. we carry out the experiment in nightlight with a simple background and high-resolution camera (720 p). Here the

total TP elements are 244 and total FP elements are 6. Hence, accuracy and error rates are as follows. Accuracy = 0.976, Error rate = 0.024. In Table IV. we carry out the experiment in nightlight with a complex background and high-resolution camera (720 p). Here the total TP elements are 231 and total FP elements are 19. Hence, accuracy and error rates are as follows. Accuracy = 0.924, Error rate = 0.076.

TABLE III. NIGHTLIGHT SIMPLE BACKGROUND (720 P)

□	A	B	C	D	E	F
1	n=250	Akash	Debalina	Sanjib	Akanksha	Unknown
2	Akash	48	0	0	2	0
3	Debalina	0	46	0	4	0
4	Sanjib	0	0	50	0	0
5	Akanksha	0	0	0	50	0
6	Unknown	0	0	0	0	50

TABLE IV. NIGHTLIGHT COMPLEX BACKGROUND (720 P)

□	A	B	C	D	E	F
1	n=250	Akash	Debalina	Sanjib	Akanksha	Unknown
2	Akash	50	0	0	0	0
3	Debalina	0	45	0	5	0
4	Sanjib	0	0	43	0	7
5	Akanksha	2	4	0	42	2
6	Unknown	0	0	0	1	49

TABLE V. DAYLIGHT SIMPLE BACKGROUND (0.3 MP)

□	A	B	C	D	E	F
1	n=250	Akash	Debalina	Vishesh	Jayeasha	Unknown
2	Akash	50	0	0	0	0
3	Debalina	0	48	0	5	2
4	Vishesh	0	0	50	0	0
5	Jayeasha	0	5	0	45	0
6	Unknown	0	2	0	0	48

In Table V. we carry out the experiment in daylight with a simple background and low-resolution camera (0.3 mp). Here the total TP elements are 241 and total FP elements are 9. Hence the accuracy and error rate are as follows. Accuracy = 0.964, Error rate = 0.036. In Table VI. we carry out the experiment in daylight with a complex background and low-resolution camera (0.3 mp). Here the total TP elements are 241 and total FP elements are 9. Hence, accuracy and error rates are as follows. Accuracy = 0.964, Error rate = 0.036. In Table VII. we carry out the

experiment in nightlight with a simple background and low-resolution camera (0.3 mp). Here the total TP elements are 243 and total FP elements are 7. Hence, accuracy and error rates are as follows. Accuracy = 0.972, Error rate = 0.028.

TABLE VI. DAYLIGHT COMPLEX BACKGROUND (0.3 MP)

	A	B	C	D	E	F
1	n=250	Akash	Debalina	Vishesh	Jayeesh	Unknown
2	Akash	45	0	0	0	5
3	Debalina	0	49	0	0	1
4	Vishesh	0	0	50	0	0
5	Jayeesh	0	3	0	47	0
6	Unknown	0	0	0	0	50

TABLE VII. NIGHTLIGHT SIMPLE BACKGROUND (0.3 MP)

	A	B	C	D	E	F
1	n=250	Akash	Debalina	Vishesh	Jayeesh	Unknown
2	Akash	50	0	0	0	0
3	Debalina	0	44	0	0	6
4	Vishesh	0	0	50	0	0
5	Jayeesh	0	0	0	49	1
6	Unknown	0	0	0	0	50

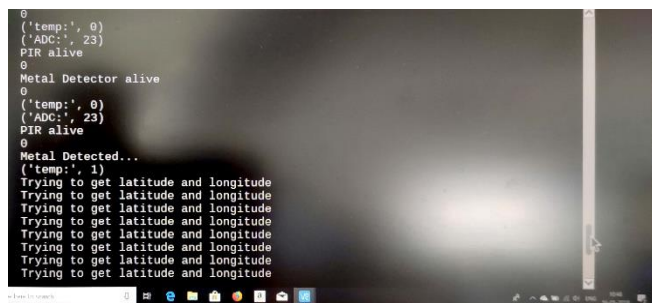


Fig. 8. Before the location is triggered on detection of threat

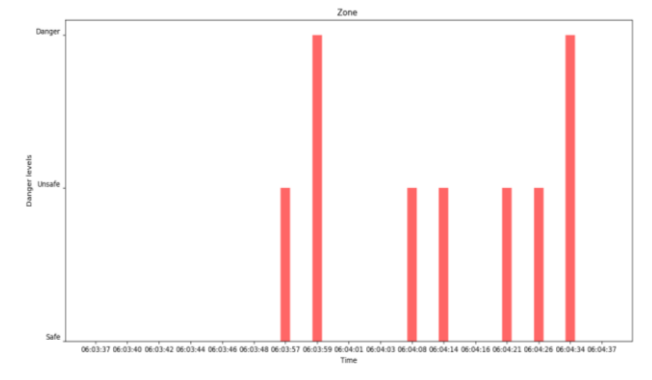


Fig. 9. Decision plot

The proposed prototype is programmed in such a manner that it displays the real-time location of the robot where it detects the threat/s. Fig. 8 shows the image of the Rpi terminal when the threat is just detected and before the location is displayed.

We have implemented sensor fusion so that the system could identify threats using the data from different sensors and define the danger zones as defined. Fig. 9 shows the decision plot defining various zones.

6. Conclusion

Here we discuss the application of sensor fusion in a military-bot using Raspberry Pi which has been implemented for hazardous conditions and found to work efficiently using video, image, and voice-based inputs. It is found to be suitable for landmine detection which can be extended to other approaches also. The proposed system, when deployed in a real situation, will allow the personnel to collect real-time data and do further improvements in the features of the whole system.

References

- [1] D.Salunkhe, A.U.Jagatap, D.R.Ghodekar, P.V.Shelke, "Wireless multifunctional robot for military applications"International Research Journal of Engineering and Technology, Vol.05,no.:03, Issue: March 2018
- [2] M. Dharaskar, V. Gupta, P. Kale, R. Chopade, A. Jaiswal, "IOT based surveillance robotic car using Raspberrypi", International Research Journal of Engineering and Technology, Vol.05, no.: 03, Issue: March 2018.
- [3] M.Madhuram, B. P Kumar, L Sridhar, N Prem, V Prasad, "Face Detection and Recognition Using OpenCV", International Research Journal of Engineering and Technology, Vol.:05, no.: 10, Issue: October 2018.
- [4] H. Gyulyustan, S. Enkov, "Experimental speech recognition system using Raspberry Pi", IOSR Journal of Computer Engineering, Vol.:19, no.: 3, Issue: May-June 2017.
- [5] A. Pajaziti, K.C. Cheok, X.Bajrami, "Semi autonomous mobile robot for mine detection",10.13140/RG.2.1.4754.3121.
- [6] L Dinalankara, "Face Detection & Face Recognition Using Open Computer Vision Classifiers" Issue: August 4, 2017.
- [7] R. C. Luo, C.C Yih and K. L. Su, "Multisensor fusion and integration: approaches, applications, and future research directions," in IEEE Sensors Journal, vol. 2, no. 2, pp. 107-119, April 2002.
- [8] P. B. Patel, S.Jadhav, M.B. Potdar, V. M. Choski, "Smart Motion Detection System using Raspberry Pi", International Journal of Applied Information Systems (IJ AIS), Vol.:10, no. 05, Issue: February 2016.
- [9] J.H. Piater, S. Richetto, and J. L. Crowley, J. M. Ferryman, Ed., "Event-based Activity Analysis in Live Video using a Generic Object Tracker", Proceedings 3rd IEEE Int. Workshop on PETS, Copenhagen, June 1, 2002.
- [10] B. Singh, D. Singh, G. Singh, N. Sharma, and V. Sibbal, "Motion detection for video surveillance," 2014 International Conference on Signal Propagation and Computer Technology (ICSPCT 2014), Ajmer, 2014, pp. 578-584.
- [11] L. Robledo, M. Carrasco, and D. Mery, "A survey of land mine detection technology", International Journal Of Remote Sensing, Vol. 30,no.:09 , Issue: 2009.

Creative Commons Attribution License 4.0 (Attribution 4.0 International, CC BY 4.0)

This article is published under the terms of the Creative Commons Attribution License 4.0
https://creativecommons.org/licenses/by/4.0/deed.en_US