INTERNATIONAL JOURNAL OF ADVANCE RESEARCH, IDEAS AND INNOVATIONS IN TECHNOLOGY

IJARIIT

ISSN: 2454-132X
Impact factor: 4.295
(Volume 5, Issue 3)
Available online at: www.ijariit.com

Automated human detection and tracking for surveillance applications

Amritha Sharma
amritha950@gmail.com
MIT Academy of Engineering, Pune, Maharashtra

Milind Pathak
milind0906@gmail.com
MIT Academy of Engineering, Pune, Maharashtra

Avaneesh Tripathi
avitri89@gmail.com
MIT Academy of Engineering, Pune, Maharashtra

Parth Vijay
parthvijay96@gmail.com
MIT Academy of Engineering, Pune, Maharashtra

Shitalkumar A Jain
sajain@comp.maepune.ac.in
MIT Academy of Engineering, Pune, Maharashtra

ABSTRACT

Automated surveillance systems are gaining importance because of their vast applications at the border while security is concerned. Various algorithms are developed and technologies are used to improve the efficiency of these surveillance systems. Efforts are being made to reduce the number of false alarms and detect any kind of suspicious activity happening in the region of suspicion within no seconds. These suspicious activities include drug smuggling, illegal immigrants crossing the borders and last but not the least, terrorist intrusion. These activities need to be detected and analyzed in order to conclude if the activity is suspicious enough to be classified as a threat. The existing systems deployed at the border are not efficient enough to detect threats and hence this paper is designed with an objective of presenting a better algorithm to make a better automated surveillance system. The sole purpose of this algorithm is to increase security at the border because safeguarding the border till date continues to remain a challenge to our country.

Keywords— *Image Processing, Thermal imagers, Machine learning, Border surveillance, Neural Networks, Star skeletonization*

1. INTRODUCTION

Security at the border has always been one of the greatest challenges to its country. It is very important to protect and safeguard our border from all kinds of malicious activities that might take place. These malicious activities if not identified at the right time can leave the whole country at risk. Nation's security depends mostly upon the kind of security that has been imposed at the border. One cannot rely on humans for surveillance because this kind of surveillance requires long hours of concentration. To overcome this drawback of manual systems which rely upon the humans, we use automated surveillance systems. The surveillance systems are designed in order to detect, analyze and generate an alarm if any suspicious activity is detected. The automated surveillance system's efficiency is increased can be used as an asset for the protection of the border and ultimately safeguarding the country. A better algorithm is the need of this hour. The algorithms that exist do not provide the desired results when two or three suspicious activities take place at the same time in the region of suspicion. The existing algorithms also fail when a single object starts demonstrating multiple behaviours. A system that overcomes these drawbacks can help to safeguard the country in the best possible way.

2. RELATED WORK

Manish Khare [1] used Haar-like features to train the image set. He then used this system to detect the humans and tracked them using particle filter. His system allowed to detect a human in poor lighting conditions, different shapes and sizes etc. His system can also detect multiple human objects in a video. Object detection is done using a 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, the next step is Haar-like feature extraction from the samples. Rectangular Haar-like features an 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 data-set. Integral image facilitates easy and fast calculation of features hence reducing the computation effort. Image data-set which consist of different images of human beings (different poses, sizes etc.) are used for training. After detection, tracking comes in the picture. For this particle filter has been used. It is having the 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 the previous frame to the 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. Color Histogram is used to represent the newly detected human object. Likelihood of the particles is calculated through this histogram in the future frames.

Nizar Zarka [2] created a system that can detect humans in both indoor as well as outdoor environments. He first designed an adaptive background model which is robust and can deal with different lighting conditions and object occlusions. He used a background subtraction method to get the foreground pixels. After obtaining the foreground pixels he applied noise cleaning and object detection techniques to detect the human objects. Once the human object was detected he used human modelling methods to recognize the activities performed by these detected objects.

Neha Gaba [3] used an advanced approach which is applied on videos for detecting motion of the objects. Using her system a complete moving object is detected which is vigorous even if the brightness varies continuously, surrounding environments and noises are dynamically varying. To build the model she used an approach which is dependent on pixel and non-parameterized and is associated with the first frame of the video. Once she captured the subsequent frame she detected the foreground which is used to represent the different detected objects and then its background. She then identified and removed the ghost objects using unique tracking methods. Shortcomings that were present in currently developed techniques were overcome by using a single set of variables.

Chun-Ming Li [4] proposed an algorithm which is used to detect the human body and is based on the information which is gained from the movement and shape of the human. He detected a moving object by using the Eigen object which is computed from the initial video frame. The human is differentiated from other objects by using their shape. Continuous multiple frames are used to check whether there is occlusion between two objects. Due to the elimination of shadows which are associated with objects minimizes their effect on the human body.

Hyukmin Eum [5] found a method to recognize human action at night. He used an infrared thermal camera which works on heat signatures emitted by the body instead of normal visual attributes. He preferred this camera as it was easier to detect human movement at night time.

Xiaoshi Zheng [6] proposed that one of the most important steps in video surveillance is the detection of moving object. These applications are achieved using frame difference algorithms. He first obtained the moving pixels of frames in the video by using an automatic threshold calculation algorithm. After that, he used morphological operation for the formation of moving regions. The distance which is the shortest between the two regions was calculated at last and this distance was used in region combination. This surveillance application is fully automatic and efficient.

Sequential multimedia data were processed using Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM). RNN and LSTM achieved a very good result in processing this type of data. These networks were used in recognizing the speech, processing of the video, digital signal processing and analyzing the text data. In this paper, Amin Ullah [7] found a method which is used for action recognition by using Convolutional Neural Network (CNN) and deep bidirectional LSTM (DB-LSTM) on the video frames. For minimizing the redundancy he used every sixth frame of the incoming video for extracting deep features. To learn the sequential information from the frame features and to increase the depth he used multiple layers which are combined together in both forward passes as well as the backward pass of DB-LSTM. This method is used in learning long term sequences and in the processing of large videos by evaluation of its features for a fixed interval of time. Action recognition is improved to a large extent by using this method.

3. OUR METHOD

The flow diagram in figure 1 depicts our system framework containing multiple stages. These stages start from taking frames of the video to the result that is suggesting an appropriate action in case of intrusion detection. Some of the major steps that we have taken in this paper include, YOLO which is used in order to detect human objects present in the frame and track their movement in the incoming video. The Haversine formula is used in calculating the position of an intruder with respect to the camera. Now the output of the Haversine formula is used in order to plot the exact location of an intruder on a map. If the intruder is detected inside of our territory then the system suggests appropriate actions and alert concerning authorities.

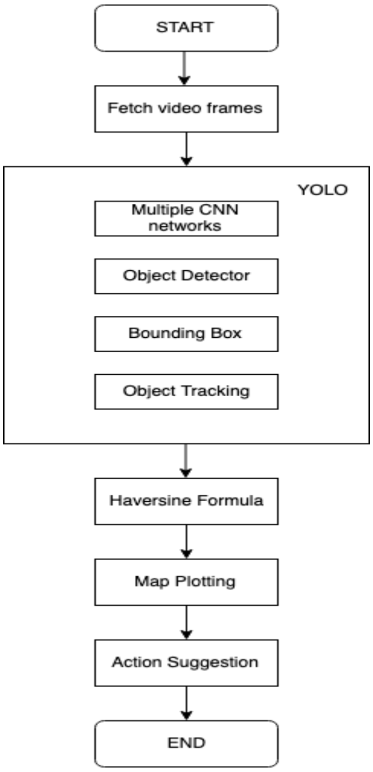

Fig. 1: Flow diagram

3.1 YOLO

YOLO stands for 'You Only Look Once' is a real-time object detection technique. In this, we have given one full frame to one single neural network. By using this network the image is divided into regions and for each region bounding box and probability is predicted. Weight is given to these bounding boxes by using these predicted probabilities. Since object from the camera is calculated at the starting and using these parameters along with the camera's coordinates we can calculate the exact location i.e. the latitude and longitude of the detected object from a given frame. Haversine is very fast in calculating the desired coordinates hence it becomes easy to plot the location of an intruder on the map. The simplified version of the Haversine formula [13]:

$$d = 2r \arcsin\left(\sqrt{\sin^2\left(\frac{\varphi_2 - \varphi_1}{2}\right) + \cos(\varphi_1)\cos(\varphi_2)\sin^2\left(\frac{\lambda_2 - \lambda_1}{2}\right)}\right)$$

It uses one single network hence it is much faster than other neural networks. Different objects are detected easily using YOLO. It helps us in two ways:

3.1.1 Problem of occlusion: By using YOLO we minimize with the problem of occlusion as it considers occluded objects as one single entity.

3.1.2 Object detection: Objects are detected uniquely and fast by using YOLO with the help of appropriate hardware components.

By using the bounding boxes we get the center coordinates of the objects which are used in Haversine. Figure 2 depicts the output of YOLO.

Fig. 2: Object is standing

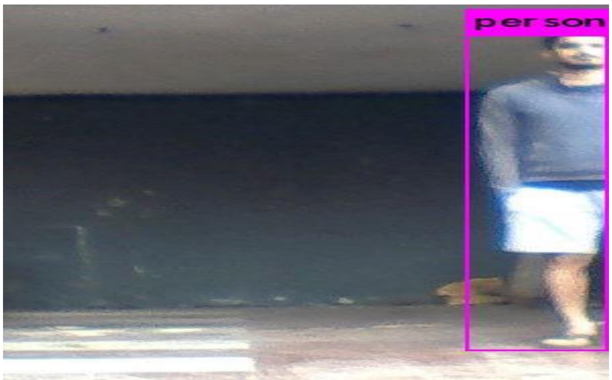

Fig. 3: Object is exiting from the scene

3.2 Haversine method

Here our aim is to obtain the coordinates that are longitude and latitude of the object. In order to do so, we have to get the location of the object with respect to the camera so that we can get the coordinates of the object and plot it on the map. For this Haversine formula has been used. Haversine is used on the surface of the sphere to get the shortest distance between two points. Range and the initial bearing of the

d = distance between the 2 points

r = 6371 km (Earth's radius)

$\phi 1$, $\phi 2$ = latitude of point 1 and latitude of point 2 (radians)

$\lambda 1$, $\lambda 2$ = longitude of point 1 and longitude of point 2 (radians)

3.3 Map

The map plays a vital part in this system. The exact location of the intruder is shown through the map and it is used to check whether the intruder is inside of our territory or not. Here we take the output from the Haversine formula that is coordinates of the object and by using these coordinates we are able to plot the correct position of an intruder on the map. The map is properly tested and built to avoid fake intrusion alerts. After every 2 seconds map is refreshed to get the current exact location of the intruder. Figure 3 shows the position of the camera and intruder in real time.

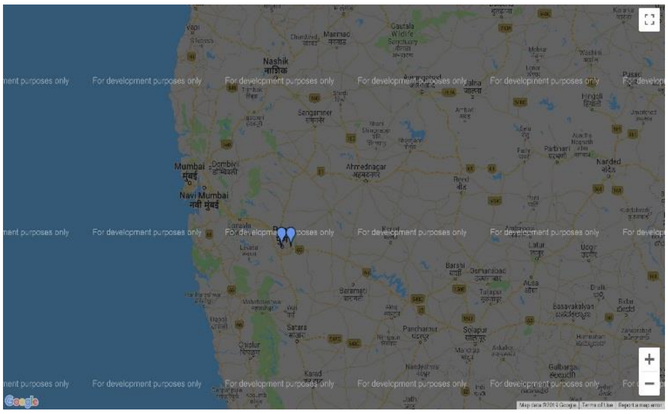

Fig. 4: Map showing positions of camera and intruder

3.4 Action suggested

The last and most important step of this system is to suggest appropriate action to the user when an intrusion activity is detected. If the intruder is inside of our territory then the system can sound the alarm and give the precise location of the intruder so that he can be apprehended without any complications. It also provides a list of appropriate actions to the user so that he can easily take the correct action and alert the authorities accordingly. Action algorithm is properly trained so that it doesn't raise fake alarms and suggest corrective actions.

4. EXPERIMENTAL SETUP

We have performed these experiments using hardware which are easily accessible. For enhancing the system performance we have used a homogeneous set of hardware components. The components needed for our system consists of Intel Core i7 processor, 16GB of RAM, 8GB Nvidia 1050Ti GPU (Graphical Processing Unit). The operating system used in our model is Ubuntu 16.04 LTS and the disk storage size required is 2GB. YOLO v3 is used for real-time object detection along with OpenCV-3.2.0 and CUDA 10. By using this hardware configuration we are able to detect a human object in real-time at the rate of 30 FPS.

5. RESULTS AND OBSERVATION

Here, we have compared the accuracy of the system in predicting human object detection in different conditions such as partial images, blurry images, different lighting conditions, varying poses and scenarios. The accuracy varies in mentioned conditions in the range of 0 to 100(%). As mentioned above to perform the experiment a homogeneous configuration of hardware is used.

Table 1: Accuracy in different environments

Different environment conditions / Poses	Accuracy (%)
Standing	99
Walking	97
Occlusion	94
Laying	90
Blurry	94
Exited / Entry	80

The figure 5 shows us the accuracy of our system in detecting a human object in different environments.

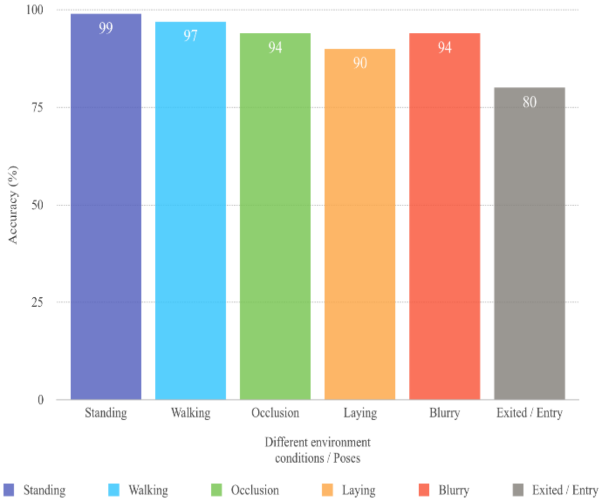

Fig. 5: Accuracy in different environments

6. CONCLUSION

In our paper, we have presented an approach which uses YOLO, OpenCV and CUDA coupled with haversine formula for the detection of human objects. The exact location of the human is plotted on the map. Moreover, we have checked the accuracy of our system for different environments. It has been observed that the object having occlusion and noise has less accuracy as compared to other objects. Thus, by using this system human detection can be done independently with minimal human intervention.

7. REFERENCES

[1] Maneesh Khare, Rajnesh Kumar Srivstav and Ashish Khare, Automatic Multiple Human Detection and Tracking for Visual Surveilance Sys- tem, IEEE/OSAIIAPR International Conference on Informatics, Electron- ICS Vision.

[2] Nizar Zarka, Ziad Alhalaha and Rada Deeb, Real-Time Human Motion Detection and Tracking, 2008 3rd International Conference on Information and Communication Technologies: From Theory to Applications.

[3] Neha Gabal, Neelam Barak and Shipra Aggarwal, Motion Detection, Tracking and Classification for Automated Video Surveillance, 1st IEEE International Conference on Power Electronics. Intelligent Control and Energy Systems (ICPEICES- 2016).

[4] Chun-Ming Li, Yu-Shan Li, Shu-Hai Wang, And Xiu-Qing Zhang Moving human body detection in video sequences, Proceedings of the Sixth International Conference on Machine Learning and Cybrnetics, Hong Kong, 22 August 2007.

[5] Hyukmin Eum, Jeisung Lee, Changyng Yon and Mignoon Parke Human Action Recogntion for Night Vision Using Temparal Templtes with In- frared Thermal Camera,2013 10th International Conference on Ubiquitus Rabots and Ambient Intelligence (URAI).

[6] Xiashi Zeng, Yaling Zhaao, Na lee, Huimean Woo An Autamatic Mov- ing Objct Dtection Algorithm for Video Surveilance Aplications,2009 International Conferences on Embedded Software and Systems 2009 International.

[7] AMIN ULAH, JAMIL AHMAD Action Recognition in Video Sequences using Deep Bi-Directional LSTM WithCNN Features, Digital Object Iden- tifier 10.1109/ACCESS.2017.2778011.

[8] Christian Herdianto Setjo, Balza Achmad, FaridahThermal Image Human Detection Using Haar-Cascade Classifier, 2017 7th International Annual Engineering Seminar (InAES), Yogyakarta, Indonesia.

[9] Palwasha Afsar, Paulo Cortez, Henrique Santos Automatic Human Action Recog- nition from Video using Hiden Markav Model,2015 IEEE 18th International Conference on Computational Science and Engineering.

[10] Sipra Ojha, Sachin Sakare Image processing techniques for object tracking in video surveillance- A survey,2015 International Conference on Perva- sive Computing (ICPC).

[11] Mr.C.S.Sanoj, Mr.N.Vijayaraj, Ms.D.Rajalakshmi Vision Aproach Of Human Detection And Tracking Using Focus Tracing Anlysis,2013 International Conference on Information Communication and Embed- ded Systems (ICICES).

[12] Christapher M. Bishap F.R.Eng Patern Recognition and Machine Learn- ing.

[13] Anang Suryna, Fareza Rynaldi, Firman Pratma, Gilang Ginanjar, Indra Indriansyh, Deni Hasman Implementation Of Haversine For- Mula On The Limitation Of E-Voting Radius Based On Android, 2018 Intrnational Conference on Computing, Engineering, and Design (ICCED).

[14] Cecep Nurul Alaam, Kherul Manaf, Aldy Rialdy Atmdja, Khrisna Arum Implementation of Haversine Formula for Counting Event Visitor in The Radius Based on Android Application,2016 4th International Conference on Cyber and IT Service Management.

[15] Hagar Mahmouud, Nadine Akari Shortest Path Calculation: A Com- parative Study for Location-Based Recommender System, 2016 World Sympousium on Computer Applications Research (WSCAR).

[16] Joseph Readmon, Santosh Divala, Ross Girshick, and Ali Farhaadi You Only Look Once: Unified, Real-Time Object Detection, 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR).

[17] Joseph Readmon, Ali Farhaadi YOLO9000: Better, Faster, Stronger, 2017 IEEE Conference on Computer Vision and Patern Recognition (CVPR).
Jing Taao, Hongbo Wang, Xinyu Zhang, Xiaoyu Lee, and Huawei Yang An Object Detection System Based on YOLO in Traffic Scene, 2017 6th International Conference on Computer Science and Network Technology (ICCSNT).