Multiple Person Detection and Tracking using Convolutional Neural Network

Pranob K Charles, SK Jasmine Sultana, P Hemalatha, E Keerti
Department of Electronics and Communication Engineering, Andhra Loyola Institute of Engineering and Technology Vijayawada,-520008, Andhra Pradesh, India.

Article Info
Article history:
Received 17 January 2020
Received in revised form
20 May 2020
Accepted 28 May 2020
Available online 15 June 2020

Keywords: Surveillance Video, Multiple person detection, Convolutional Neural Network, detection accuracy, robust detection, occlusion.

Abstract: Tracking multiple persons is a challenging task when persons move in groups and occlude each other. Existing group based methods have extensively investigated how to make group division more accurate in a tracking-by-detection framework. However, few of them quantify the group dynamics or consider the group in a dynamic view. Inspired by the sociological properties of pedestrians, this work proposes a network that tracks the moving persons.This work uses CNN algorithm by extracting the ROI based HOG features to track more accurately without any interference and to obtain a robust free result. It can be done by considering the thresholding values of the person and based on the thresholding bboxes are assigned to the persons to keep the track s of the persons being detected. Implementation of algorithm, creation of user Interface, leads to observe the performance criteria of the persons being tracked which will gives us accurate and robust free results and widely used in video surveillance and generates direct response. As in the case of any accidental decisions taken by a manual observation can be replaced by using this kind of network .
CNN based tracking can overcome the problem of manual observation very accurately based on region of Interest.

1. Introduction

The Human detection and tracking is one of the important tasks in Computer Vision. As Safety is one among basic human need that need to be fulfilled and as in Computer Vision it mainly includes methods for acquiring, processing and understanding of digital images. Public safety is one among major task problem faced within the world. As rate rising, needs of safety on public place is additionally becoming an enormous demand. Commonest used solution for this problem is surveillance video. Surveillance video allows us to record images or videos on certain location. With this application of technology, feel watched then give us sense of security.

As treat tracking as a learning problem of estimating the situation and therefore the scale of an object given its previous location, scale also as current and former image frames. Given a group of examples, train convolutional neural networks (CNNs) to perform the above estimation task. Different from other learning methods, the CNNs learn both spatial and temporal features jointly from image pairs of two adjacent frames. Introducing multiple path ways in CNN to raised fuse local and global information.

As surveillance videos that are widely used today only ready to capture image or record video, there's no additional information except that pixel combination provided by surveillance video device. Surveillance video device only send images or video to watch in security room. This condition led to wish of human resource to watch the image or video footage recorded by surveillance video device. While the device is recording non-stop it also means surveillance video operator must watch the monitor continuously. By watching the monitor continuously, the operator can suffer fatigue which will reduce effectiveness of surveillance video. Therefore, there's a high demand to automatically process footage from surveillance video device and extract additional information which will be useful for security officer.

In this study to unravel the problem on person detection and tracking, using AI supported Convolutional Neural Networks (CNN) to detect and track the human position inside surveillance video footage. This framework has been trained from arranged dataset contains thousands of images to extend performance on detecting human in various condition and it is trained using Convolutional neural network to detect human and added with regional proportion layer to localize the human condition. After to acquire the human location inside the footage then using tracking algorithm to trace the human and record its movement. The experimental results of this method shows excellent results both on detecting and tracking by considering features of a person which includes HOG and ROI based HOG and on comparative illustration acquired the desired results.

2. Related Work

As there are many detection and tracking results exists they are foreground detection method, the feature detection method and many classifier algorithms are used to track the persons. Research on computer vision especially in surveillance video is growing fast in recent years. Human detection, act recognition, motion tracking, scene modeling, and behavior understanding are growing popularity in computer vision and machine learning researcher and communities. This led to discussion about the way to maximize performance on advanced surveillance video.

Deep learning method [1] are successfully improved various visual detection and recognition tasks. Example of this application used for image classification [2], [3], image segmentation [4], and object detection [5], [6]. Deeper network features a main advantage of the power to find out effective feature representation automatically, which make appealing for practitioners. All the network parameters are solely learned from the training data.

As for surveillance video, many researchers were trying to extract information from video footage. For surveillance video topics, its objective is to detect, recognize, or learn interesting events. This results in research on action recognition, suspicious event detection [13], irregular behaviour [14], unusual activity [15], anomaly [16], and abnormal behavior[17]. To detect and locate human in our video footage. During this work, by utilizing deep technique on human detection combined with tracking algorithm to show it's possible to detect an track human movement from video footage.

3. Existing system

As the existing method uses multiple vehicle tracking by using two classifier technique. The proposed method is explained by the key components of person detection, prediction of persons in future frames, track let associations, and managing the life span of identities for tracked objects.

In order to overcome the matter that occlusion and interfaces causes wrong multiple vehicle tracking, the two classifier method is supposed. As this method enables the tactic to possess both time efficiency advantage of the first classifier SVM and high accuracy advantage of the second classifier CNN. Firstly, the improved ViBe is used to extract the connected areas to detect moving vehicles, and then the SVM classifier along with the combined LBP features is used to scan on the image. If the

***Corresponding Author,**
E-mail address: shaikjasu2511@gmail.com

threshold level may be a smaller amount than the low threshold Tmin, it is determined not to be the thing vehicle, and it is discarded directly. If the threshold value is greater than Tmin and fewer than the high threshold Tmax, then the connected region could even be the candidate vehicle tracking frame. The CNN classifier with CNN features [17] is used because the second classification to remove the interference region. If the threshold value is greater than Tmax, then it can be judged to be the right tracking area .It is added into connected region table directly. After the foreground detection and classification of auto tracking rectangle, the connected region matching algorithm is used to undertake the correlation analysis for the motion of vehicles front and rear frame to understand stable multiple vehicle tracking.

Therefore, the proposed technique to firstly as we are interested in tracking of humans and to do the overall tracking and detection by using only a single classifier technique, which will reduce the complexity of implementation and by using a single classifier technique can be able to effectively track multiple persons. The process involved in it is given in figure1.

4.Proposed System

Block Diagram of the Proposed System has been shown in the figure1 as:

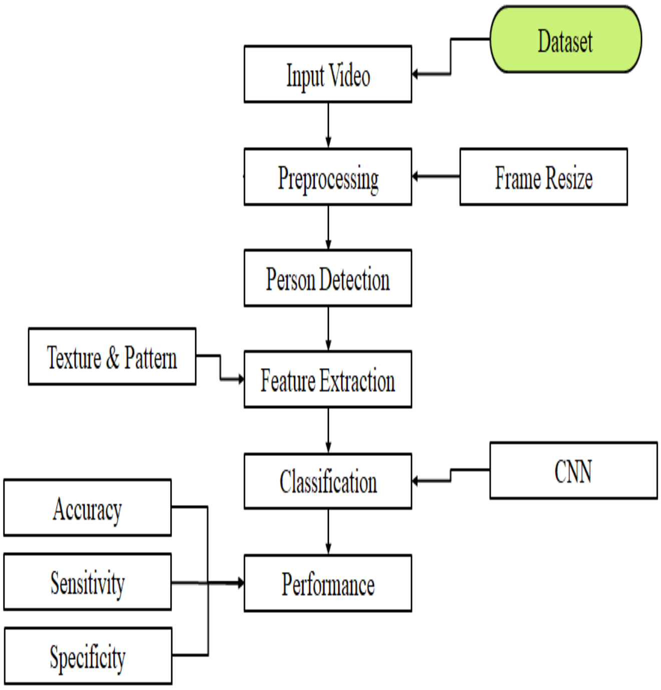


Fig.1: Block Diagram

4.1.Video Footage

Footage utilized in this paper consists of a dataset,which contains people walking inside it. The video footage size can be of any size based on the application, so as to track and improve theaccuracy of detection.

4.2. Pre- Processing

In this stage the video is divided into multiple frames say 1000 frames and can be resized by using frame resize parameter ratio.

4.3. Person Detection by extracting features:

Reading the frame is done in this stage and whether the person is there not is identified by using this network. Features can be extracted by using ROI which helps us to assign the boundary box around the person based on the thresholding value of the person and which contains information indicating human position. Main core of the person detection framework is Convolutional Neural Networks (CNN). CNN as a deep learning method has shown significantly great performance for detection and classification. Therefore, CNN as base of our neural networks structure.

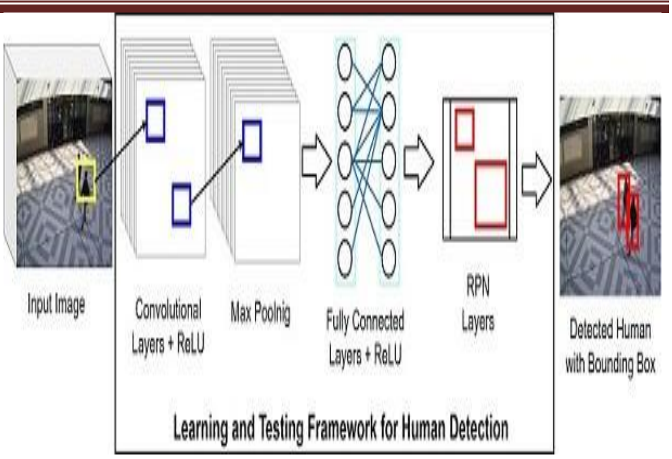

Input Image
Convolutional Layers + ReLU
Max Poolnig
Fully Connected Layers + ReLU
RPN Layers
Detected Human with Bounding Box
Learning and Testing Framework for Human Detection

Fig.2: Person detection framework

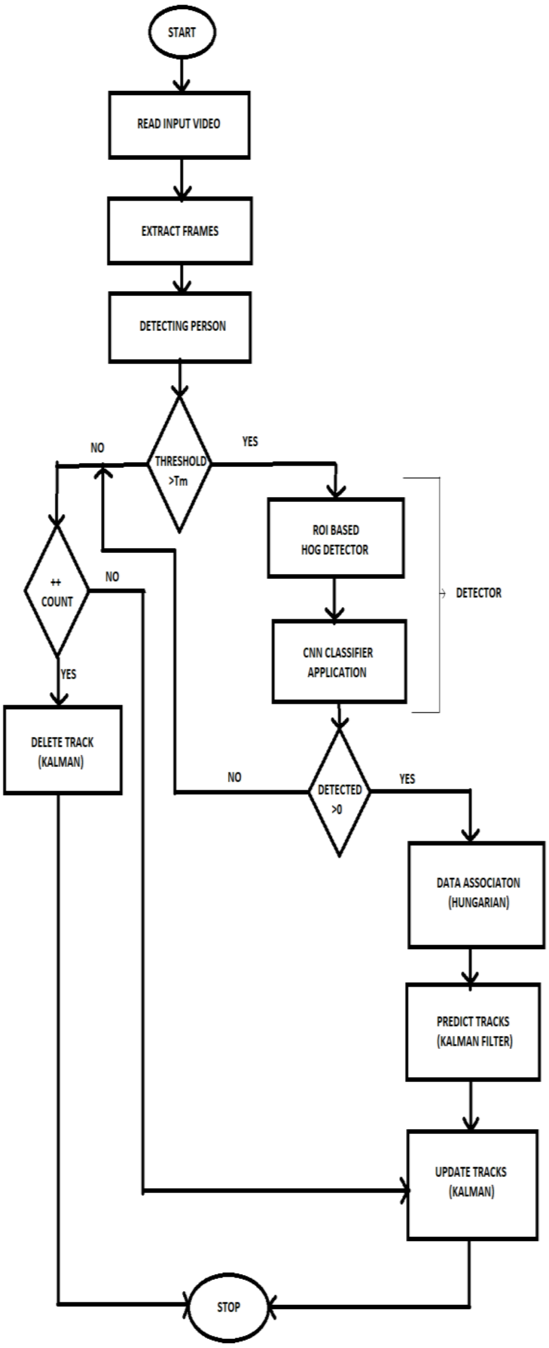
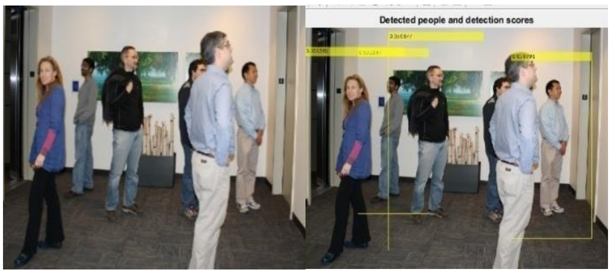


Fig.3: Flow Chart of CNN Classifier with ROI based HOG detection Results

4.4. CNNclassifier

CNN is essentially several layer staged together a bit like another neural network structure. A layer in CNN commonly consists of convolutional, max pooling, and fully connected layers which have different roles for every layer. A convolutional layer contains linear filter which is followed by a non-linear activation function. This work used an activation layer like the Rectified linear measure (ReLU).During this convolutional layer, a CNN utilizes kalman filter to convolve the entire image also because the intermediate feature maps, generating various feature maps. The feature map contains A width, B height, C channels to point size of feature map. To scale back dimensions of feature maps used for pooling layer then followed by convolutional layer. Overview of person detection framework used in this research. The framework consists of convolutional layer, ReLU, max pooling, and fully connected layer. This framework given output of human location with bounding boxes.

Pooling layers are invariant to translation, it takes the neighboring pixels of feature maps. Max pooling is just taking the utmost value from predetermined window. A fully- connected layer performs similar as feed forward neural network. It converts previous multi-dimensional feature maps into a pre-defined length. These layers acts as classification layer and will be used as feature vector for next processing. The flow chart consists of the steps of operation involved in the process of the framework, which is shown in the figure 3. Input and the detected people with the score has been shown in the below frames 1 and 2 as follows:

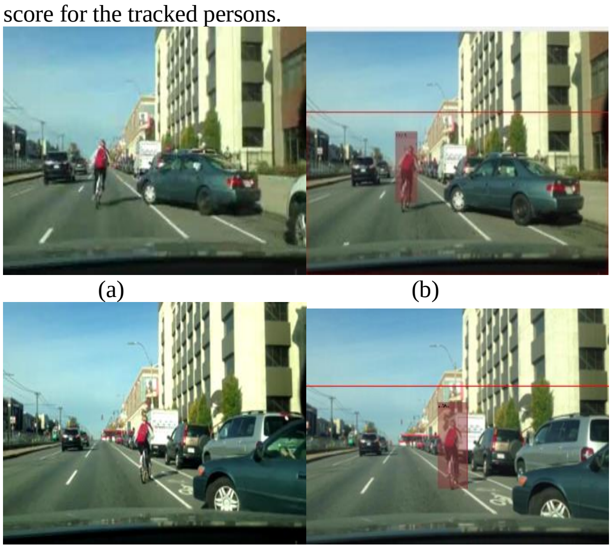


Frame1 Frame2

Fig.4: Frame with the detected people result

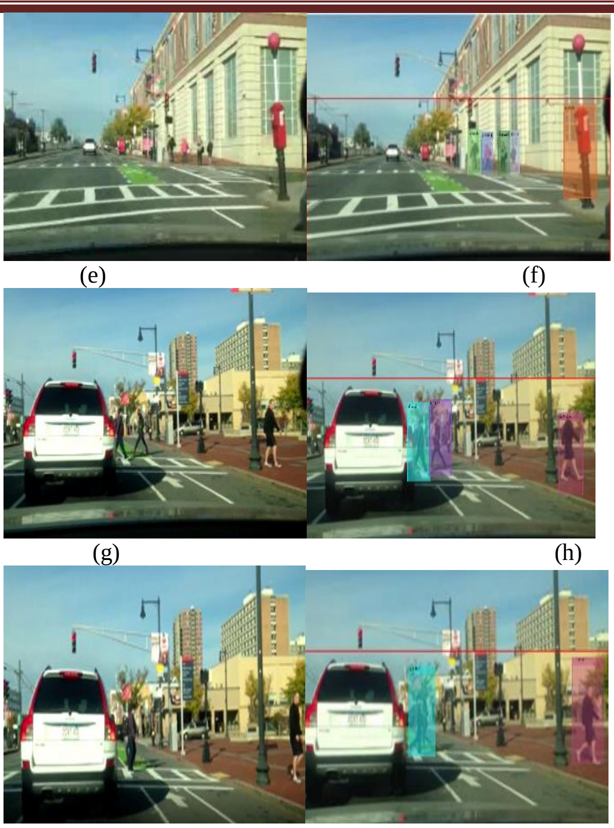
As shown in the figure 4 the persons having the same pixel values are considered as a single person. So as by considering the region of interest of a person the drawback can be overcome.

Video footage is of 780X1280p size and the frames are extracted from the video footage and the processing is done by using the convolutional neural network withROI based HOG features of the persons, and the results obtained as shown in the fig5. For the frame (a) the corresponding result is the frame (b) with boundary and the score for the tracked persons.



(a) (b)

(c) (d)



(e) (f)

(g) (h)

(i) (j)

Fig. 5: The results for the read images from the footage and thecorresponding detected results with scores has been shown.

Figure 5, representing the existing CNN and the results that are obtained by using CNN with ROI. The graph depicts the rates of detection of the existing method and the proposed method that uses CNN with ROI, and x axis gives the frames and accordance to that how the scores i.e., the detection rates are occurred is shown. As the CNN based tracking does not track in the frames from 50 to 65, while CNN with ROI will track it. Hence improving the accuracy of detection and tracking.

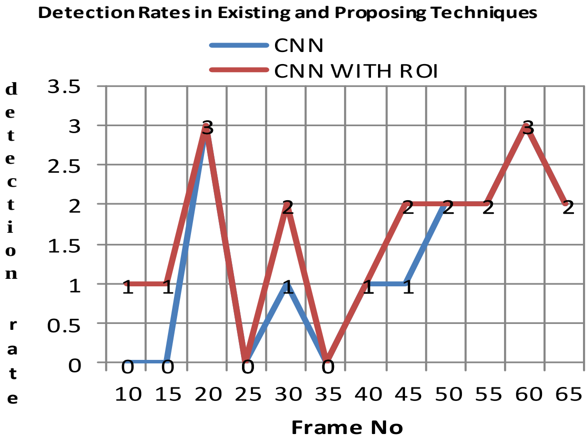


Fig.6: Detection rates and the corresponding frames

5. Conclusion

This study have provided research work on multiple person detection and tracking for surveillance video footage, using Convolution Neural Network for person detection and tracking algorithm provided by ROI based HOG features to track the detected person .As in the previous detections only by using CNN detection takes place even though it has high accuracy it has Computational complexity of CNN is higher and it is time consuming. Therefore, by considering its ROI, can improve the accuracy of the system and the multiple persons are effectively tracked. As it has the disadvantage of Tracking the objects whose height and width look like that of a person it would misinterpret and consider it as a person and track it .So it need to be overcome in the future to avoid error in tracking of a person.

References

[1] Y LeCun, Y Bengio,G Hinton. Deep learning, Nature, 521(7553), 2015, 436–444.

[2] A Krizhevsky, I Sutskever, GE Hinton. Imagenet classification with deep convolutional neural networks, Advances in neural information processing systems, 2012, 1097–1105

[3] K Simonyan, A Zisserman. Very deep convolutional networks for large-scale image recognition, arXiv preprint arXiv, 2014, 1409-1556.

[4] C Farabet, C Couprie, L Najman, Y LeCun. Learning hierarchical features for scene labeling, IEEE transactions on pattern analysis and machine intelligence, 35(8), 2013, 1915–1929.

[5] R Girshick. Fast r-cnn, Proceedings of the IEEE International Conference on Computer Vision, 2015, 1440–1448.

[6] R Girshick, J Donahue, T Darrell, J Malik. Rich feature hierarchies for accurate object detection and semantic segmentation, Proceedings of the IEEE conference on computer vision and pattern recognition, 2014, 580–587.

[7] Ferryman, James, A Shahrokni. Pets2009: Dataset and challenge. IEEE International Workshop on Performance Evaluation of Tracking and Surveillance. 2009.

[8] He, Kaiming. Deep residual learning for image recognition. Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.

[9] RK Satzoda, MM Trivedi. Multipart vehicle detection using symmetry-derived analysis and active learning, IEEE Trans. Intell. Transp. Syst., 17(4), 2016, 926–937.

[10] O Barnich, M Van Droogenbroeck. ViBe: A universal back ground subtraction algorithm for video sequences, IEEE Trans Image Process., 20(6), 2011, 1709–1724.

[11] Everingham, Markl. The pascal visual object classes (voc) chalenge. International journal of computer vision 88(2), 2010, 303-338.

[12] M Abadi. Tensor Flow: Large-scale machine learning on hetero- geneous systems, 2015. Software available from tensorflow.org

[13] GLayee, L Khan, Bthuraisingham. A framework for a video analysis tool for suspicious event detection”,Multimedia Tools Appl, 352007, 109-123.

[14] Yzhang, Z Liu. Irregular behaviour recognition based on treading track, in Proc,Int Conf.Wavelet Anal Pattern Recog 2007 .

[15] Kristan, Matej. The visual object tracking vot2017 challenge results. Proceedings of the IEEE International Conference on Computer Vision. 2017.

[16] HZhong, JShi, Mvisontai. Detecting unususal activity in video, Proc.2004.

[17] YBenezeth, P Jodoin. Abnormal events detection based on soatio-temporal c0-occurences, Proc IEEE Conf 2009.

[18] A Lukezic, T Voj’ir, LC Zajc, J Matas, Matej Kristan. Discriminative correlation filter tracker with channel and spatial reliability. International Journal of Computer Vision.

[19] Anton Milan, Laura Leal-Taixe, Joint Tracking and Segmentation of Multiple Targets. IEEE 2015.

[20] MD Breitenstein, F Reichlin. Robust Tracking-by-Detection using a Detector Confidence Particle Filter. International Conference on Computer (ICCV).

[21] S Yu, Y Wu, L Wei, Z Song, W Zeng. A model for fine-grained vehicle classification based on deeplearning, Neurocomputing, 257(27), 2017, 97–103.

[22] DT Lin, YH Chang. Occlusion Handling forPedestrain Tracking Using Partial Object Template-based Component Particle Filter. In IADIS 8(2).

[23] J Berclaz, F Fleuret. Multiple Object Tracking Using K-Shortest Paths Optimization. IEEE on pattern Analysis and Machine Intelligence 33(9), 2011.

[24] X Shi, H Ling. Multi-Target Tracking by Rank-1 Tensor Approximation. IEEE Conference On Computer Vision and Pattern Recognition 2013.

[25] W Hu, W Li. Single and multiple object Tracking using a Multi-Feature joint Sparse Representation. IEEE 0162-8828, 2013.

[26] A Milan, K Schindler. Detection- and Trajectory- Level Exclusion in Multiple Object Tracking. IEEE 2013.

[27] L Zhang, Y Li. Global Data Association for Mult-Object Tracking Using Network flows.

[28] Y Xiang, A Alahi. Learning to Track: online Multi-Object Tracking by Decision Making, IEEE Conference 2015.

[29] W Choi. Near-Online Multi-Target Tracking with Aggregate Local Flow Descriptor. IEEE 2015.

[30] J Xing, Haizhou. Multi-Object Tracking through Occlusion by Local Tracklets Filtering and Global Tracklets Association with Detection Responses. IEEE 2009.