Real Time IOT Based Smart Surveillance Alerting System with Motion and Face Detection

Rakesh Kalshetty
P.G Student
PDA College of Engineering
Kalaburagi, Karnataka
rakeshk.work@gmail.com

Prof. Chandrakanth Biradar
CSE Department,
PDA college of Engineering
Kalaburagi, Karnataka
csbiradar.pda@gmail.com

Abstract - Video Surveillance has been used significantly with lot of effectiveness in past decade or so for monitoring Home, School and other public places. However, one of the major issues raises from such a Surveillance system requires the data to be logged which means that video of the feet coming from the camera is recorded it is important to know that a year back Apple faced a huge loss in such kind of monitoring system of the user activities Julian Assanage also reviled an important analysis system by FBI where FBI monitored almost 2 Billion people on the Earth in order to know more about that therefore new rylm in the security system is required which prevents the authorities from knowing the private information about the people being monitored in this work therefore we focus on developing a new generation surveillance system assisted with IoT that is based on intelligent system such a system should not compromise the users identity at the same time offer the highest level of security our designed system would not lock the entire data rather various sensors will be placed along with motion sensor to analyze the abnormality observed within a specific context in a house or a building only in case of such abnormality the events will be recorded and that recorded event will be saved in the cloud will further introduce a cloud based face detection and recognition mechanism in order to enhance this system and this system can further be used in a home security use case specification as a demonstration of the application of this system.

The build system therefore offer

i) privacy preservation of the people.

ii) at produce high level of security by facilitating the video logging capability when abnormality is detected.

KEYWORDS

IoT, Real Time, Privacy Preservation, Motion, Face Detection, Surveillance, etc.

1. INTRODUCTION

The Internet of Things (IoT) is the inter-networking of physical devices, vehicles, buildings, and other items embedded with electronics, software, sensors, actuators, and network connectivity which enable these objects to collect and exchange data. Typically, IoT is expected to offer advanced connectivity of devices, systems, and services that goes beyond machine-to-machine (M2M) communications and covers a variety of protocols, domains, and applications. The interconnection of these embedded devices (including smart objects), is expected to usher in automation in nearly all fields, while also enabling advanced applications like a smart grid, and expanding to areas such as smart cities. "Things", in the IoT sense, can refer to a wide variety of devices such as heart monitoring implants, biochip transponders on farm animals, electric clams in coastal waters, automobiles with built-in sensors, DNA analysis devices for environmental/food/pathogen monitoring, or field operation devices that assist fire fighters in search and rescue operations. Legal scholars suggest to look at "Things" as an "inextricable mixture of hardware, software, data and service".

These devices collect useful data with the help of various existing technologies and then autonomously flow the data between other devices. Current market examples include home automation (also known as smart home devices) such as the control and automation of lighting, heating (like smart thermostat), ventilation, air conditioning (HVAC) systems, and appliances such as washer/dryers, robotic vacuums, air purifiers, ovens, or refrigerators/freezers that use Wi-Fi for remote monitoring. As well as the expansion of Internet-connected automation into a plethora of new application areas, IoT is also expected to generate large amounts of data from diverse locations, with the consequent necessity for quick aggregation of the data, and an increase in the need to index, store, and process such data more effectively. IoT is one of the platforms of today's Smart City, and Smart Energy Management Systems.

Surveillance is the monitoring of behavior, activities, or other changing information for the purpose of influencing, managing, directing, or protecting people. This can include observation from a distance by means of electronic equipment (such as closed-circuit television (CCTV) cameras) or interception of electronically transmitted information (such as Internet traffic or phone calls). It can also include simple no- or relatively low-technology methods such as human intelligence agents and postal interception. Surveillance is used by governments for intelligence gathering, prevention of crime, the protection of a process, person, group or object, or the investigation of crime. It is also used by criminal organizations to plan and commit crimes, such businesses to gather intelligence, and by private investigators.

2. RELATED WORK

The system detects any kinds of motion and also detects face. The system operates in the following sequence. We install sensors in the houses, schools, colleges, private properties, restricted areas, etc. Cameras detect behaviour such as motions and human faces. Cameras information is sent as a alert to the administrator via mail. The system analysis motion and saves images in local storage. Determined result is sent over the cloud (YouTube). Where the administrator can view live streaming of actual footage in real time. Hence necessary action can be taken by administrator. We create an real time IoT based smart surveillance alerting system with motion and face detection.

3. PROBLEM STATEMENT

Privacy preserving surveillance and monitoring system is extremely critical in today's world there is an immense amount of mistrust between the authorities and the public as per as using their private data is concerned many of the people are hesitant to obtain for their mobile sensor or any other public sensor monitoring because they fear critical information about their private life may be reviled therefore new techniques are required that can give comfort to public to participate in a secured system and yet be confident that their private data should not be reviled.

(For Example: The security camera is implemented in buses not everybody travelling by bus want to be known that I am travelling there might be other concern with whom you are travelling so under that situation one may not opt for such system because no security technique in the modern age exist that does not reveal the identity of the person either partially or completely it is extremely important to develop new architecture that meet this call)

The overall problem can be summarized as to use a cloud based processing technique in order to develop an effective video surveillance system that can protect the privacy of the participating people and yet offer a high degree of security to the system.

4. PROPOSED SYSTEM

i) In order to avoid such privacy concern and save the bandwidth and space required for the system we propose a novel privacy preserving video processing based security system where the camera feet will be processed using a for computing frame work first by a local system the local system will analyze frame by frame and try to detect abnormality.

ii) The local system will also be integrated with the cloud data base where it will try to find any abnormal person presence within the video frame such changes are called events whenever there is an event the system will start tracking the video feet and start storing at the end of the event the video is being packed re-encoded and stored in cloud storage and the notification is sent to specific authority.

5. SYSTEM ARCHITECTURE

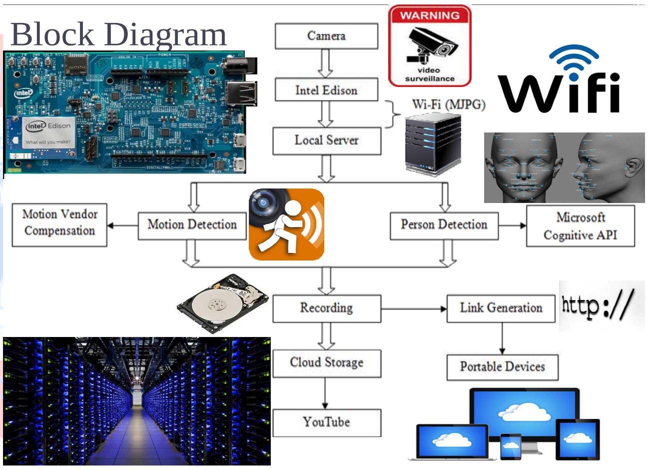


Fig 1. Shows the architecture of the Smart Surveillance system

5.1. Cameras

The role of Cameras is to monitor the targeted or focused areas which must be capable to monitor selected area regardless of climatic conditions and on any given time of any specific day.

5.2. Intel Processor

After getting the data form cameras the video/photos sent to the processor where the data is separated where as the videos are sent to the cloud and the photos are stored in the cloud this sorting is done by the instructions given to the processor via visual studio program.

5.3. Local Server

Here with the help of an local server the IP cameras are being configured an also allotted the IP address based on an local server through which all the setup is configured say IP cameras and main Module (Intel Processor) are connected to one another

5.4. Motion Detection

The Method/Technique used here is Background Subtraction where the system observe for an significant change in the background or we can say that there must be noticeable amount of change in test are so that the system detect motion when compared to the original or fed image.

5.4.1 Background subtraction

Background subtraction is a general technique for the separation of motion within fixed images. It detects regions in motion by deducting the needed current image pixel-by-pixel from an attributed background image. This is established by equating multiple images from the start. The main point of background subtraction is the detection of objects in motion. This is done by initially setting a background, then deducting the present or current frame which contains the objects in motion that are to be detected. This process is easy and it acts in accordance to a simple protocol. It accurately aids in obtaining features of target data, nevertheless, it had sensitivity towards small changes within the external environment. Hence, its only usage is within situations in which the background can be predicted.

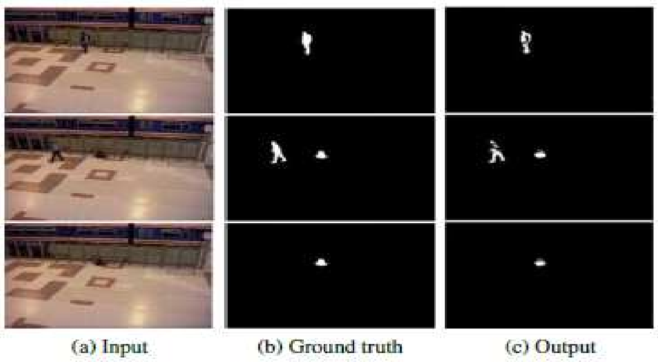


Fig 2. Shows the Background Subtraction method

5.4.1.1 Background Image Initialization

There are numerous other means to gain the first background image. Some of these ways include the background being the first frame, the common brightness of the pixel of first few frames being the background or utilizing background image series lacking the faith of the substances in motion to guess the parameters of the background model. Of the various previously stated methods, the most common method is used for the initialization of a background image. However, there are underlying issues that will be present; this can be eliminated by the median method. Below shows the equation1:

$$\beta init(x, y) = median fk(x, y)k = 1, 2, ...n \quad ...(1)$$

beta init is depicted as the initial background while n is depicted as the sum of the selected frames.

5.4.1.2 Background Update

The background model has a great ability to adapt to the changes of time. There needs to be an update for the background in order to extract the object in motion. When detecting, the pixels are considered to be a part of the object in motion. They manage the original background grey values and are not updated. As it relates to the pixels that are considered the background, the following equation (2) is used to update the background module:

$$\beta k + 1(x, y) = \beta\beta k(x, y) + (1 - \beta)Fk(x, y) \quad ....(2)$$

Where beta belongs to (0, 1), a renewed parameter, and Fk(x,y) is referred to as the pixel grey value within the existing framework. Betak(x, y) is the background value of the existing frame, while beta k+1(x,y) is the background value of the other frame. Using the static camera, the background model has the ability to remain fixed for a long time period. This method, with great effect, can avert the unforeseen unpredicted occurrence of the background. Examples include the immediate form of an object within the background, which is not a part of the novel background. Nonetheless, by the pixel grey value update of the background has the possibility to be adapted effectively, through the light impact and variations of the external environment which includes the weather.

5.4.1.3 Moving Object Extraction

Background subtraction is a regular method which separates the important objects within a frame. The method includes deducting an image containing the object found in the earlier background image and has no important foreground objects. The region found within the image plane where a noticeable contrast can be seen in these images shows the location of the pixel of the object in motion. Expressed by clusters of pixel, the objects are disjoined from the background image with the use of the threshold method. When the background image beta k(x, y) is retrieved, it is deducted from the existing frame Fk(x, y). In the event that the pixel distinction is large compared to the set threshold Th, the pixels then appear in the objects in motion. If this does not occur, they appear as background pixels within the frame. After this process, objects in motion are able to be detected. The following shows the expression 3:

$$Dk(x,y) = 1 if|Fk(x,y) - \beta k(x,y)| > T, Dk(x,y) = 0 \quad ....(3)$$

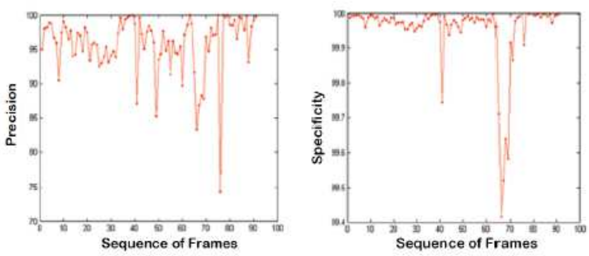
Depicting the complementing human part deriving out of the region of motion is the main purpose of human body detection. The region of motion can be correlated to multiple objects in motion. These include birds, vehicles, swaying trees, and floating clouds. This is further decided on whether or not the object in motion is a human being. The shape features of regions in motion may be used. The following shows the criteria for judging:

—The set threshold is smaller than the object area.

—The region of the objects aspect ratio needs to adhere to the set ratio. Once these prerequisites are met, the confirmation is made of whether or not the object in motion is a mobile human body.

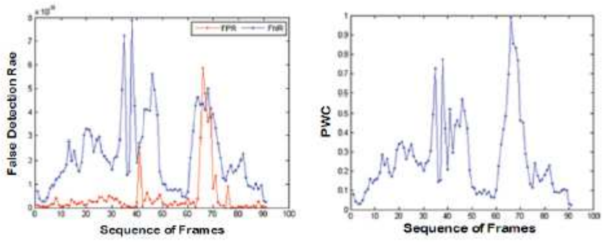
5.4.1.4 The Removal of Noise

Owing to the issues relating to the background subtraction method, the variation of the image obtained includes the noise amount in addition to the region of motion. These noises may or may not be contained in the illumination changes or environmental factors during transmitting the video from the camera for further processing. Hence, the removal of noise is needed. The noise may be filtered by adopting a median filter. The region of motion includes flying birds, swaying tress, human beings, non-body parts, and also flying clouds. Further processing may be done using morphological methods. Corrosion operation is done to filtrate the majority of non-body regions of motion and in addition, without injury, maintain the structure of human motion. After this, the inaccessible areas of the image and the intervention of small pieces are disregarded in an effort to gain a precise human motion region.



(a) (b)

Fig. 3: Performance evaluation: (a) Precision (b) Specificity



(a) (b)

Fig. 4: Performance evaluation: (a) FPR and FNR (b) PWC

5.4.1.5 Understanding Behaviour

After the success of the detection of mobile humans from multiple frames within an image sequence, there is a natural issue to understand the behaviours of humans from image sequences. The understanding of behaviour includes description and action identification. This can aid in the enhancement of numerous analysis systems of human motion. This process recognizes and analyses patterns within human motion and produces a top-standard description of interactions and actions. This is also the key area of any future study in human motion analysis.

5.4.1.6 ANALYSIS OF BACKGROUND SUBTRACTION PERFORMANCE

Programmed/automatic visual recognition of items is an essential undertaking for a huge scope of home, business, and modern applications. Video cameras are among the most generally utilized sensors for reconnaissance to savvy spaces for feature conferencing. Moving target identification intends to recognize moving items from the background picture to the persistent feature video. Following moving target items intends to discover different areas of the moving question in the feature. There is a necessity to create calculations for undertaking, for example, moving items identification. Presently utilized strategies as a part of in-motion item location are basically the frame subtraction technique, the foundation subtraction system, and the optical stream strategy. Frame subtraction technique is the contrast between two continuous frames to focus the vicinity of in-motion items. Its estimation is straightforward and simple to create. For a mixture of element situations, it has solid flexibility; however it is generally hard to acquire a complete diagram of in-motion items. Thus, the discovery of moving items is not precise. Optical flow systems ascertain the picture optical stream field and do much preparing, as indicated by the optical stream appropriation components f picture. This strategy facilitates the complete development data and identifies the in-motion items from the background more superior, because of a huge amount of computation, affectability to commotion, and poor hostile to calmer execution, making this system not appropriate for constant requesting events. Background subtraction is conventional when it comes to identify moving objects in videos. This method is accomplished by identifying the difference among the current frame as well as a background model, or background image, along with an understandable and uncomplicated algorithm. As a consequence, it will also produce accurate information about the specific object when the case of that specific background is known beforehand. This procedure is adequate and efficient to upgrade the development of identifying the moving object. We have mentioned an exact as well as a real-time background subtraction approach in this article, which adequately excludes the collision of changes of light. This simple and intelligible algorithm is capable of identifying the moving object greater as well as it has a wide relevance. This procedure is often used in video surveillance applications.

5.5. Face Detection

The Technique used here is Viola-Jones Haar Cascade method where with the Haar-Like features such as different Haar pattern are taken into consideration and kept over an image of the face with the combinations to white and gray cells the layout of an face is identified her to identify an face thousands of faces are taken into consideration where the face must be of all kinds of human expressions which are required to perform Haar operations over them and at last to conclude that the feed object is an face.

5.5.1 Viola Jones Face detection Method

The Viola-Jones algorithm uses Haar-like features, that is, a scalar product between the image and some Haar-like templates. More precisely, let I and P denote an image and a pattern, both of the same size N × N (see Figure 5). The feature associated with pattern P of image I is defined by eq 1

$$\sum_{1\le i\le N}\sum_{1\le j\le N} I(i,j)1_{P(i,j)\text{ is white}} - \sum_{1\le i\le N}\sum_{1\le j\le N} I(i,j)1_{P(i,j)\text{ is black}}.$$

......(1)

To compensate the effect of different lighting conditions, all the images should be mean and variance normalized beforehand. Those images with variance lower than one, having little information of interest in the first place, are left out of consideration.

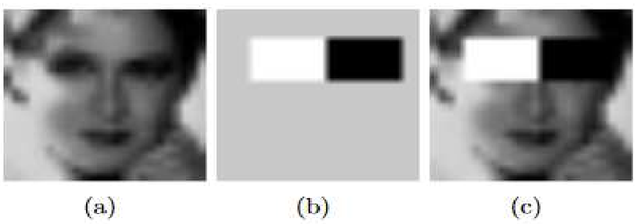


Figure 5: Haar-like features. Here as well as below, the background of a template like (b) is painted gray to highlight the pattern's support. Only those pixels marked in black or white are used when the corresponding feature is calculated.

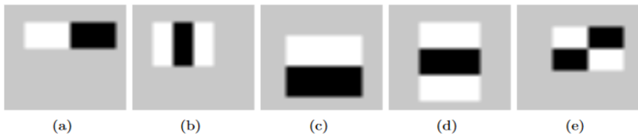


Figure 6: Five Haar-like patterns. The size and position of a pattern's support can vary provided its black and white rectangles have the same dimension, border each other and keep their relative positions. Thanks to this constraint, the number of features one can draw from an image is somewhat manageable: a 24 × 24 image, for instance, has 43200, 27600, 43200, 27600 and 20736 features of category (a), (b), (c), (d) and (e) respectively, hence 162336 features in all.

In practice, five patterns are considered (see Figure 6). The derived features are assumed to hold all the information needed to characterize a face. Since faces are by and large regular by nature, the use of Haar-like patterns seems justified. There is, however, another crucial element which lets this set of features take precedence: the integral image which allows to calculate them at a very low computational cost. Instead of summing up all the pixels inside a rectangular window, this technique mirrors the use of cumulative distribution functions as shown in eq 2. The integral image II of I

$$\mathrm{II}(i,j) := \begin{cases} \sum_{1\le s\le i}\sum_{1\le t\le j} I(s,t), & 1\le i\le N \text{ and } 1\le j\le N \\ 0, & \text{otherwise} \end{cases},$$

.......(2)

is so define as eq 3 that

$$\sum_{N_1\le i\le N_2}\sum_{N_3\le j\le N_4} I(i,j) = \mathrm{II}(N_2,N_4) - \mathrm{II}(N_2,N_3-1) - \mathrm{II}(N_1-1,N_4) + \mathrm{II}(N_1-1,N_3-1), \quad (1)$$

.....(3)

holds for all N1 ≤ N2 and N3 ≤ N4. As a result, computing an image's rectangular local sum requires at most four elementary operations given its integral image. Moreover, obtaining the integral image itself can be done in linear time: setting N1 = N2 and N3 = N4 in (1), we find I(N1,N3) = II(N1, N3) − II(N1, N3−1) − II(N1−1, N3) + II(N1−1,N3−1).

5.2.2 Feature Selection with Adaboost

How to make sense of these features is the focus of Adaboost. Some terminology. A classifier maps an observation to a label valued in a finite set. For face detection, it assumes the form of f : Rd → {-1,1}, where 1 means that there is a face and -1 the contrary (see Figure 7) and d is the number of Haar like features extracted from an image. Given the probabilistic weights w. € R+ assigned to a training set made up of n observation label pairs (xi, yi), Adaboost aims to iteratively drive down an upper bound of the empirical loss

$$\sum_{i=1}^{n} w_i 1_{y_i \neq f(x_i)}, \quad .....(4)$$

under mild technical conditions. Remarkably, the decision rule constructed by Adaboost remains reasonably simple so that it is not prone to over fitting, which means that the empirically learned rule often generalizes well. Despite its groundbreaking success, it ought to be said that Adaboost does not learn what a face should look like all by itself because it is humans, rather than the algorithm, who perform the labelling and the first round of feature selection, as described in the previous section.

(a) (b)

Figure 7: Some supervised examples: (a) positive examples (b) negative examples. All of them are 24 X 24 greyscale images.

The building block of the Viola-Jones face detector is a decision stump, or a depth one decision tree, parameterized by a feature *f* € {1, . . . , d }, a threshold t € R and a toggle T € {-1, 1}. Given an observation x € Rd, a decision stump h predicts its label using the following rule

$$h(x) = (1_{\pi_f x \geq t} - 1_{\pi_f x < t})T = (1_{\pi_f x \geq t} - 1_{\pi_f x < t})1_{T=1} + (1_{\pi_f x < t} - 1_{\pi_f x \geq t})1_{T=-1} \in \{-1,1\}, \quad .....(5)$$

where πfx is the feature vector's f-th coordinate. Several comments follow:

1. Any additional pattern produced by permuting black and white rectangles in an existing pattern (see Figure 4) is superfluous. Because such a feature is merely the opposite of an existing feature, only a sign change for t and T is needed to have the same classification rule.

2. If the training examples are sorted in ascending order of a given feature f, a linear time exhaustive search on the threshold and toggle can find a decision stump using this feature that attains the lowest empirical loss

$$\sum_{i=1}^{n} w_i 1_{y_i \neq h(x_i)}, \quad ......(6)$$

on the training set. Imagine a threshold placed somewhere on the real line, if the toggle is set to 1, the resulting rule will declare an example x positive if πfx is greater than the threshold and negative otherwise. This allows us to evaluate the rule's empirical error, thereby selecting the toggle that fits the dataset better.

6. RESULT

When the Cameras and Main System is connected and configured with one another using internet then with the help of visual studio the program will be launched on the main system and the following activity window is popped out and the Real Time Surveillance System starts working and it follows following steps successively as shown below

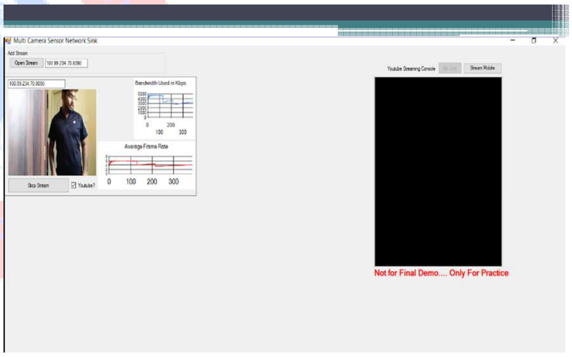


Fig 5. Above Screenshot we can see that the software is collecting the motion in test area and sound values continuously.

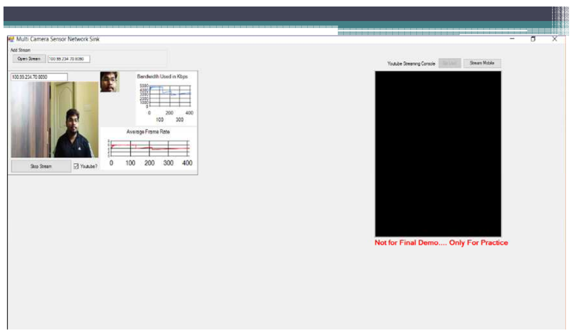


Fig 6. The above Scenario occurs whenever the face is detected and immediately it triggers an email alert. The graph shows bandwidth used.

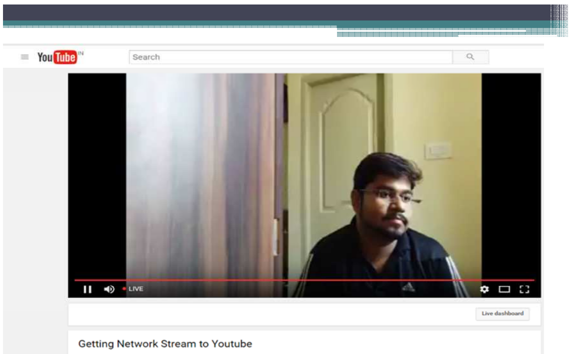


Fig 7. Above Scenario occurs whenever the motion and face is detected and the triggering system start continuous streaming of test area footage into cloud in this case YouTube is live streaming the content

7. CONCLUSION

The developed project is an Real Time Smart Surveillance Alerting system where the main objective covered is that the system which must work in real time and must be Affordable and the system must provide only the useful information to the Administrator so that the necessary action can be taken immediately. Here the two important mechanism used are motion detection and face detection for motion detection we

use background subtraction method and for face detection we use Viola-Jones Haar-Cascade method, here it's taken care about privacy preservation of the people and with the advanced alerting system such as emailing administrator on an arrival and live real time streaming of audio/video helps an immediate possible precautionary measures can be taken, here another important characteristic is that no continuous monitoring of camera footages is required and no wastage of storage medium occurs where as to alerting less bandwidth and bandwidth is utilized only during the time of interruption of someone into the system implemented area. This kind of surveillance system can be implemented for safe guarding houses, school, colleges, private properties, restricted areas, etc..and in future this work can be extended to detect faces of people and with cloud database integration the identification of the person by name or any specific characteristic can be implemented and their by using different technique it can also be implemented such that if any suspicious activities performed by an person can be distinguished with other activates and immediately alert can be sent to an concerned authority to perform an suitable action.

REFERENCES

[1] Michael Bramberger , Roman P. Pflugfelder, Arnold Maier Bernhard Rinner, Bernhard Strobl, Helmut Schwabach. A Smart Camera for Traffic Surveillance.
[2] Sharad Mehrotra, Alfred Kobsa, Nalini Venkatasubramanian. TIPPERS: A privacy cognizant IoT environment.
[3] Mohamed F. Abd-el-Kader, Rama Chellappa, Qinfen Zheng. Integrated Motion Detection and Tracking for Visual Surveillance.
[4] Nan Lu, Jihong Wang, Q.H. Wu and Li Yang. An Improved Motion Detection Method for Real-Time Surveillance
[5] Weiming Hu, Tieniu Tan, Fellow, IEEE, Liang Wang, and Steve Maybank. A Survey on Visual Surveillance of Object Motion and Behaviors.
[6] Brahmanandha Prabhu R, Arul Prabhar A, Garima bhora, Implementation of webcam based system for surveillance monitoring.
[7] Ms. Moharil R. S, Dr. Mrs Patil S. B, Implementation of Surveillance Monitoring System
[8] Ms. Renuka Chuimurkar, Prof. Vijay Bagdi. Smart Surveillance Security &Monitoring System Using Raspberry PI and PIR Sensor
[9] S. Khawandi, B. Daya, P.Chauvet Implementation Of An Intelligent Surveillance System For Elderly
[10] Yi-Qing Wang. An Analysis of the Viola-Jones Face Detection Algorithm
[11] Arwa Darwish Alzughaibi, Hanadi Ahmed Hakami, Zenon Chaczko. Review of Human Motion Detection based on Background Subtraction Technique.