

# AUTOMATED HUMAN DETECTION AND TRACKING FOR SURVEILLANCE APPLICATIONS

Avaneesh Tripathi, Milind Pathak, Amritha Sharma, Parth Vijay and S.A. Jain

*School of Computer Engineering and Technology, MIT Academy of Engineering, India*

## Abstract

*Safeguarding a country's border is very vital to its nation. A nation's security is determined by the kind of security imposed at the border. Proper surveillance system needs to be imposed at border because of numerous reasons which include drug smuggling, illegal immigrants crossing the borders and last but not the least, terrorist intrusion. To protect our border from these activities we need to monitor, filter and detect if there is any motion or any kind of activity that might lead to intrusion in our borders. For this we need 24x7, surveillance and hence an automated surveillance system needs to be developed. The existing systems are not efficient enough to detect threats if the intruders are using some kind of camouflage or cover. To overcome these drawbacks of the existing systems we intend to use thermal imaging cameras. The focus is to use thermal imaging cameras to capture the images for detection and tracking of the human/object, so as to analyze and predict their behaviour and intentions. The use of thermal imaging unlike the optical camera will help the system to perform detection of suspects which are hidden behind an obstacle for e.g. a bush. The thermal imaging will perform the detection at night with the same efficiency as in the daytime. The surveillance system is designed in order to achieve the objective of safeguarding the border, thus satisfying the aim of detecting threats and minimizing false alarms.*

## Keywords:

*Image Processing, Thermal Imagers, Machine Learning, Border Surveillance, Neural Networks*

## 1. INTRODUCTION

Security at borders has always been a major concern for a country. The surveillance task is not suitable for a human as it requires concentration over a very long period of time. Hence automated surveillance systems are developed with the view of detecting potential threats which include drug smuggling, illegal immigrants and terrorists. Automated surveillance systems are gaining importance because of the increased level of security that is achieved on its implementation. The automated surveillance system enables remote monitoring which means the system can be implemented anywhere in this world i.e. on the mountains, sea banks, plains etc. The objective is to make these systems more and more efficient [5]. Such systems play a major role in detection of any kind of intrusion and thus are used for border security. Thermal imaging camera is used because thermal cameras are not dependent on light rather they are dependent upon the heat emitted by a body. Since the thermal cameras are not dependent on light they can be used in day as well as night. The video sequence obtained from this thermal imaging camera is processed and the activity of the moving object is tracked and analyzed [6]. If the analyzed activity is found to be malicious, the system displays a detected intrusion message. The existing automated surveillance systems do not provide desired results when the same object or person exhibits multiple changes in its behavior within

a very short span of time. In this case it becomes very difficult to analyze the activity and to label it as a threat or not. Also, it is easy to track the activity of a single person/object but tracking multiple humans or objects is a challenging task [7].

## 1.1 THERMAL IMAGING CAMERAS

Thermal imaging cameras have an ability to detect the electromagnetic radiations which are in the infrared zone of the spectrum and are emitted by all the objects as a function of its temperature [8]. The main advantage of using thermal imaging camera is that it can help in detecting activity/motion through smoke, darkness and heat permeable obstacles. The sensor information obtained from thermal camera gives a very distinctive profile of the person in the thermal image. This distinctive profile which is obtained helps into easier differentiation and labeling of the objects [9].

## 2. RELATED WORK

### 2.1 AUTOMATIC MULTIPLE HUMAN DETECTION AND TRACKING FOR VISUAL SURVEILLANCE SYSTEM

Kushwaha et al. [1] focuses on the problem of detecting human automatically and tracking them for visual surveillance. This paper uses trained human detector for detecting the objects and particle filters are used for tracking them. Several steps and techniques have been used to accomplish the goals.

#### 2.1.1 Sample Creation and Human Detector Training:

- Object detection is done using machine learning approach which is based on Haar-like features. While training the human detector binary adaptive boosting is to speed up the process. 2000 positive images (contains human) and 2700 negative images (non-human) are collected to train the detector.
- After the collection of the samples, next step is Haar-like feature extraction from the samples. Rectangular Haar-like features and integral image representation has been used for fast feature extraction. Haar-like features are used because; as the data sets become large it is going to be more difficult and complex to extract the features from the dataset. Integral image facilitates easy and fast calculation of features hence reducing the computation effort. Image dataset which consist of different images of human beings (different poses, sizes etc.) are used for training.

#### 2.1.2 Particle Filter Creation for Tracking:

After detection, tracking comes in the picture. For this particle filter has been used. It is having capability to represent a non-linear object tracking system in presence of non- Gaussian nature

of noise. Multiple objects can be tracked in two ways, first by creating multiple particle filters for each track and another one is by having a single particle filter for all tracks. In particle filter set of weighted particles are used to represent the uncertainty in humans state. Using a motion model, filter reproduces particles from previous frame to current frame, computes a weight for each propagated particle using an appearance model, then re-samples the particles according to their weights. Steps involved in this process are prediction, Likelihood and re-sampling.

### 2.1.3 Appearance Model:

Color Histogram is used to represent the newly detected human object. Likelihood of the particles is calculated through this histogram in the future frames.

### 2.1.4 Advantages:

- Training of classifier is less complex as it uses Haar-like features.
- Binary Boosting has been used which increases the speed of the performance.

### 2.1.5 Disadvantages:

- The system fails at night as it is using optical cameras.

## 2.2 REAL-TIME HUMAN MOTION DETECTION AND TRACKING

Zarka et al. [11] focuses on the problem of detecting human and tracking them at real time. The system will be able to detect human both in outdoor as well as indoor environment. Several steps which have been used to achieve this are: Background modeling, Foreground and feature extraction, Object tracking, human modeling, and Human motion analysis.

### 2.2.1 Background Modeling:

This is done in two steps; in first we will obtain a background image which is having non-moving objects. This can be done through Highest Redundancy Ratio (HRR) algorithm. Second is to maintain this background against the changes such as illumination change, motion change etc.

### 2.2.2 Foreground and Feature Extraction:

Background subtraction method is used for foreground region extraction. From the current frame the background image is subtracted to get foreground pixels. After this, position of object is detected in the foreground image. To achieve this, we use 2D image of foreground by its 1D projections histogram which help us in developing the boundary box for the objects. Size filtering is applied to remove the small objects. Images which are produced through object detection may have holes or gaps which lead to split the object in to two or more parts hence making single object to appear as multiple object. Two dilation's and one erosion are applied to store the object to its current state. Good features are extracted for tracking. Centre of gravity gives the object location in the frame, velocity of object is used to predict the next position of the object, and Number of foreground pixel is obtained through object size.

### 2.2.3 Object Tracking:

Object which are entered or exited the scene are detected. Occlusion is detected by checking the boundary box of the old and new regions groups.

### 2.2.4 Human Modelling:

Star Skeletonization in which a star like representation is used to represent the human object. Skeletonization can be done by applying different steps of abstraction on the Object boundary.

### 2.2.5 Human Motion Analysis:

By using star representation analysis of human motion becomes much faster. The lower extremal points in star are considered as legs and uppermost will be considered as head, angle theta is taken between the vertical line and centroid and at last frequency of theta is seen to detect the motion.

### 2.2.6 Advantages:

- Star skeleton method has been used which has low computation cost and is robust when the dealing with noise.
- Adaptive background model has been used which can adapt the different conditions which occur in the environment hence making the system to work smoothly.

### 2.2.7 Disadvantages:

- Camera has to be fixed at one place while tracking hence the system fails in some condition.
- Different kind of activities such as jumping, falling etc. cannot be detected through this system.
- Other objects such as vehicles, animals cannot be detected.

## 2.3 MOTION DETECTION, TRACKING AND CLASSIFICATION FOR AUTOMATED VIDEO SURVEILLANCE

Gabal et al. [2] focuses on automatically detecting and tracking motion of an object to extract information from a video based surveillance system. The algorithm used in this paper tries to achieve complete detection of a moving object in changing brightness, variations in environment and noise in background.

- *Initial Background Modeling and Background Subtraction and Normalized Cut Segmentation:* Current background model value is calculated by taking an average of selected frames using modified moving average to get initial background model. Back- ground subtraction is performed by comparing current frame with the estimated background. To minimize the similarity of pixels that are to be split we use segmentation based on finding minimum cuts in a graph.
- *Tracking of the Detected Object:* Tracking of a detected object is used to predict the objects path moving in the scene. The method proposed in the paper only focuses on the foreground pixels only. It uses optical flow method to distinguish between ghost and abandoned objects along with tracking and detecting stationary and moving objects.
- *Update Background Model:* A simple moving average method is used to smooth the background model even with changes in the scene with respect to video noise, illumination, moved background objects etc. This model is constantly updated if any changes hap- pen in the background.

### 2.3.1 Advantages:

- This system requires next to no human interference.
- It uses optical flow method which is capable of tracking stationary as well as moving objects.

- Using background subtraction along with segmentation has reduced processing requirements and reduced processing time.

### 2.3.2 Disadvantages:

- The detected object can't be classified into categories of humans or vehicles. The system needs to have a fixed camera for background modeling.

## 2.4 MOVING HUMAN BODY DETECTION IN VIDEO SEQUENCES

Li et al. [3] focuses on one of the sub-problems of object detection i.e. human body detection in the static camera, no crowd outdoor environment, and end long bodies. Humans exhibit some definite patterns in their motion and on the basis of the motion exhibited on the basis of which the human is recognized. The paper aims at using the human body detection algorithm based on the combination of moving and shape information of an object. Firstly, the moving object is detected by the Eigen objects computed by the three frames of the first video. Eigen object is a kind of geometrical features that encompass the shape, location and motion information of the original image sequences. The paper is aimed at achieving the detection and tracking of a human using Eigen object and moving object segmentation methods. To achieve this we use the human body detection algorithm which is based on the combination of moving information with shape information.

In this paper, two methods are used for an object detection which is Eigen-object and moving Object segmentation method. In the method which is based on Eigen-object, we first compute the Eigen-object from three frames in the initial video sequences which are used to detect the moving object and then the shape information is used to classify human and different objects. In order to improve the result, we use the object segmentation method which is also based on the Eigen-object and this method is used in large applications where an object appears in a largely stable background. Eigen-object is computed by the Principal Component Analysis from the original image sequences. By using this algorithm the shadow of a moving object will also be eliminated. Moving object detection is considered as a classification problem that classifies the objects in image sequences and this classification is achieved by eigenvectors. In moving object extraction, we use Sobel operator to extract the edge of Eigen-object and then we use a morphological operator to perform fill and smoothness. There are two approaches for classifying objects namely Shape-based Classification and Motion-based classification.

### 2.4.1 Advantages:

- The advantage of the algorithm is that the moving object detection is accurate, and the result does not effect by body pose.
- The shadow of a moving object is also eliminated in the detection step.
- The algorithm can also detect human bodies in complex situations.

### 2.4.2 Disadvantages:

- The method is capable of tracking a human in the short distance.

- Chances of inconsistencies will increase in case of multiple human detections.
- When the camera lens is not stable without tingle then we face the problem of noise.

## 2.5 HUMAN ACTION RECOGNITION FOR NIGHT VISION USING TEMPORAL TEMPLATES WITH INFRARED THERMAL CAMERA

Eum et al. [4] focuses on the problem of detecting human action at night. The paper aims at using infrared thermal camera for video surveillance. Temporal templates, HoG and SVM are used to achieve our goals.

- *Segmentation*: The outline of the object is extracted through the background subtraction method. Already saved background image is subtracted from current input image. After background subtraction the image which is formed also contains noise, so it becomes difficult to extract features from this image. So the image is converted into a binary image for faster evaluation.
- *Feature Extraction*: Motion History Image (MHI) is used against series of video images. MHI requires less computation to recognize which action has occurred in the image. Median filter are used to remove the noise from the obtained MHI image. Histogram and oriented Gradients (HoG) are used to detect the object.
- *Classification*: Classification is done through support vector machine (SVM). SVM is type of supervised learning. Hyperplane is chosen to differentiate between different classes. Centroid is calculated for each class and outliers are calculated for that class.

### 2.5.1 Advantages:

- MHI is used which make the system much faster than any other system.
- Due to thermal camera human can be detected even if they are hiding behind trees or bushes.
- SVM are used which makes the classification much faster.

### 2.5.2 Disadvantages:

- The system is using fixed camera for background modelling.
- This system always required a background image without any objects in it.

## 2.6 VISION APPROACH OF HUMAN DETECTION AND TRACKING USING FOCUS TRACING ANALYSIS

Sanoj et al. [10] focuses on one of the sub-problems of object detection i.e. human motion detection. Humans exhibit some definite patterns in their motion and on the basis of the motion exhibited on the basis of which the human is recognized. The paper aims at using thermal imaging based video surveillance systems. The five main tasks of detecting and tracking human are identified which are data acquisition, pre-processing, target detection and object representation. The paper is aimed at achieving detection and tracking of human using computer vision approach. To achieve this a thermal imaging camera is used, the advantage of using a thermal imaging camera is that it can be used over a wide variety of night vision applications. FLIR (Forward

Looking Infrared) Thermal Imaging cameras can be used to perform detection in complete darkness, through smoke and light fog. The system in this case consists of five important modules which are Data Acquisition, Preprocessing, target detection, tracking target and object representation. Image acquisition is done by capturing the thermal images from the camera. The second stage is pre-processing which involves two sub processes which are filtering and thresholding. In filtering process, we use a Prewitt operator to measure the features of an image. The Prewitt operator component is calculated with kernel  $k_x$  and horizontal component of edges with kernel  $k_y$ . After filtering, thresholding is carried out wherein a grayscale image obtained from the thermal imaging camera is converted into a binary image using a threshold value. Adaptive threshold is used when different threshold is used for different regions. The third stage is the Human Detection where the moving object (human) has to be detected. Moving target detection can be carried out by following three different approaches which include temporal differencing, background subtraction and optical flow. In this paper, we are using background subtraction. Background subtraction is the process of modeling the background and determining the foreground by comparison of the frames in sequence. The next stage is human tracking, where Kalman filter is used. The advantage of using sensor information from a thermal camera is that a person in the thermal image has a very distinctive profile so that the person can be clearly separated from the background. The video focus is also analyzed. The system can be further enhanced by incorporating into it technologies which will make it function in case of plural people and in case of long distances.

#### 2.6.1 Advantages:

- The use of thermal imaging cameras makes the application of the system very wide especially in case of night vision applications. Also the sensor information obtained from thermal camera has a very distinctive profile of the person in the thermal image. Because of this distinctive profile, the person can be easily subtracted from the image by using background subtraction.
- The use of Canny edge detector for edge detection has an advantage that it uses multiple threshold values.

#### 2.6.2 Disadvantages:

- The method is capable of tracking a human in the short distance.
- Chances of inconsistencies will increase in case of multiple human detection.

### 2.7 AN AUTOMATIC MOVING OBJECT DETECTION ALGORITHM FOR VIDEO SURVEILLANCE APPLICATIONS

The paper deals with the algorithm that is used in automatic detection of a moving object. The paper describes how frame difference and region combination when used together can result into the detection of the moving object. Frame difference method is chosen as it has less computational complexities as compared to the rest of the methods such as background subtraction, optical flow etc. The frame obtained from the video sequence first undergoes 2D median filtering and then one of the color components from the RGB model is selected for further processing. The next step is the selection of a threshold value,

adaptive threshold method is used. The frame difference i.e. the difference between the position of one pixel in two different frames is calculated. Morphological operations are applied to make the moving pixels continuous and filling the isolated pixels. If the distance between the position of a pixel in different frames is below threshold value, the regions are combined and if it is greater than threshold value motion is detected.

#### 2.7.1 Advantages:

- Frame difference method involves very less computational complexities as compared to other algorithms.
- The major aim of frame difference algorithm is that it focuses on differentiating the static and the moving objects hence is very less sensitive to light changes.

#### 2.7.2 Disadvantages:

- The whole efficiency of the system depends upon the calculation of threshold value.

### 3. ALGORITHM

**Step 1:** START

**Step 2:** Get positions of camera i.e its longitude and latitude values.

**Step 3:** Obtain the horizontal and vertical camera angles.

**Step 4:** From the incoming video stream store a base frame that has no moving objects.

**Step 5:** Maintain the stored background image with the changes happening in the video.

**Step 6:** Obtain the silhouette of objects from the current frame using background subtraction.

**Step 7:** Create bounding box around each of the detected objects.

**Step 8:** Using bounding box check whether some new objects has entered or exited the scene.

**Step 9:** Using Haar-cascade check whether the detected object is a human or not.

**Step 10:** if human detected

**Step 11:** Detect action using Star Skeletonization method if suspicious action is detected

**Step 12:** Beep alarm else

**Step 13:** GOTO step 5

**Step 14:** END

#### 3.1 PSEUDO CODE

##### 3.1.1 Function for Bounding Box Coordinates:

left is equal to (box x-coordinate - box width)/2 × image width

right is equal to (box x-coordinate + box width/2) × image width

top is equal to (box y-coordinate - box height/2) × image height

bot is equal to (box y-coordinate + box height/2) × image height

IF left is less than zero left = 0

IF right is greater than image width-1 right = image width - 1

IF top is less than 0 top = 0

IF bot is greater than image height - 1 bot = image height - 1

```
//Print bounding box values print Bounding Box coordinates
"Left, Top, Right, Bottom"
```

```
OPEN output file in append mode IF file not found
```

```
print "Error opening file!"
```

```
print in file Mid Points in Pixels, Class and Accuracy
```

```
IF the Class in output file matches any entry in file classes
```

```
print in output file
```

### 3.1.2 Function for Real Time Detection:

```
function demo_for_human_detection()
```

```
IF filename is valid
```

```
open video stream(video_filename) ELSE IF
```

```
open webcam
```

```
ELSE
```

```
Error "Couldn't connect to webcam" Initialize buffer, buffer
letter and count=0
```

```
IF no predefined window found
```

```
make window "Demo" of 1352*1013 pixels, fullscreen
Initialize demo_time = current time
```

```
End
```

## 4. PROPOSED METHOD

The desired result is obtained when the tracking, analyzing and the identification of the object is done properly along with the proper filtering of the images which are obtained from the video sequence. Firstly a background free of moving objects is obtained. This frame is stored as a permanent background and it's maintained against frequent changes. All the new frames are compared to the permanent background in order to get a silhouette. This can be done by various methods and is known as object detection. After the silhouette is obtained, tracking of objects comes in the picture. Tracking helps in determining whether the object has entered or exited from the scene or occlusion has occurred. Haar-Cascade method is used to identify the Object as human. Once the Object is identified as a human the system moves forward i.e. the activity done by the human is recognized and can be predicted.

### 4.1 BACKGROUND MODELLING

Main aim of background modeling is to capture a background that is free of moving objects in a video. This is basically done in two steps: I. Background Initialisation, this is used to obtain a background image at any specific time of the video. II. Background Maintenance, it updates the background with respect to changes in the real scene. Algorithms like Median or Highest Redundancy Ratio (HRR) can be used for background model initialization. The main drawback of the Median algorithm is that it's based on an assumption that during the training sequence every pixel of the background should be visible at fifty percent of the time or more. This means it can produce wrong results if any of the object that is moving stops for more than fifty percent of the training time. In order to overcome this drawback we switched to a more flexible algorithm i.e. High Redundancy Reduction (HRR). HRR works on the intensity of the pixel. It considers intensity values of the pixels which are extracted from the training

sequence. The pixels' intensity which are having the highest redundancy ratio are considered as a part of background image. Hence HRR algorithm is used over the median algorithm. Background Model Maintenance is done because there are changes that keep occurring which needs to be incorporated as changes to the background scene. The changes can reflect from various sources like illumination changes, motion changes etc. Sometimes, the changes can be because of sudden arrival of an object into the frame. The background model should be maintained in such a way that it is insensitive to the kind of changes mentioned. The model should be able to differentiate a static Object from a moving object.

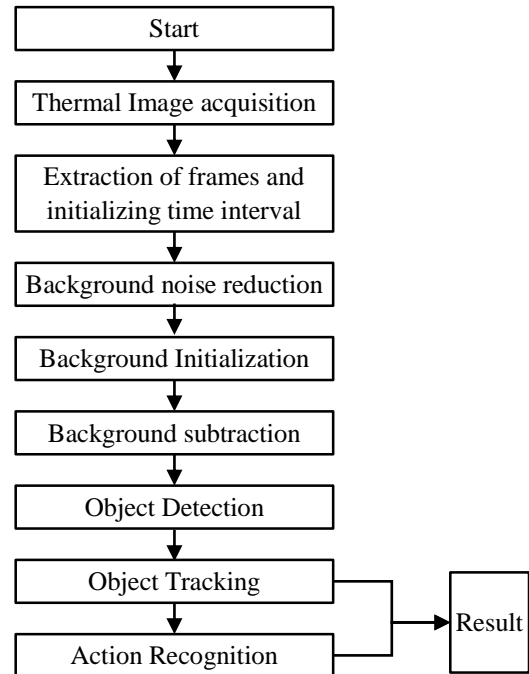


Fig.1. Architecture of Proposed System

### 4.2 OBJECT DETECTION

There are many kinds of objects in the environment. We are going to develop a model which will be detecting objects i.e. humans in real time. For object detection we have various techniques available which include segmentation, frame difference method and optical flow method. Out of these three, segmentation is chosen. Segmentation is done in two steps:

#### 4.2.1 Background Subtraction:

Background subtraction which is the process of modeling the background and determining the foreground by comparison of the frames in sequence. In background subtraction a permanent background is chosen and stored in the system. This system uses Least Median of Squares method to estimate the motion going on in the background. In this a permanent background is chosen and stored in the system which is then subtracted from the current input image to get the silhouette.

#### 4.2.2 Morphological Operation:

After performing background subtraction we get a silhouette. It becomes a bit difficult to do different operations on this image as it has slight noise in it. Hence this image is converted into binary image to reduce our computational complexity time and

thus making our work much easier. There are several morphological functions, out of which erosion, dilation and hit-or-miss are the primary operations. Continuous moving regions are obtained using these operations. Sometimes these images contain holes and isolated pixels of the mask which can be removed using DILATE, FILL, and CLOSE operations in sequence.

### 4.3 OBJECT TRACKING

In the field of computer vision, object tracking is considered as one of the important task. It helps to analyze the motion of object. Using Object Tracking it is easy to track what objects have entered or exited the current active scene. It is impossible to determine the object activities unless they are tracked. For tracking, first Region of Interests (ROI) are identified with the use of bounding boxes. Different algorithms can be easily applied on the objects once we find the appropriate bounding box which is drawn around them. Most of the system uses Kalman filter to track the objects position by comparing the position of their bounding box in the previous frame to their position in the current frame. But this algorithm fails when occlusion occurs. Hence another approach which we have used works efficiently in different condition like occlusion. While tracking different situations like, if the numbers of objects in the current scene are less than the number of objects in the previous scene or there were more objects in the previous scene than the current scene may occur. This happens because some objects have either exited or entered the scene or occlusion may have occurred. Hence there is a need to make the connections or form relationship between the objects position in the current and previous scene. These problems are solved using Object tracking. Three important tasks which are done through tracking are:

#### 4.3.1 Detection of Exiting Object:

Centroid of object in the previous scene which don't have correspondence in the current scene are estimated by using the velocity and previous position of the object. The coordinates of the centroid are stored and then they are compared with coordinates of scene. If any of the coordinates is outside or larger than the scene then the object is considered as exiting object.



Fig.2. Detection of Exiting Object

#### 4.3.2 Detection of Entering Object:

To decide if an object has entered the scene, we consider the edges of the bounding box of those objects which didn't have any correspondence in the previous scene. If any of the edges of the bounding box overlaps with the edge of the scene then the object is considered as an entering object.



Fig.3. Detection of Entering Object

#### 4.3.3 Detection of Occlusion:

Bounding Box for object is predicted if any edge of the box intersects with another object bounding box then occlusion occurs. First object will be occluded by the second object forming a different bounding box for both the objects. By automatic camera calibration we are able to extract lines from the image and detect the presence of humans. These extracted lines are given as parameters for camera calibration. Now that we know the number of humans present in the image we perform depth estimation. Using the information provided by depth estimation we detect whether occlusion has occurred.

### 4.4 ACTION RECOGNITION

This is the last and the most important task in the field of computer vision. Activity recognition is done after we have tracked the object. It helps to predict the objects motion and identify its action. Before moving to Activity recognition it is important to detect whether the object is a human or not. There are many algorithms to detect if the object is a human or not, but as the number of objects in the scene increases the complexity increases and thus these algorithms fail to recognize multiple objects. Hence to overcome this problem we use an efficient algorithm known as Haar Cascade.

In HAAR Cascade method detection of object is done through Haar-like features; these features are combined with boosted cascade classifier. As it is a combination of two things therefore it is known as Haar-cascade method. Steps to detect if an object is human or not are:

#### 4.4.1 Haar-like Features:

It is a 2-D function which provides fast extraction methods. A template is formed by using multiple rectangles placed adjacent to each other.

#### 4.4.2 Integral Image:

Integral Image is a technique in which each pixel value is changed to another value to form a new image representation. This newly formed image representation is known as Integral Image. It is done for faster calculation of haar-like features.

#### 4.4.3 AdaBoost:

While detection many kind of features having different possess, scale and position come into picture, making the system slow while performing these calculations. For the system to work efficiently AdaBoost algorithm is used. AdaBoost removes irrelevant features by choosing the best features which have the highest boundaries between the object and non-Object.

#### 4.4.4 Cascade Classifier:

Speed of the detection is increased by using Cascade Classifier method. In this method different classifier are combined together that work in a multilevel structure. Sub images are given as an input to classifier and if the image is containing the characteristics of object than the process is continued using further classifier otherwise the image is discarded. Once the object is detected the system moves further i.e. to activity recognition. There are many algorithms which are used to detect the activity but all these algorithms are computationally expensive and fails as the complexity increases.

To overcome this problem a new algorithm Star Skeletonization is used which has low computational cost and is more efficient when the scene is containing noise. Star Skeletonization is a star like representation of human object that resembles five main parts of the human body (legs, hands and head). The bounding contour is unwrapped into a Euclidean distance signal for defining the position of human body (extremal points). These points are the distance between objects centroid and point in the objects contour. The distance signal may be uneven so in order to make it even, median filter is applied. Zero crossing in the amplitude values of the distance signal are used for extracting the local maxima. After extracting the entire local maxima star skeleton is formed by joining each local maxima and the object centroid.

Once the Star Skeleton is formed it becomes easier to determine the gait of a human. Cyclic motion of the skeleton components is analyzed for analyzing the gait of a person. For a person who is in the standing position, the lower extremal points will be assumed as legs and head will be represented by the uppermost point. An angle  $\theta$  is defined between the vertical and the line which is connecting the centroid and the left-most lower extremal points. The  $\theta$  will change rapidly if the person is walking or running. Hence while walking the frequency will be higher. An angle  $\psi$  is defined between the line joining the centroid and the head and vertical. Whether the human is running, walking etc. is detected using this angle.

## 5. EXPERIMENTAL SETUP

Heterogeneous hardware components have been used for performing the experiments in order to enhance the performance of the system.

The components we used for our system consists of Intel Core i7 processor, 16GB of RAM, 8GB Nvidia 1050Ti GPU (Graphical Processing Unit). For real time analysis the camera used is FLIR A325sc. It has a lens of 4mm along with resolution of 320×240. The operating system used in our model is Ubuntu 18.04 LTS and the disk storage size required is 2GB. By using these hardware configuration we are able to detect human object in real-time at the rate of 30 FPS.

Here, we compared the accuracy of the system in human object detection. We have considered different scenarios such as partial images, blurry images, various lighting conditions, different poses and environment. The range of accuracy lies between 0 and 100%. All the hardware components used to perform the experiment have been mentioned above.

Table.1. Accuracy of the developed system in detecting human object in different environment

Various Conditions	Accuracy (%)
Standing	93
Lying	83
Walking	86
Exiting/Entering	80
Occlusion	85

The Table.1 shows the accuracy of the developed system in detecting human object in different environments.

## 6. CONCLUSION

The system designed achieves the aim of detecting human automatically and the activity that is performed by that human is successfully analyzed and detected. This system therefore can be termed as an efficient system and can be used for surveillance purpose. However, it has been observed that accuracy of the system decreases in case the human is leaving or entering the frame i.e. only partial human body is visible. Thus, by using this method we can implement an autonomous system fully capable of detecting humans and tracking their actions.

## REFERENCES

- [1] Alok K. Singh Kushwaha, Chandra Mani Sharma, Manish Khare, Rajneesh Kumar Srivastava and Ashish Khare, "Automatic Multiple Human Detection and Tracking for Visual Surveillance System", *Proceedings of International Conference on Informatics, Electronics and Vision*, pp. 12-19, 2012.
- [2] Neha Gabal, Neelam Barak and Shipra Aggarwal, "Motion Detection, Tracking and Classification for Automated Video Surveillance", *Proceedings of 1<sup>st</sup> IEEE International Conference on Power Electronics Intelligent Control and Energy Systems*, pp. 1-5, 2016.
- [3] Chun-Ming Li, Yu-Shan Li, Shu-Hai Wang and Xiu-Qing Zhang, "Moving Human Body Detection in Video Sequences", *Proceedings of 6<sup>th</sup> International Conference on Machine Learning and Cybernetics*, pp. 19-22, 2007.
- [4] Hyukmin Eum, Jeisung Lee, Changyong Yoon and Mignon Park, "Human Action Recognition for Night Vision using Temporal Templates with Infrared Thermal Camera", *Proceedings of 10<sup>th</sup> International Conference on Ubiquitous Robots and Ambient Intelligence*, pp. 1-5, 2012.
- [5] Xiaoshi Zheng, Yanling Zhao, Na Li and Huimin Wu, "An Automatic Moving Object Detection Algorithm for Video Surveillance Applications", *Proceedings of International Conference on Embedded Software and Systems*, pp. 21-27, 2009.
- [6] Amin Ullah, Jamil Ahmad, Khan Muhammad, Muhammad Sajjad and Sung Wook Baik, "Action Recognition in Video Sequences using Deep Bi-Directional LSTM With CNN Features", *IEEE Access*, Vol. 6, pp. 1155-1166, 2017.
- [7] Christian Herdianto Setjo and Faridah Balza Achmad, "Thermal Image Human Detection using HAAR-Cascade

- Classifier”, *Proceedings of International Annual Engineering Seminar*, pp. 1-3, 2017.
- [8] Palwasha Afsar, Paulo Cortez and Henrique Santos, “Automatic Human Action Recognition from Video using Hidden Markov Model”, *Proceedings of IEEE 18<sup>th</sup> International Conference on Computational Science and Engineering*, pp. 21-23, 2015.
- [9] Shipra Ojha and Sachin Sakhare, “Image Processing Techniques for Object Tracking in Video Surveillance- A Survey”, *Proceedings of IEEE International Conference on Pervasive Computing*, pp. 1-5, 2015.
- [10] C.S. Sanoj, N. Vijayaraj and D. Rajalakshmi “Vision Approach Of Human Detection and Tracking using Focus Tracing Analysis”, *Proceedings of IEEE International Conference on Information Communication and Embedded Systems*, pp. 12-17, 2013.
- [11] Nizar Zarka, Ziad Alhalah and Rada Deeb, “Real-Time Human Motion Detection and Tracking”, *Proceedings of International Conference on Information and Communication Technologies from Theory to Applications*, pp. 1-6, 2008.