341

Smart Surveillance System for Abnormal Activity Detection Using CNN

Sathiyavathi V [a], Jessey M [a], Selvakumar K [b] and SaiRamesh L [c,1]

[a] *Department of Computer Science, JBAS College for Women, Chennai*
[b] *Department of Computer Applications, NIT, Trichy*
[c] *Department of Information Science and Technology, CEG Campus, Anna University, Chennai*

Abstract. Crime is the factor increases day by day and also needs the solution for identifying these activities in an efficient and quick manner. Many surveillance systems use artificial intelligence and image processing are incorporated with them to implement an intelligent surveillance system. But most of the systems are provided the alarm or identifies the crime after it happens. To solve this problem, camera footage-based theft detection will be used with the help of machine learning to detect theft occurrence. System will detect the human activity with the help of open-pose algorithm and convolution neural networks. Then the footage will be examined based on the pretrained model and it will be classified into three categories namely safe, abnormal or crime. Convolution neural network is used to classify the motion and an alert message will be sent to the owner along with captured image and options such as neglect or call the police.

Keywords. surveillance, convolution neural network, open pose, activity detection.

1. Introduction

In rapid movement of life, everyone can't get the time to monitor all the things happening around them. And everyone busy with their outside works and they don't even watch their own belongings or their house security in their daily routine activities. The only solution for this is continuous monitoring the behaviour or activities of people around as and to get the information about any abnormal activities. Smart surveillance is the solution for this which provides continuous monitoring of any one behaviour comes under the surveillance system. In general, surveillance system includes some electronic equipment like closed-circuit television (CCTV), sensors with alarms for identify unauthorized access, etc., In addition, surveillance system used for monitoring the behaviour of the people under surveillance and inform the admin about the behaviour of the people's come under the system installed for protection. Sometimes, this also used for negative purpose but we have to neglect that for further proceedings.

Many systems are in use currently to detect the crime activities before by identifies the new user to the place or by the identify the activity of the person which considers as

---

[1] SaiRamesh L, Department of IST, CEG Campus, Anna University, Chennai, India.
E-mail: sairamesh.ist@gmail.com.

abnormal one belonging to the place. Sometimes, false alarm also given by the system and that has to be avoided in the surveillance system. To overcome all these things, we need the system that provides continuous monitoring and detect the situation of crime before it happens. It also has not to give alarm publicly without identifies it properly.

The implementation of anti-theft environment with smart surveillance using CCTV which captures the live video and analysing the frames using machine learning approaches. This system gives the alarm to the admin if they found any suspicious activity which is not come under the normal behaviour of the people. The alarm sent with the captured image in which frame it find the activity was suspicious. The proposed system involves in real time crime detection which helps the people to protect their belongings in any kind of environment.

The overall objectives of this work are to detect the abnormal activity like theft in live video processing. This system also aims to notifying the respective person with image when user is online. Eliminating manual intervention and reduce the storage overhead for increasing security. The need for a system which automatically detects the crime and warns the admin / user is of great demand in current situation as in a fast-running world where trust is no more. The smart way is to get a instant notification about the abnormal activity in the place at absence of human. The existing system can detect the activity but not on real-time and it doesn't notify about the activity too.

2. Related Works

The literature survey was done regarding object detection, human activity detection and theft detection. Detection deals with the different objects which come in camera action with certain classes like immovable objects or movable objects like human, vehicles, animals, etc.,. The movement among the objects is detected through the frames and analysis carried out among them to identify the crime.

The system proposed by Bharath and Dhivya [1] detects moving object and it classifies and evaluates its parameters by alternating the algorithm in effective way. It tracks the object frame by frame and parameters like speed, velocity of the object are calculated. The system proposed by Edgar et al. [2] gives the systems that detect the overlapped objects in the frame. Their system considers both local and global cues using probabilistic top-down segmentation. Sharieff et al. [3] proposed the system for detect the hided image or object in the frame using inpainting technique.

Human activity detection deals with detecting various activities performed by human. Those activities include walking, talking, standing, sitting, doing some activities, etc., The system proposed by Xinyu et al. [4] can detect human abnormal behaviours from real-time video surveillance.

The system proposed by Gowsikhaa, Manjunath et al. [5] works on the sequence of frames form where the abnormal activity detected. The system proposed by Lazaros and Anastasios [6] helps to categorize the behaviour of the person is abnormal or not in a crowded environment. The system proposed by Debaditya and Krishna [7], detects snatch thefts using gaussian mixture model with a large number of mixtures known as universal attribute model to capture variations of attributes in diverse scenarios. The image retrieval is helps to extract the target objects from the available frames as mentioned [8,9].

The system proposed by Anjum and Babu [10], detects theft using raspberry pi, using image processing on live video to detect theft using motion and also to highlight the area where motion occurred. The system proposed by Munagekar [11] detects theft taking place in an enclosed environment. It uses canny edge detection algorithm for detecting objects and to prevent theft. If there is any variation in the count of pixel then the system detects there is an intrusion. Then it raises an alarm. Ayyasamy et al. [12] using the image processing techniques to locate or trace the object using IClique Cloak approach which helps to locate the disabled object.

The system proposed by Rupesh and Nupur [13] deals with automation of video surveillance in ATM machines and detects any potential crimes. The captured video is fragmented into smaller frames and then the vector graphics and image processing techniques are implemented.

3. System Design

The system takes in live-video or recorded video as input as shown in Figure 1. The first step is to split the video into frames and process it for human's detection. The detected number of humans are displayed along with the number of Frames per Second (FPS) being processed. Individual human activity is detected and the points specific to the particular human is measured. The predicted points are stored and is given to the trained
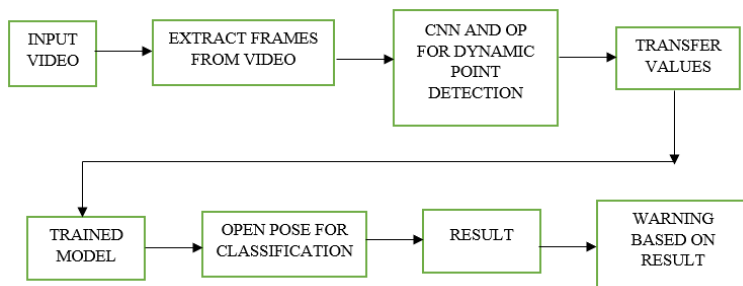
Figure 1. Architecture Diagram of Smart Surveillance for Theft Detection

model as input to detect the current activity based on the detected points it predicts the activity in the frame. At the end, the output on the screen is a classification of the video in real-time, a classification of the frame either safe, suspicious, or criminal activity. The system consists of the following modules:

- Input video Extract frames from video.
- CNN for image processing
- Open-pose algorithm for classification of activity Classified Result (Safe / Abnormal / Crime)
- Notifying the user / admin
- Sending the screenshot of the detected frame to user / admin

Each frame in the video will be taken and open-pose algorithm runs for human detection and then it plots the points on the human. The number of frames to be processed in a

second can be defined. Less number of frames per second leads to more accuracy in activity detection. CNN identifies the pose of the human through the plotted points by open pose through the model which is got by training the points of various poses and it's displayed for user / admin. Once the identified activity is found to be theft, then the user / admin will be warned about it through a text message to his verified mobile number in in case that he is not connected to internet. If the admin / user has a live connection then the screenshot the frame that is identified to be theft will be send to him.
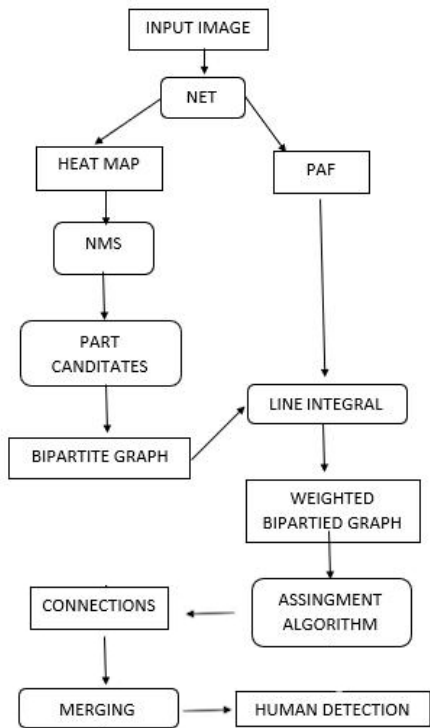
Figure 2. Flow of open-pose algorithm stimulation

4. Implementation Methodology

The dataset that is going to be trained in video format. A collection of videos for different types of theft is downloaded from online sources. The training phase takes data of about 2950 samples of about 5 classes. It also takes 329 samples to validate the data. The number of epochs given was 20. It shows, 20 number of times that the learning algorithm has gone through the entire training dataset. It also displays value accuracy, value loss at each step which are used to analyze the training phase and the efficiency. At the end of the training phase, a graph is generated to visualize the value loss, value accuracy, training accuracy and training loss.

4.1. Human Pose Estimation

Open pose is a human pose estimation algorithm. Figure 2 shows the flow of open – pose algorithm which gets input image and generate the key points based on heat map for the human body. The bipartite graph generated through the heat map merged with the assignment algorithm to get the human structure. Goal is to extract the position of each of the body parts of every person appearing in a frame with no more sensors than a digital camera. Open-Pose is a library that allow us to do so.

4.2. Parts and Pairs

The human body are the combination of parts such as neck, shoulder hip and arm. There is pairs in between the parts to shoe that they are joint according to the specific pairs as shown in Figure 3. These pairs are used to identify the position of the human parts while doing any activities and to recognize the activity. A body part is an element of the body, like neck, left shoulder or right hip. A pair is a connection between parts. These skeletons show the indices of parts and pairs on the COCO dataset.
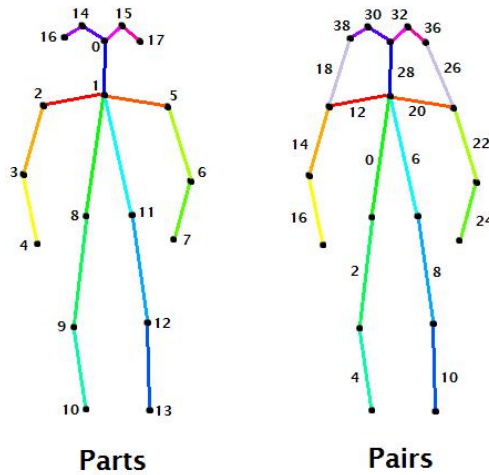

Figure 3. Difference Between Parts and Pairs

4.3. Pre-processing

In pre-processing each frame of the video is converted from [0, 255]to [-1, 1].

$$img = 1 + img(2:0) \tag{1}$$

4.4. Neural Network

The last operation of the neural network returns a tensor consisting of 57 matrices. However, this last operation is just a concatenation of two different tensors: heatmaps and Part Affinity Fields (PAFs). Here, heatmap stores the pixel values which one involved in pairs and parts of the body which given to the PAFs. From the associated PAFs the direction of the body part movement detected for activity detection.
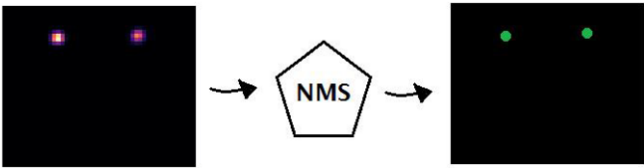
Figure 4. Non-Maximum Suppression

4.5. Non-Maximum Suppression

To extract parts locations out of a heatmap the local maximums should be needed. Then apply a non-maximum suppression (NMS) algorithm to get those peaks as shown in Figure 4. After all the process, the non-zero pixels denote the location of the part candidates.

Algorithm 1 Non-Maximum Suppression

Input: Blurred pixel images
Output: Sharp images

 1: Start in the first pixel of the heatmap.
 2: Surround the pixel with a window of side 5 and find the maximum value in that area.
 3: Substitute the value of the centre pixel for that maximum.
 4: Slide the window one pixel and repeat these steps after the entire heatmap is covered.
 5: Compare the result with the original heatmap. The pixels with same value are the peaks and suppress the other pixels by setting them with a value of zero.

4.6. Bipartite graph

For connection of identified body parts to form pairs graph theory should be used (bipartite graph) as given in Figure 5. To find the best matching between vertices of a bipartite graph assignment problem should be used, each edge on the graph should have a weight.
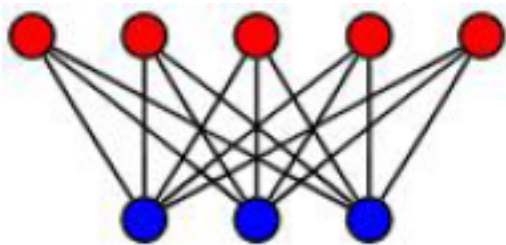

Figure 5. Bipartite graph for mapping body parts

4.7. Line Integral

The PAFs enter the pipeline. The line integral is computed along the segment connecting each couple of part candidates, over the corresponding PAFs (x and y) for that pair as

given in Figure 6. A line integral measures the effect of a given field (Part Affinity Fields) along a given curve (possible connections between part candidates). The line integral will give each connection a score, that will be saved in a weighted bipartite graph and helps to solve the assignment problem.
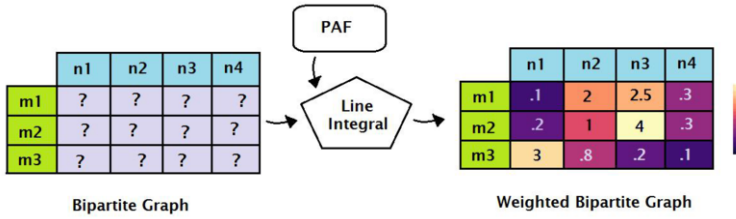

Figure 6. Updation of bipartite graph after line integral

Algorithm 2 Line Integral Algorithm

Input: An empty table of bipartite graph
Output: Weighted bipartite graph

1: dx = x2-x1
2: dy = y2-y1
3: normVec = $\sqrt{(dx^2 + dy^2)}$
4: vx = $\frac{dx}{normVec}$
5: vy = $\frac{dy}{normVec}$
6: numsamples = 10
7: xs = Create an evenly spaced numeric values (start=x1, end=x2) within an interval(dx/numsamples)
8: ys = Create an evenly spaced numeric values (start=y1, end=y2) within an interval(dy/numsamples)
9: pa f Xs = pafX[ys, xs]
10: pa fY s = pafY[ys, xs]
11: score = sum(pa f Xs*vx + pa fY s*vy) / numsample

4.8. Assignment

The weighted bipartite graph shows all possible connections between candidates of two parts, and holds a score for every connection. To find the connection the assignment problem should be solved.

4.9. Merging

The final step is to transform these detected connections into the final skeletons. Assume that every connection belongs to a different human. This way the number of humans is same as the connections. Let humans be a collection of sets H1, H2 and Hk. Each one of these sets that is, each human contains, at first, two parts (a pair). And let's describe a part as a tuple of an index, a coordinate in the x direction and a coordinate in the y

---

Algorithm 3 Assignment Suppression Algorithm


Input: Points of human detected.

Output: Pairs connected from point Sort each possible connection by its score.

1: The connection with the highest score is indeed a final connection.
2: Move to next possible connection. If no parts of this connection have been assigned to a final connection before, this is a final connection.
3: Repeat the step 3 until we are done.

---

direction. If humans H1 and H2 share a part index with the same coordinates, they are sharing the same part! H1 and H2 are, therefore, the same humans. So, merge both sets into H1 and remove H2. Continue for every couple of humans until no couple share a part.

5. Simulation Results & Discussions

The trained model produces a graph which shows the training accuracy, training loss, value accuracy and value loss as mentioned in Figure 7. The model takes input for ac-
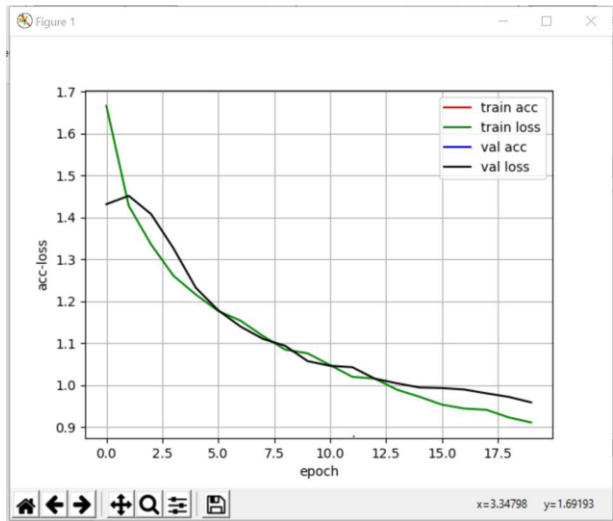
Figure 7. Analysis of Training

tivities likes walking, standing, sitting, operate and two member activity like handshaking, the train and a model is generated. The generated model is used to predict the activity. The generated model is loaded into the takes input video, and predicts and displays the type of activity it recognized in each frame in the live-video itself along with the frame number and the number of human detected in the frame. It recognizes activities like standing, walking, sitting, operate. It can also recognize two member interaction activity like handshake and theft activity. The result displays various body points for all the humans detected. Based on the points detected, the model recognizes which activity has been occurred in the frame. As the video proceeds, the result changes as per the recognized activity.

References

[1] Bharath RR, Dhivya G. Moving object detection, classification and its parametric evaluation. InInternational Conference on Information Communication and Embedded Systems (ICICES2014) 2014 Feb 27 (pp. 1-6). IEEE.

[2] Leibe B, Seemann E, Schiele B. Pedestrian detection in crowded scenes. In2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) 2005 Jun 20 (Vol. 1, pp. 878-885). IEEE.

[3] Sharieff AH, Sabena S, Sathiyavathi V, SaiRamesh L. Intelligent framework for joint data hiding and compression using SMVQ and fast local image in-painting. Int. J. Sci Technol. Res. 2020;9(2):2267-71.

[4] Wu X, Ou Y, Qian H, Xu Y. A detection system for human abnormal behavior. In2005 IEEE/RSJ International Conference on Intelligent Robots and Systems 2005 Aug 2 (pp. 1204-1208). IEEE.

[5] Gowsikhaa D, Abirami S. Suspicious Human Activity Detection from Surveillance Videos. International Journal on Internet & Distributed Computing Systems. 2012 Jul 1;2(2):141–8.

[6] Lazaridis L, Dimou A, Daras P. Abnormal behavior detection in crowded scenes using density heatmaps and optical flow. In2018 26th European Signal Processing Conference (EUSIPCO) 2018 Sep 3 (pp. 2060-2064). IEEE.

[7] Roy D. Snatch theft detection in unconstrained surveillance videos using action attribute modelling. Pattern Recognition Letters. 2018 Jun 1;108:56-61.

[8] Sabena S, Yogesh P, SaiRamesh L. Image retrieval using canopy and improved K mean clustering. InInternational conference on emerging technology trends (ICETT) 2011 (pp. 15-19).

[9] Raghuraman G, Sabena S, Sairamesh L. Image retrieval using relative location of multiple ROIS. Asian Journal of Information Technology. 2016;15(4):772-5.

[10] Anjum U, Babu B. Iot based theft detection using raspberry pi. Int. J. Adv. Res. Ideas Innov. Technol. 2017;3(6):131-4.

[11] Munagekar MS. Smart Surveillance system for theft detection using image processing. International Research Journal of Engineering and Technology. 2018 Aug;5(8):232–4.

[12] Ayyasamy A, Ramesh LS, Sathiyavathi V. Iclique Cloak Approach for Protecting Privacy of Mobile Location with Image Processing Techniques. InInternational Conference on Image Processing and Capsule Networks 2020 May 6 (pp. 160-168). Springer, Cham.

[13] Mandal R, Choudhury N. Automatic video surveillance for theft detection in ATM machines: An enhanced approach. In2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom) 2016 Mar 16 (pp. 2821-2826). IEEE.