Moving object detection and tracking Using Convolutional Neural Networks

Shraddha Mane
Department of Electronics and Telecommunication
MKSSS's Cummins College of Engineering for Women,
Pune, India
Shraddha.sanjay.mane@cumminscollege.in

Prof.Supriya Mangale
Department of Electronics and Telecommunication
MKSSS's Cummins College of Engineering for Women,
Pune, India
Supriya.mangale@cumminscollege.in

***Abstract*— The object detection and tracking is the important steps of computer vision algorithm. The robust object detection is the challenge due to variations in the scenes. Another biggest challenge is to track the object in the occlusion conditions. Hence in this approach, the moving objects detection using TensorFlow object detection API. Further the location of the detected object is pass to the object tracking algorithm. A novel CNN based object tracking algorithm is used for robust object detection. The proposed approach is able to detect the object in different illumination and occlusion. The proposed approach achieved the accuracy of 90.88% on self generated image sequences.**

Keywords—CNN, Object detection, TensorFlow, Tracking,

I. Introduction

In last decade, number of approaches were proposed and demonstrated by different researchers for foreground detection and tracking [1]. However these approaches were failed to resolve the problems like radical changes and target drift during tracking. The main challenge for moving object detection and tracking is to estimate the object position more accurately. Moving object detection is important step of the computer vision algorithm. It is used in different applications like video analysis, medical imaging and military application.

Usually, frame contains background and foregrounds information [2]. In this foreground object is represented by features points in the ROI and remaining features are consider as background. In general, surveillance system consists of two major steps such as moving object detection and motion estimation. The first step is the object detection and it is influenced by the background pixels information's.

Video is irrelevant and redundant to the space and time hence the video data need to be compressed. Compression can be done with the spatial and temporal information minimization in the video.

A number of researchers have a lot of methodologies that focus on detecting objects from a video sequence. Many of them use multiple techniques and there are combinations and intersections between different methodologies. Background subtraction is the method which extracts the interested moving object from the video frames [3]. The background subtraction is affected by mostly non-stationary background and illumination changes. In practice, this drawback can be removing by the optical flow algorithm but it is produces false alarm for tracking algorithms under cluttered conditions. In most of the cases of background subtraction, the object trackers are influenced by background information but it lead to the misclassification. Further the selection of robust classifier being the challenge to increase the accuracy of the algorithm.

To overcome this limitation, in this approach a novel and generalized Tensor flow based object detection and CNN based object tracking algorithm has been presented. These approaches are robustly detected and track the object in complex scenes and complicated background conditions. CNN is based learner because it is demonstrated to extract the local visual features and they are used in the recognition algorithms. CNNs require the extraction of local characteristics by limiting the receptive fields of the hidden units as local, based on the fact that the images have strong local two-dimensional structures. The convolutional neural network combines three architectural ideas to guarantee a certain degree of invariance of change and distortion: local receptive fields, shared weights (or pending replication) and sometimes, spatial and temporal sub sampling.

Tracking objects is a fundamental problem in computer vision. Traditional methods based on feature, such as those based on color [4] or blobs movement [5]-[7], follow-up maintaining a simple model of the objective and adapting this model over time. However, real situations in practice pose enormous challenges to these techniques because: 1) over time, the model of the object can deviate from its original, and 2) do not have a discriminating model that distinguishes the category of interest from the others

II. Literature Survey

There are different approaches had been presented by different researchers starting from background subtraction to CNN. Some of the human tracking methods have been presented in this section.

Human tracking consist of three basic steps for pedestrian tracking: Human detection from sequence of frame, tracking and analysis of the tracking for particular purpose.

There are three fundamental aspects of pedestrian tracking that are analogous to object tracking: 1) Detection of the pedestrian in the video frame, 2) Tracking of the detection, and 3) Analysis of the tracks for the specified purpose [8].

978-1-5386-2842-3/18/$31.00 ©2018 IEEE

In this literature survey, object feature point detection, background subtraction, segmentation and classification algorithms of previous research have been discussed. For tracking to be perfect, features which described the object is most important, hence the object detection is plays vital role. This can be achieved by using deterministic or probilistic motion models and appearance based model. To achieve the better accuracy adaptations of the model have been presented over time. The feature points were trained and update in the process of tracking. Only problem to track the object is that it requires large number of features which cannot be always be possible [9].

Recently, the CNN is used to image classification and recognition to improve the significant performance. CNN is trained with millions of images of different classes [10]. CNN are the learning method which exploits the spatial information of an image and learn the complex features automatically. CNN is intrusive to the variation of an input [11].

Fan et al. [12] presented the CNN based object tacking algorithm with shift variant architecture. In this algorithm, the features were learned during online process. The spatial and temporal features are considered using pair of images instead of single image.

Hong et al. [13] presented the approach where the output of the last layer of the pre-trained CNN module is cascade with the on-line SVM to learn discriminative appearance models. The tracking is performed using Bayesian network with target specify saliency map.

Wang et al. [14] used pre-trained CNN model for online tracking. The CNN is used after parameter tuning to adjust the appearance of the object in the scene and probability map is created to instead of creating labels.

Wang and Yeung [15] create a stack denoising auto encoder offline model which learned from the offline training. This model transfers the knowledge from offline model to track the online object.

III. Methodology

The proposed CNN based moving object detection algorithm consists of two phase: Object detection and tracking. The generalized block diagram of the proposed system is shown in Fig. 1.

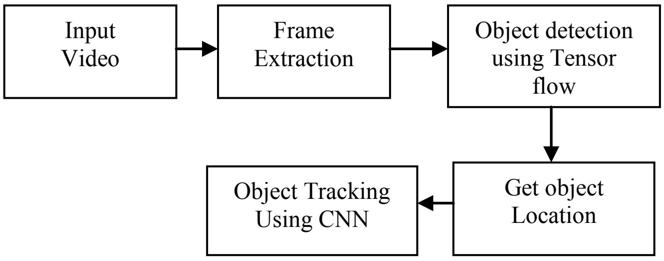

Fig. 1. Block Diagramof proposed system

In this system, the video is feed to the system as an input. Frames are extracted for further processing. The two main algorithms object detection and object tracking is process through deep learning methods. The object detection is explained in detail in below flow.

The object detection using computer vision algorithm is affected by different aspects like light variation, illumination, occlusion and system has difficulty to detect the multiple objects. Hence in this paper, Tensor flow based object detection algorithm has been used.

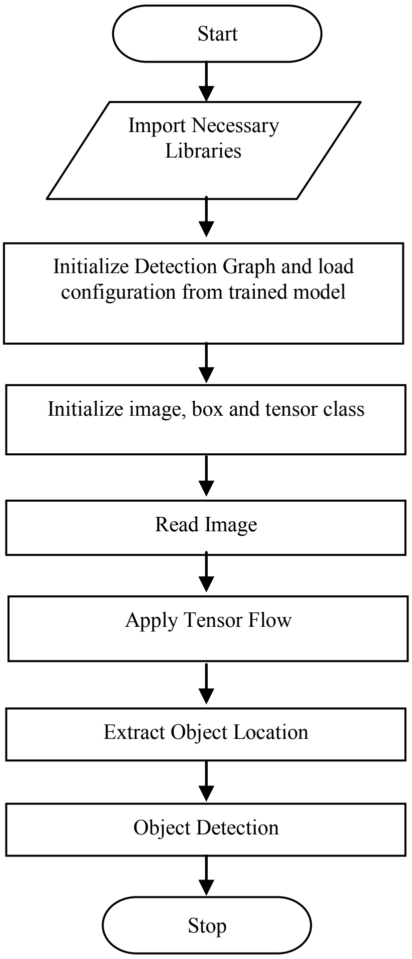
Start
Import Necessary Libraries
Initialize Detection Graph and load configuration from trained model
Initialize image, box and tensor class
Read Image
Apply Tensor Flow
Extract Object Location
Object Detection
Stop

Fig. 2. TensorFlow Based Object detection flowchart

TensorFlow based object detection API is an open source platform. It is built on the top of TensorFlow which make simple to construct, train and detection models. The process of tensor flow based object detection is presented in Fig. . In this approach, firstly the necessary libraries are imported. Then import the pre-trained object detection model. The weights are initializing along with box and tensor class. After initialization of all the parameters of the tensor flow model, the image in which object to be detected is read. Apply the loaded tensor flow model on the image, the TensorFlow based model test the image and return the location (x, y, w, h) of the object in the image. This is the process of object detection of TensorFlow object detection algorithm. The success rate of this approach is better and it is applicable to RGB images.

After detecting the object, their locations are important to start the tracking process. Instead of using conventional computer vision based algorithm, in this approach Convolutional Neural Network (CNN) based tracking algorithm is used. The Flow of CNN based object tracking algorithm is as shown in Fig.3.

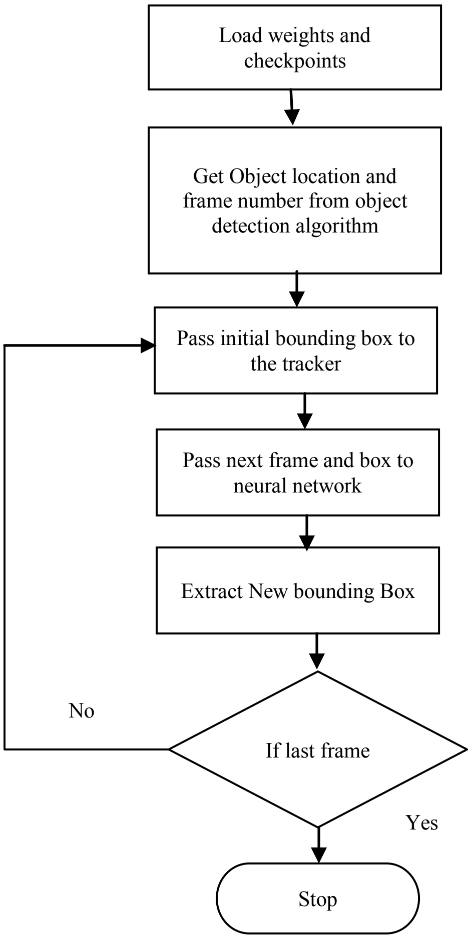
Load weights and checkpoints
Get Object location and frame number from object detection algorithm
Pass initial bounding box to the tracker
Pass next frame and box to neural network
Extract New bounding Box
No
If last frame
Yes
Stop

Fig. 3. Flowchart for object detection

The object tracking is the important step in computer vision algorithms. For tracking to be robust, requires object knowledge and understanding like motion and its variation over time. Tracker must be able to its model and adopted for new observations.

In this approach first load the weights of the pre-trained model. The model is capable of incorporating the temporal information. Rather than focusing on the objects in the testing time, the pre-trained model which is trained on large variety of objects in real time. This lightweight model has ability to track the object at the speed of 150 frames per second. Also it is able to remove the remove the barrier of occlusion.

In this approach, the object locations obtained from the TensorFlow based object detection algorithms are passed to the CNN based object tracking algorithm. The initial positions are learned by the model and the same points are search in the net frames by testing process of CNN model.

IV. Results

The proposed algorithm is tested on variety of video sequences. The experimentation is divided into two parts, the object detection and tracking. The algorithm is implemented in python and tested on nine different video sequences, 3.6GHz Laptop. Without optimization the algorithm runs with good FPS.

The results of the proposed algorithm are presented in qualitative and quantitative manner. The qualitative analysis is as shown in Fig. 4 for different sequences.

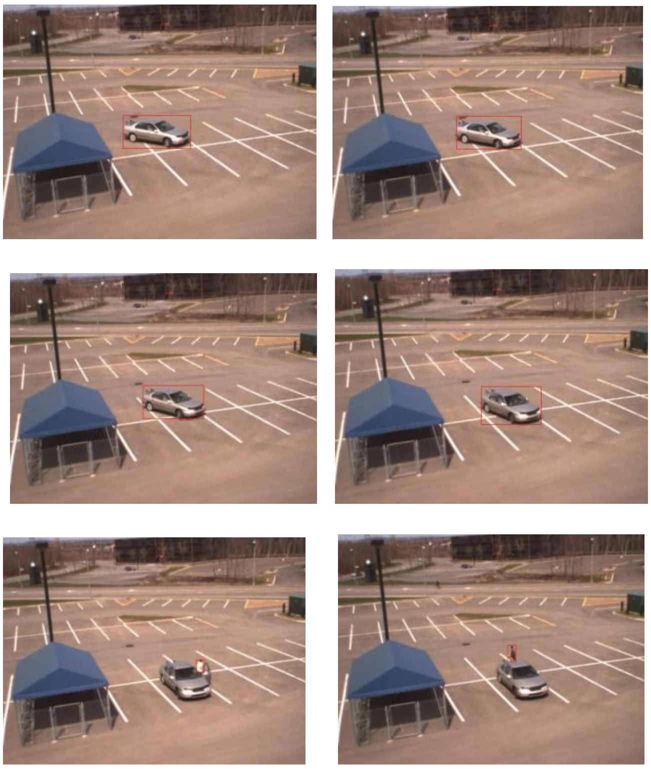

Fig. 4. Qualitative analysis of proposed system of the cdv sequence.

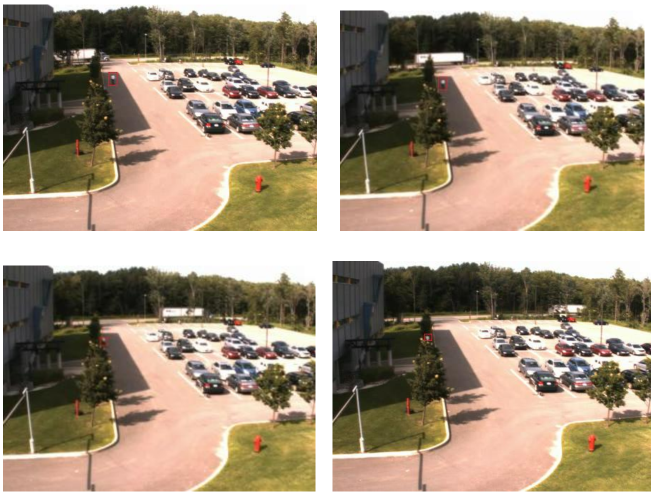

Fig. 5. Qualitative analysis of proposed system of the mdv sequence.

The quantitative analysis is performed using sensitivity, specificity and accuracy parameter. These parameters are calculated using True Positive (TP), True Negative (TN), False Positive (FP) and False Negative (FN).

- TP: moving object correctly identified moving object
- FP: Stationary object incorrectly identified as moving object
- TN: Stationary object correctly identified as Stationary object
- FN: moving object incorrectly identified as Stationary object

The mathematical representation of the quality metrics is given as:

A. Sensitivity

It is the ratio of truly object present in the scene who are correctly identify as an object. This term present the number of positive samples correctly identified. Higher the true positive element higher is the sensitivity.

$$Sensitivity = \frac{TP}{TP+FN} \tag{1}$$

B. Specificity

It is the ratio of truly stationary object present in the scene that are correctly identify as a stationary object. This term present the number of negative samples correctly identified. Higher the true positive element higher is the Specificity.

$$Sensitivity = \frac{TN}{TN+FN} \tag{2}$$

C. Accuracy

Accuracy is the overall performance of the system including sensitivity and specificity.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{3}$$

TABLE I. QUALITATIVE ANALYSIS OF THE PROPOSED SYSTEM

Database video sequences	No. of frames	Sensitivity	Specificity	Accuracy
Brv	1201	0.9885387	0.9837067	0.9775
Cdv	2030	0.9597742	0.9018933	0.928079
Cpv	239	0.8874172	0.797619	0.841004
Gfv	1482	0.9730539	0.9766537	0.946694
Mdv	2400	0.9160305	0.8788789	0.915833
Mev	551	0.8856089	0.8931034	0.905626
Pev	820	0.9467593	0.9331395	0.88916
Psv	2938	0.8630665	0.9254984	0.905037
T	386	0.8732394	0.9217391	0.870466
Average		0.9214987	0.9124702	0.908822

V. CONCLUSION

In this paper, novel approach for object detection and tracking has been presented using convolutional neural network. The moving object detection is performed using TensorFlow object detection API. The object detection module robustly detects the object. The detected object is tracked using CNN algorithm. Considering human tracking as a special case of detection of objects, spatial and temporal classes the facilities were learned during offline training. The shift variant architecture has extended the use of conventional CNNs and combined the global features and local characteristics in a natural way. The proposed approach achieves the sensitivity of 92.14%, specificity of 91.24% and accuracy of 90.88%.

References

[1] A. Elgammal, R. Duraiswami, D. Harwood, and L. S. Davis, “Background and foreground modeling using nonparametric kernel density estimation for visual surveillance,” Proc. IEEE, vol. 90, no. 7, pp. 1151– 1163, 2002.

[2] Enrique J. Fernandez-Sanchez *, Javier Diaz and Eduardo Ros, “Background Subtraction Based on Color and Depth Using Active Sensors”, Sensors 2013, 13, 8895-8915; doi:10.3390/s130708895.

[3] Prajakta A Patil, Prachi A Deshpande, “Moving Object Extraction Based on Background Reconstruction”, International Journal of Innovative Research in Computer and Communication Engineering, Vol. 3, Issue 4, April 2015

[4] Jialue Fan, Wei Xu, Ying Wu and Yihong Gong,” Human tracking using Convolutional Neural Network,” in Proc. IEEE transactions on neural networks, vol.21, NO.10, Oct 2010

[5] Junda Zhu, Yuanwei Lao, and Yuan F. Zheng, “Object tracking in structured environment for video surveillance applications”, IEEE transactions on circuits and systems for video technology, vol.20, February 2010.

[6] D. Koller, J. Weber and J. Malik, “Robust multiple car tracking with occlusion reasoning,” Proc. Third European Conference on Computer Vision, 1994, pp. 189-196, May 2-6 1994

[7] Logesh Vasu and Damon M. Chandler, “Vehicle tracking using a Human-Vision-based Model of Visual Similarity” IEEE 2010.

[8] Chen, Y, X. Yang, B. Zhong, S. Pan, D. Chen, and H. Zhang, “Cnn tracker: Online discriminative object tracking via deep convolutional neural network”. Applied Soft Computing, 2016.

[9] Feris, R., A. Datta, S. Pankanti, and M. T. Sun, “Boosting object detection performance in crowded surveillance videos”. In IEEE Workshop on Applications of Computer Vision, pp. 427-432, 2013.

[10] Krizhevsky, A., I. Sutskever, and G. E. Hinton, “Imagenet classi_cation with deep convolutional neural networks”, 2012

[11] Ji, S., W. Xu, M. Yang, and K. Yu, “convolutional neural networks for human action recognition”, IEEE Transactions on Pattern Analysis and Machine Intelligence 35(1), 221-231, 2013.

[12] Fan, J., W. Xu, Y. Wu, and Y. Gong, “Human tracking using convolutional neural networks”. IEEE Transactions on Neural Networks 21(10), 1610-1623, 2010.

[13] Hong, S., T. You, S. Kwak, and B. Han, "Online tracking by learning discriminative saliency map with convolutional neural network", arXiv preprint arXiv:1502.06796 .

[14] Wang, N., S. Li, A. Gupta, and D. Yeung, "Transferring rich feature hierarchies for robust visual tracking", Computing Research Repository abs/1501.04587, 2015.

[15] Wang N., S. Li, A. Gupta, and D. Yeung, "Transferring rich feature hierarchies for robust visual tracking". Computing Research Repository 2015, abs/1501.04587.