



DETECTION-LEARNING-TRACKING OF MULTIPLE OBJECTS IN VIGILANCE VIDEO

¹Mr. Ankit Pandey, ²Mr. Parmatma Pandey, ³Prof. Geetha Narayanan

Department of Electronics and Telecommunication Engineering Vidyalankar Institute of Technology, Mumbai, India^{1,2}, Department of Electronics and Communication Engineering J.J.T.U. , Jhunjhunu-Rajasthan, India³

ABSTRACT

Tracking, Learning & Detection of a real time objects in video stream plays a very crucial role in case of surveillance video. An object is defined by its position & extent in next frame. Efficiently designed object tracking systems can provide better efficiency but sometimes they fail due to loss of information caused by complex shapes, rapid motion and illumination changes. There are different types of algorithms that can be used like DNN, MOSSE, BOOSTING, MIL, MEDIANFLOW, TLD, KCF, CSRT. The proposed system uses Deep Neural Network (DNN) in order to detect an object. using Open CV trackers. This paper describes a method for tracking multiple objects and moving in real time of tracking in video stream. Initial results shows that the different object tracking methods have different properties with respect to number & trajectories of the objects

Keywords - Multiple object tracking, TLD , KCF, Median Flow, CSRT.

INTRODUCTION

In the field of computer vision, object detection, tracking & recognition have become challenging & important tasks for a long time. Consider a video stream taken by a hand-held camera depicting various objects moving inside and outside the camera's view. Bounding box defines object of interest in various frame & the aim is to determine that the object is not visible in any of the frame that follows. Video stream is to be processed at frame rate & this process should run continuously for a long time. We refer to this task as long-term tracking. The frame work of Tracking learning & detection was initiated by kalal . This technique is very effective during the situation where tracking happens for restricted scenarios, such as pedestrians, by using back ground subtraction. Where subtraction is applicable on N-type objects to provide false alarm when target is constantly under motion and can create illusion or when out of vision. Cam shift framework provides good tracking accuracies when object is under slow or constant motion, but at the time when object is in rapid motion tracking system get confused due to background illumination and distraction. This framework can provide better performance by using object classification method using object location.

This improved framework can judge whether targeted object has lost. This methodology works on five different parameters to recognize targeted objects like, 2-D central location, width, height & orientation of selected objects. A technique called template matching is successful one. This technique is independent from shape and size of target. This technique provides template for each available video frame and extract matching elements out of it. In case of targeted object if matching elements are not found in frame, then system can easily claims, object is not available in available video frame. Background subtraction is another technique which is use for

motion segmentation. It detects moving regions by subtracting current video frame from previous video frame pixel by pixel. These pixels are used to enhance detection region of targeted object and to avoid background illusion. Every time the background gets updated which new frame over the time .

To enable long-term tracking, there are a number of problems which need to be addressed. The main difficulty is, detection of object during the time of reappearance in camera's view. This problem is aggravated by fact that the object may change its presence, thus making the appearance from the initial frame irrelevant. Next, a successful long-term tracker should handle scale and illumination changes, back-ground clutter, and partial occlusions and operate in real time .

The availability of high quality & inexpensive video cameras along with interesting need for automated video analysis has generated a great deal of interest in the field of object tracking.

The tracking can be defined as a process of following an object through successive image frames to determine its relative movement with respect to other objects.

Tracking can be simplified by imposing pressure on the motion or appearance of objects. One can further constrain the motion of object to be of constant velocity or acceleration based on prior information. Prior knowledge regarding the number & the size of objects, or the object appearance & shape can also be used to simplify the problem.

Different approaches for object tracking have been proposed. These primarily differ from each other based on the way they approach the following question: which object representation suitable for tracking? Which image features should be used? How should be the appearance & the shape of object? The answers for these questions depend on context/environment in which tracking is performed. There are huge number of tracking methods that have been proposed which attempt to answer these questions for variety of scenarios.

E-ISSN NO:2349-0721

LITERATURE SURVEY

Xiao Liu, Dacheng Tao, Mingli Song, Luming Zhang, Jiajun Bu, and Chun Chen "Learning to Track Multiple Targets" IEEE Transaction on Neural Networks And Learning System. This paper have proposed a framework for detection and tracking of multiple objects.

- In this authors are trying to implement algorithm for detecting, learning and tracking of multiple object. Each of this algorithm will work simultaneously for the object.
- This paper deals with the problem of automatically tracking a variable number of targets by using high definition camera to capture video and processes using video processing techniques in congested scenes camera.
- This method decreases the post processing mistake risk and improves the performance in tracking. It is efficient in many cases like vehicle tracking in congested traffic.

Bin Sun, Zhi Liu, Yulin Sun, Fangqi Su, Lijun Cao and Haixia Zhang "Multiple Objects Tracking and Identification Based on Sparse Representation in Surveillance Video" IEEE International Conference on Multimedia Big Data.

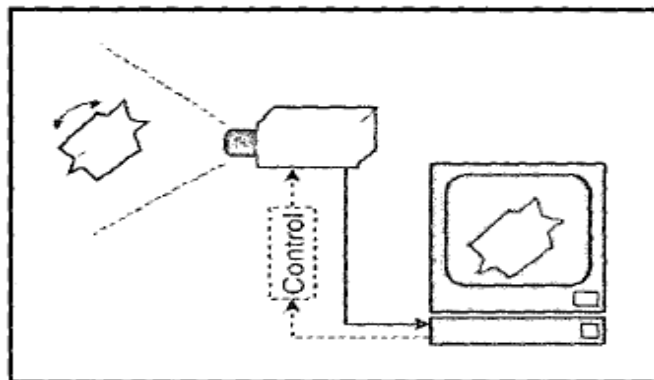
- In this paper, authors have proposed an efficient multiple objects tracking and identification method based on sparse representation in surveillance video.
- Color histogram and Hash code are used as the feature for detection, and discriminative sparse similarity map method is used to guarantee the accuracy of tracking result.

Shen-Fu Hsiao, Guan-Fu Yeh, and Je-Chi Chen“Design and Implementation of Multiple-Vehicle Detection and Tracking Systems with Machine Learning”17th Euromicro Conference on Digital System Design. In this paper, authors will present a new vehicle detection system based on support vector machine (SVM), a machine learning technique.

- In this paper authors have also discussed two stages for vehicle detection system :
 - a. Hypothesis generation (HG)
 - b. Hypothesis verification (HV).
- HG uses frame division and shadow detection to find possible candidates of vehicles within a suitable region of the image frame. HV is used to eliminate unreasonable hypotheses.
- In this paper, a driver-assistance system is developed that can detect the existence of multiple vehicles. Here HG finds vehicle candidates using shadow information based on cumulative density function (CDF).
- HV removed non-vehicle hypotheses using ratio-constraint and a machine learning technique called (SVM).The system is implemented by both software and hardware running on a FPGA development board.

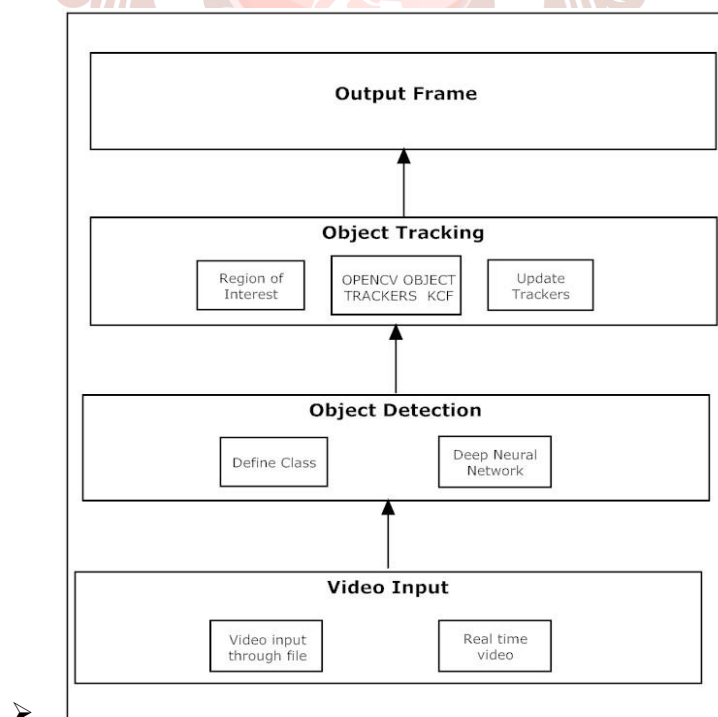
PROPOSED SYSTEM

- Figure 1 shows the schematic of a generic object tracking system. As can be seen, visual input is usually achieved through digitized images obtained from a camera connected to a digital computer. This camera can be either stationary or moving depending on the application. Beyond image acquisition, the computer performs the necessary tracking and any higher-level task using the tracking result.



- Fig 1:Schematic of a generic object tracking system.

- Here the Tracking, Learning & Detection of any real time objects can be done. Video streams are processed at frame rate and this process will run indefinitely long. Here for object detection in the video stream we use deep learning and caffe model with Open CV Object Trackers. The proposed method consists of three part which are tracking , learning & detection. System is tested using open CV Library. First the camera is initialized on system for capturing the video or select a saved video file and select one of the Open CV trackers for tracking the object in the given video stream. After resizing and pre-processing the frame is converted to the rgb colour channel ordering. Now object detector which is trained to detect presence of an object is used for detection followed by a filter to reject weak detections by checking the threshold level. If detection is above threshold then it is placed in a rectangular box and the box is passed to the tracker .This box is needed to create a Dlib. For tracking object of our choice we have to select Region Of Interest (ROI)that is the object that has to be tracked.The figure below shows the architecture of the proposed system which is consist of different parts like video input ,objects detection, objects tracking and final output in the output frame. Here the input is
- provided either from the stored file in the system or from the real time video , and this input is passed to object detection, where the passed input is detected by using deep neural network (DNN) and then this detected output is given for tracking where the region of interest (ROI) is created for each object. Consider sum threshold value in order to find the brightest spot and its intensity. If the object is not found completely in match region then we will consider it as new object and store in the data base for tracking the same object in future. If the object is matched then we update the tracker along with ROI Region and the process repeats. Thus this matched object is displayed at the output frame.



➤ Figure 2 .Architecture of the Proposed System

➤ Algorithm Used In Moving Object Detection

➤ **1.Detection Using Deep Learning:**

- Object detection in a video surveillance applications is the major task these days. Object detection technique is used to determine the required objects in video sequences & to cluster pixels of these objects. This detection of an object in video sequence plays a major role in several applications specifically as video surveillance applications .
- In a video stream Object detection can be done in different ways like pre-processing, segmentation, foreground and background extraction, feature extraction.
- The field of artificial intelligence is essential when machines can do tasks that typically require human intelligence. It comes under the layer of machine learning, where machines can acquire skills and learn from past experience without any involvement of human. Deep learning comes under machine learning where artificial neural networks, algorithms inspired by the human brain, learn from large amounts of data. The concept of deep learning is based on humans' experiences; the deep learning algorithm would perform a task continuously so that it can improve the outcome. Neural networks have various (deep) layers that enable learning .

➤ **2 Caffe Model:**

- Caffe is a framework for Deep Learning it was used for the implementation & to access the following things in the object detection system .
- **Expression:** Models & optimizations are defined as plaintext schemas in case of caffe model unlike others which use codes for this purpose.
- **Openness:** Common code, reference models, and reproducibility are the basic requirements of scientific and applied progress.
- **Speed:** For research and industry alike speed is crucial for state-of-the-art models and massive data .
- **Modularity:** Flexibility and extension is majorly required for the new tasks and different settings.

IMPLEMENTATION OF TLD

- Step 1: A camera is fitted over a laptop .
- Step 2: Moving objects video is taken into the system .
- Step 3: Objects are selected by the user.
- Step 4: For learning image array is generated.
- Step 5: The Cropped image is stored at array index 0.
- Step 6: Crop the ROI. ROI can be extended upto 20 pixels & if object is not obtained in the current ROI.
- Step 7: Generate next Frame from video stream.
- Step 8: Apply template matching in ROI at index 0 in order to obtain the Object location.

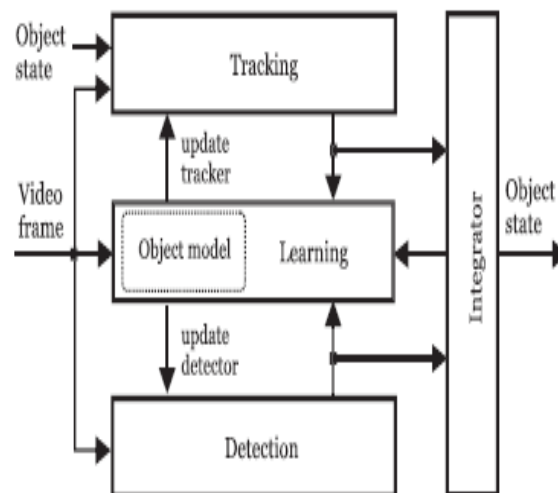


Fig-3: Detailed block diagram of the TLD framework.

Step 9: In case match is found then check the percentage of matching. If matching percentage is less than stored percentage, new object in the learned array at the index 0 and shift other images in the array.

Step 10: If match is not found then Identification of detector errors and learning from it can be done from by P-N learning. P-N learning find out the errors with the help of experts. P-experts finds the missed detection whereas N-experts finds the alarms.

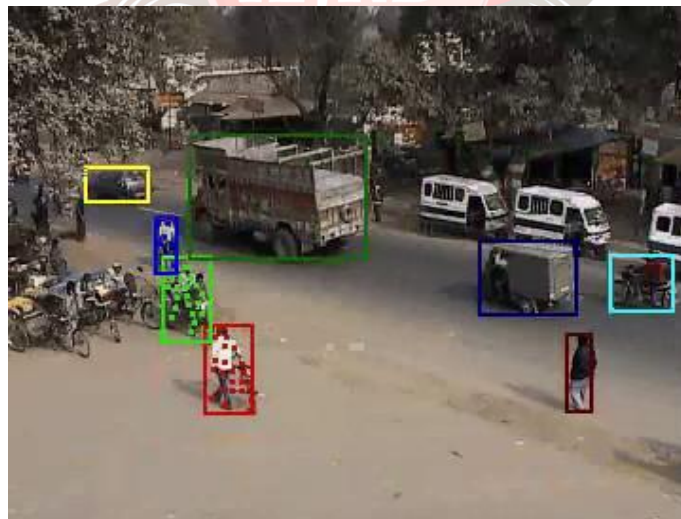


Figure 4: Multiple object tracking

APPLICATIONS

1. Multiple tracking can be used in detecting number of cars, cycle etc. in a particular time period.
2. Video indexing: automatic annotation and retrieval of in multimedia databases.
3. Traffic monitoring: real-time gathering of traffic statistics to direct traffic flow

4. Vehicle navigation: video-based path planning and obstacle avoidance capabilities.

CONCLUSION

In this project we have given the expressivity of DNNs for object detector. We show that the simple formulation of detection as DNN-base object mask regression can yield strong results when applied using a multi-scale course-to-fine procedure. These results come at some computational cost at training time – one needs to train a network per object type and mask type. As a future work we aim at reducing the cost by using a single network to detect objects of different classes and thus expand to a larger number of classes.

This paper has described methods and problems with respect to the task of Multiple Object Tracking (MOT). Here we have used various types of object tracking algorithms such as :- KCF, CSRT, MIL, Median Flow, TLD. Among this all object tracking algorithms we find that TLD works better than all and has better accuracy.

REFERENCES

- [1] A. Blum and T. Mitchell, “Combining labeled and unlabeled data with co-training,” Conference on Computational Learning Theory, p. 100, 1998.
- [2] B. D. Lucas and T. Kanade, “An iterative image registration technique with an application to stereo vision,” International Joint Conference on Artificial Intelligence, vol. 81, pp. 674–679, 1981.
- [3] Zdenek Kalal, Krystian Mikolajczyk, and Jiri Matas “Tracking-Learning-Detection”, IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, VOL. 6, NO. 1, JANUARY 2010.
- [4] Kumar S. Ray and Soma Chakraborty “An Efficient Approach for Object Detection and Tracking of Objects in a Video with Variable Background” arXiv:1706.02672v1. 11 May 2017.
- [5] M. Sahasri , C. Gireesh “Object Motion Detection and Tracking for Video Surveillance” International Journal of Engineering Trends and Technology (IJETT) – Special Issue – April 2017 .
- [6] Vaishali, Shilpi Singh “Real-Time Object Detection System using Caffe Model” International Research Journal of Engineering and Technology (IRJET) Volume: 06 Issue May 2019.
- [7] L. Wang, W. Hu, and T. Tan, “Recent developments in human motion analysis,” Pattern Recognition, vol. 36, no. 3, pp. 585–601, 2003.
- [8] Peter Janku, Karel Koplík, Tomas Dulík, and Istvan Szabo “Comparison of tracking algorithms implemented in OpenCV,” MATEC Web of Conferences 7604031, 2016