IOSR Journal of Computer Engineering (IOSR-JCE)
e-ISSN: 2278-0661,p-ISSN: 2278-8727, Volume 22, Issue 3, Ser. II (May - June 2020), PP 20-25
www.iosrjournals.org

Vigilance system to track a person in real-time in a closed network of interconnected cameras

Dr. Jayshree Pansare[1], Mr. Reshikesh Dhanrale[2], Ms. Shriya Jamadarkhana[3]
Mr. Akshit Keoliya[4]
[1](Computer, MESCOE/ Savitribai Phule Pune University, India)
[2](Computer, MESCOE/ Savitribai Phule Pune University, India)
[3](Computer, MESCOE/ Savitribai Phule Pune University,India)
[4](Computer, MESCOE/ Savitribai Phule Pune University, India)

Abstract: *In today's modern world surveillance system has become essential for our daily life, due to ever-increasing crime rates security authorities rely on this surveillance system for catching the suspects. Trailing of suspects can be done through a surveillance system but it is a very long and tiring process as each camera produces several hours of feed which is to be processed by the operator. To ease the effort of searching through hours of camera feed a robust method is been proposed to track a person in a video feed and then connecting his tracks further in the network using a trajectory prediction algorithm which saves time and speeds up the process. It aids in dipping the time invested for searching an entity in a lengthy video stream and finding its track through multiple camera feeds.*

Key Word*: Surveillance System, Tracking, Trajectory, Lstm.*

Date of Submission: 12-05-2020 Date of Acceptance: 24-05-2020

I. Introduction

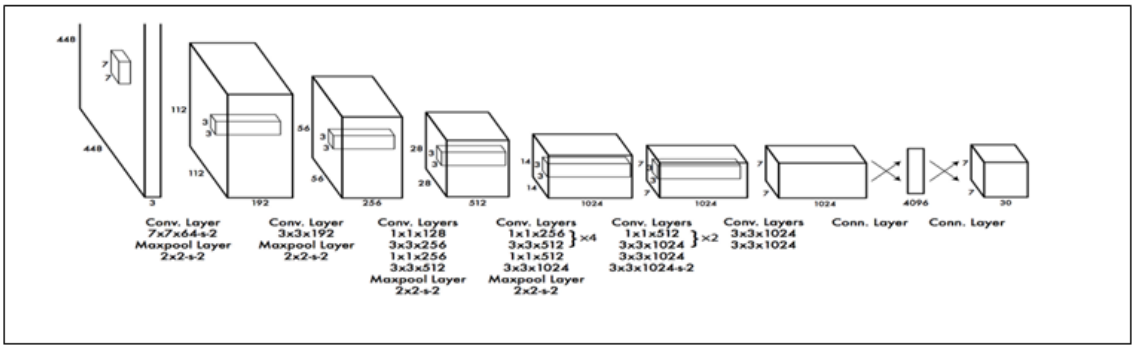
Now a days, surveillance systems are part of our everyday life environments. Consequently, these camera networks represent a massive source of information for monitoring human activities and to propose new services to the users. So, an automated model to support people investigating long videos other than a human inspection is a necessity. Entity tracking and detection are part of the significant problems in computer vision tasks. In recent years, most attention is concentrated on visual tracking, and several works from different perspectives in this area have been reviewed. In this context, to build the best video surveillance system the basic requirement is developing a sustainable and robust algorithm, with the dictum of obtaining a result on the basis of fast, consistent and strong mobile person detection and tracking system. The key suggestions of recent tracking systems are about the object detection method in image processing that detects objects frame by frame. Binary classification is the basis of this type of object tracking algorithm. Deterministic regions are chosen from multiple regions generated by the model. Further, the object and the surrounding background are determined by a trained classifier. Many models have been presented for object detection and tracking. Although, substantial results have been achieved by those algorithms, tracking an object through video sequences is still a challenging job. According to the object detection method, numerous movements of the object is tracked from each frame along with its position information to predict several object trajectories. The reason is that this method neglects the continuousness of the tracked object in the visual tracking process. The vigilance system proposed in this paper divided into two phases namely; object detection & tracking phase and person re-identification phase.

Object detection and tracking are divided into 2 parts: -
1)Object Recognition and Detection: - Object recognition and detection progressed with a basic algorithm, Viola-Jones algorithm that uses Haar-like features to detect face, nose, and eyes that is further optimized using Adaboost. Various algorithms such as CNN, R-CNN, Faster-RCNN, YOLO, etc. have been developed since then.

YOLO (You Only Look Once): YOLO is a state-of-the-art model for object detection and recognition. Before YOLO object detection used repurpose classifiers or localizers for detection. The model was applied on multiple images and varying scale, each region was scored and the highest scoring regions in the image were considered as detections.

In YOLO single CNN predicts multiple bounding boxes for an image or frame. YOLO divides the image into an m * m grid-like structure and every cell in a grid detects a single object, for every cell multiple bounding box are drawn on to the image.

Each bounding box has a fixed size and has 5 parameters height, width, x-coordinate, y-coordinate, box confidence score. x and y act as offsets for the bounding boxes, height and width are used to normalize the bounding box and the box confidence score shows the likeliness of the box containing an object in it.



Figure(I):*You Only Look Once (YOLOv3) architecture*

Kalman filtering: Kalman filtering is used for object tracking in real-time. It uses data regarding the state of an object, inaccuracies, and noise then predicts the new state of the object. There are multiple variations based on the assumption of the parameters of the filter namely constant velocity model where the velocity is considered as constant during sampling.

2)Object Re-identification: Traditionally, person re-identification was implemented in two phases,
1) feature extraction
2) distance metric learning.

CNN combined these phases into a single phase that created an end to end solution for person re-identification. A large family of person re-id study focuses on metric learning loss. Some approaches associate identification loss with verification loss, others imply triplet loss with hard sample mining. Numerous current works employ pedestrian attributes to apply more supervisions and perform multi-task learning. Deep learning models are developed for person re-identification based on evolutionary models. The re-id embeddings are improved by using the generated data which expands the search space and makes the system faster and more efficient.

II. Existing System

Cameras are used by security officials and police departments in form of surveillance system to be of vary of illegal activities and cameras can also be used by individuals for purpose of security.

The current surveillance systems use facial recognition, object detection and many other algorithms to detect and track criminals but all this work is done with supervision of an operator this increases the discrepancy in output also introduces human error. All the work is done by a human operator who manually searches through CCTV footage for the suspect.

The data gets recorded and then stored in the main database, after collecting the data analysis is done on it to segregate the areas into zones with highest probability of crime. Security officials are using algorithms to analyses crime data in order to predict where the probability of an offence taking place is high.

After an incident if the suspect is not identified by the system the person is tracked through various footages by the operator for purpose of identification which is a long process and have many drawbacks such as human error, low resolution feeds, blind spots, etc. Also, tracking a person without interference of a human being is not possible yet.

By developing a well-structured system that can effectively trace a person's track through an interconnected network of cameras we overcame these obstacles.

III. Literature Survey

Deep attention network for person re-identification withmulti-loss: Identification and verification task are combined. Two images are provided to the deep attention network, features extraction for pedestrians is

performed which enhances the generalization performance. Then ID's of two images are predicted and the verification phase determines the similarity using a cross-entropy function [1]

Unsupervised detection and tracking of moving objects forvideo surveillance Applications: - Only one filter is created as the number of objects are initially unknown. When an object enters the view, its location is detected by the initialized filter. A new filter is created to detect an object every time it enters a frame. In the

case where an object stops moving orleaves the scene, the filter responsible for tracking this object is removed [2]

A novel model based on deep learning for Pedestrian Detection and Trajectory prediction: -Trajectory prediction is a difficult task as multiple factors affect the way a person crosses an obstacle based on various factors namely path taken, objects in the path, surrounding objects adouble-layered lstm model is used with YOLOv3 for object detection and Kalman filtering for object re-identification.[3]

Object detection by Spatio-temporal analysis and trackingof the detected objects in a video with variable background - Kumar S. Ray, Soma Chakraborty proposes an approach for detecting and tracking objects in videos captured by moving cameras without any additional sensor. In the workpresented, moving objects are detected as clusters of Spatio-temporal blobs generated by Spatio-temporal analysis of the image sequence using a three-dimensional Gabor filter and merged using Minimum Spanning Tree. Theproblem of data association during tracking is solved by Linear Assignment Problem and occlusion are handled by the application of the Kalman filter.[4]

T. Mahalingam, M. Subramoniam contributed a method that is divided into three phases namely detection, tracking and evaluation phase.The detection phase comprises of foreground segmentation and noise reduction. Foreground segmentation is achieved by using a Mixture of Adaptive Gaussian (MoAG) model. Noise removal in the foreground segmented frames is implementedby the fuzzy morphological filter model. Under the tracking phase, a moving object is tracked by blob detection. The final evaluation phase contains feature extraction and classification.[5]

Venkata Prasad.V, Chandra Sekhar Rayi, Cheggoju Naveen, Vishal R. Satpute provided a technique for detecting an object's action by using the GMM algorithm. Primarily, the GMM algorithm is applied and subsequently by using the foreground image from GMM, the object's action is detected i.e. whether the object is static or moving. [6]

Minxian Li, Fumin Shen, Jingya Wang, Chao Guan, Jinhui Tang has proposed a re-id method based on the spatial-temporal features of a person and then uses a hierarchical spatial-temporal model to reduce the gallery size for person re-id in large scale surveillance system. Visual cue and spatial-temporal cue are used for person re-identification, suitable assumptions on the behavior of objects in the blind gaps are made [7]

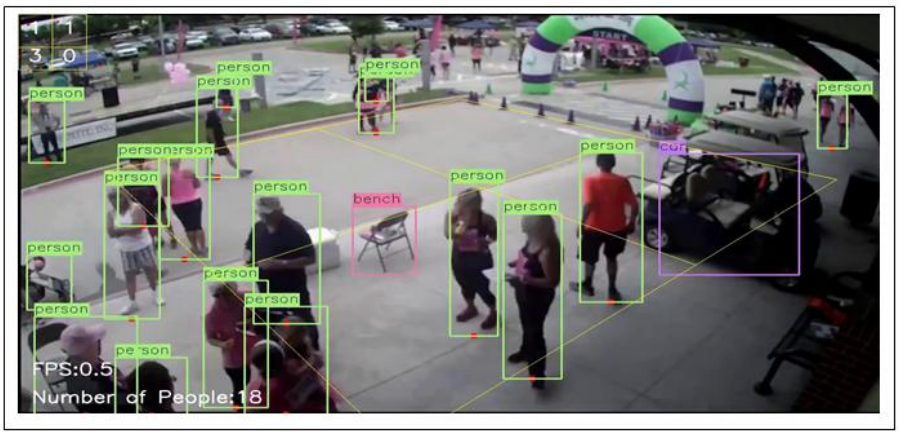
Based on ResNET50 a hybrid task CNN is built, HTCNN is divided into two parts first the ranking task which uses weighted triplet loss to learn global features and second uses cross-entropy to learn local features. [8]

Multiple methods such as mentioned above are developed for object tracking and detection in real-time. We are trying to make a robust the system that will combine object detection, tracking, re-identification and the prediction which will help in tracking an object's trajectory throughout the entire grid of interconnected surveillance cameras. The idea behind the proposed system is built upon previously proposed models for different tasks namely detection, re-identification, trajectory prediction using YOLOv3, Kalman filters, Lstm respectively.

IV. Methodology

The system is divided into three phases: -

*1)**Person Recognition and detection***: - Numerous objects exist in the feed collected from the surveillance system. Recognition of person of interest (POI) is necessary after that, all other recognized objects are discarded from the system as they are not required. For the purpose of object recognition and detection, we are using YOLO v3 as it can detect the objects in real-time with great accuracy.



Figure(III) *Real-Time Human Identification with kalam filters*

YOLO v3 (You Only Look Once)*:* - YOLO works as a real-time entity detection algorithm. YOLO has its own architecture which is based on convolutional neural networks. YOLO v3 divides the frame in a 13 x 13 grid. In YOLO v3 detection is done by applying a 1 x 1 kernels on feature maps of three different sizes at three different places in the network. YOLO v3 makes prediction on basis of 3 different scales, which are precisely given by down sampling the dimensions of the input image by 32, 16 and 8 respectively. YOLO v3 uses 3 anchors per scale i.e. total 9 anchors for an image. Bounding boxes are generated around the objects detected by the algorithm the basic YOLO v3 is trained on COCO dataset for multiple object detection it can be trained on other datasets as well according to requirements

In our model we are solely detecting pedestrians in a video feed for that we have retrained the yolo model and restricted the number of classes to 1 during the detection.

Figure(II) Pedestrian detection through surveillance system using YOLO v3 on ETH Dataset

*2)****Person Re-identification***: - We are using Kalman filtering for feature extraction of a person and detecting it in the consecutive frames as a single entity as YOLO can detect multiple objects at a time but cannot keep track of a single person at a time. Instead of extracting features of a person we are trying to build a series of co-ordinates that will determine the position of the POI (person of interest) in each frame after some time. This helps in keeping track of that person in every frame before his/her appearance

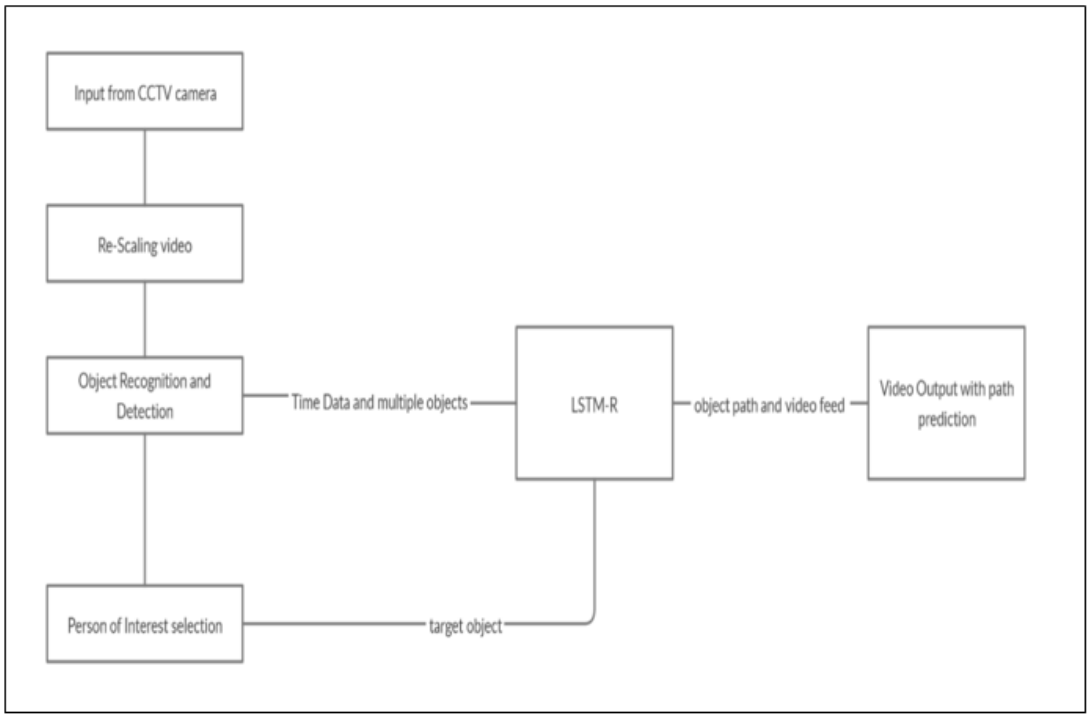
For each object detected there are 3 types of flags assigned to it:

A. *Tentative*: - Object is detected but it is not sure about it
B. Confirmed: - Object is detected and assigned an id
C. *Deleted*: - Object no longer exists.

Two bounding boxes are drawn surrounding the POI, one for object detected and second for the predicted position of the object. The centroid of the object is calculated and it is fed to the trajectory prediction algorithm which predicts the further motion of the object inside the frame if an object exits the frame the trajectory object is predicted which helps in the activation of the cameras in the network in the predicted direction

Trajectory Prediction:Long Short-Term Memory (LSTM) for path prediction is used. Lstm takes the input of the previous states and predicts the next few states based on the input given to it. The basic concept behind Lstm is based on recurrent neural networks where a previous state is remembered by the network but long-term dependencies are the major problem faced by RNN this problem is solved by Lstm as it keeps track of previous states for a longer period.

In our model, Co-ordinates of the object are feed to the Lstm algorithm which predicts the co-ordinates for the next 5-10 frames in a video stream. This is helpful when the objects exit the frame and we want to track the person in real-time, so the cameras in the predicted direction are activated instead of searching for the person in each direction. This reduces the latency in the original system where human operators perform all these actions manually.



Figure(IV) *System Architecture*

1) Input: **-** Input is taken in the form of the video stream from surveillance cameras present in the premises, the feed is present in different formats, dimensions and aspect ratio depending upon the camera used. Input can be taken from traffic signals, college premises, parking lots, streets covered by CCTV cameras

2)Re-Scaling: **-** Like the video, the stream is taken from various types of cameras their resolution differs from each other. Therefore, rescaling of the video frames is necessary, each and every frame is re-scaled depending on the original resolution of the video stream.

3)Object Detection and Recognition: **-** After the video is normalized or rescaled it is fed as input to the yolov3 model for the purpose of object recognition and detection. Model traverse's each frame to find a person of interest in it, this process takes place in a batch where each batch contains multiple frames. Each entity is characterized with an exclusive code after the existence of that entity is assured by the Kalman filters and flagged as confirmed. Two bounding boxes are drawn surrounding the object one by the YOLOv3 model and second by the Kalman filter model. The centroid of each entity surrounded is calculated and the time of detection is recorded and then both the entities are stored in a CSV file.

4)Person of Interest (POI): **-** A solitary being that is recognized by the user and who is to be tacked is termed as a person of interest (POI). It can be anyone a normal pedestrian, a criminal or even an official. It is passed as a parameter to the LSTM-R model for its prediction.

LSTM-R (Long Short-Term Memory - Responder): - A LSTM based model with the extra feature of responding to previous data and providing an output in the form of the predicted path that was used by the entity and triggering the cameras in that course for tracking an individual in that surrounding.

5)Output:-A series of frames containing the bounding boxes, predicted path and direction of the path taken by the POI in the form of video

V. Conclusion

In this paper, a well-structured system for detecting and tracking a person using interconnected networks of cameras is presented. Using computer vision techniques to detect the person also tracking that person and predicting the path of that person helps in following the tracks of an individual which indeed helps in tracking a suspicious person in a crowded area through several cameras present in the surveillance system. It aids to track any doubtful person trying to damage the society or has already committed a crime where cameras are present as eyewitnesses.

References

[1]. Rui Li a ,Baopeng Zhang a , Dong-Joong Kang b , Zhu Teng a,"Deep attention network for person re-identification with multi-loss", Computer and Electrical Engineering, vol 79, 2019, https://doi.org/10.1016/j.compeleceng.2019.106455

[2]. IssamElaÞ , Mohamed Jedra , Noureddine Zahid," Unsupervised detection and tracking of moving objects for video surveillance applications", Pattern Recognition Letters, vol. 84, pp. 70-77, 2016, https://doi.org/10.1016/j.patrec.2016.08.008

[3]. K. Shi, Y. Zhu and H. Pan, "A novel model based on deep learning for Pedestrian detection and Trajectory prediction," 2019 IEEE 8th Joint International Information Technology and Artificial Intelligence Conference (ITAIC), Chongqing, China, 2019, pp. 592-598.

[4]. Kumar S. Ray, Soma Chakraborty, "Object detection by spatio-temporal analysis and tracking of the detected objects in a video with variable background", Journal of Visual Communication and Image Representation, vol. 58, pp. 662-674, 2018, https://doi.org/10.1016/j.jvcir.2018.12.002

[5]. T. Mahalingam, M. Subramoniam, "A robust single and multiple moving object detection, tracking and classification", Applied Computing and Informatics,2018, https://doi.org/10.1016/j.aci.2018.01.001

[6]. Venkata Prasad.Va, Chandra Sekhar Rayia, CheggojuNaveena, Vishal R. satpute, "Object's Action Detection using GMM Algorithm for Smart Visual Surveillance System", International Conference on Robotics and Smart Manufacturing, 2018, vol. 183, pp. 276-283, https://doi.org/10.1016/j.procs.2018.07.034

[7]. Minxian Li, Fumin Shen, Jingya Wang, Chao Guan, Jinhui Tang," Person Re-Identification with Activity Prediction based on Hierarchical Spatial-Temporal Model", Neurocomputing, vol 275, pp. 1200-1207, 2018, https://doi.org/10.1016/j.neucom.2017.09.064

[8]. Shuang Liu, WenminHuang , Zhong Zhang, "Person re-identification using Hybrid Task Convolutional Neural Network in camera sensor networks", Ad Hoc Networks, vol 97, 2020, https://doi.org/10.1016/j.adhoc.2019.102018

Mrs. Jayshree Pansare, et. al. "Vigilance system to track a person in real-time in a closed network of interconnected cameras." *IOSR Journal of Computer Engineering (IOSR-JCE)*, 22(3), 2020, pp. 20-25.