Multiple Object Detection and Tracking

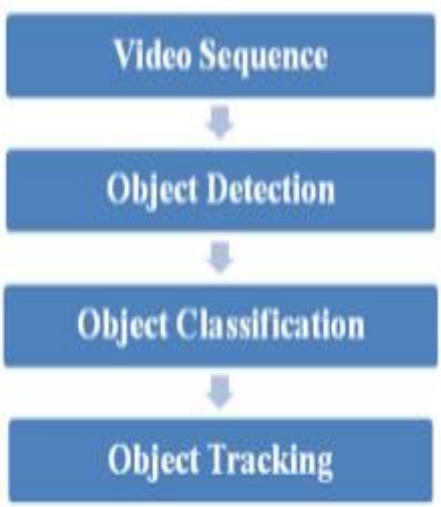
Suresh Kumar[1], Prof. Yatin Kumar Agarwal[2]
[1]M.Tech. Scholar, [2]Professor
[12]*Department of CSE, Greater Noida Institute Of Technology (GNIOT)*
[12]*Greater Noida, Uttar Pradesh, India*

Abstract- Multiple Object Tracking is the process of locating multiple objects over time in a video stream. Object detection and classification are two prior steps before performing tracking over video scene. Object detection is the process of locating an object of interest in a single frame. So, in other words we can say that multiple object tracking is the process of associating detected objects in consecutive video frames. The detected objects may belong to various categories such as vehicles, humans, swaying trees or other moving objects. So, object classification is the process to classify these objects using different approaches. This project deals with the information extraction from an image where the key idea is to detect multiple objects with all the essential elements present in an image and classify the object category and localize the object. The methodology we adopted for this project was a modular approach where we conquered the different aspects by implementing the project in various phases. The knowledge base for the same is from convolution neural networks. We successfully created our object detection model and tested it against various parameters that can affect the performance of the model. The tools we used for this project were python, Jupyter Notebook, Labelmg(for labelling the dataset), OpenCv. Hence, we got to learn to work on an environment (with a computer language) that was not very familiar to us.

Keywords- *Object Detection, Object Tracking, Video Surveillance, Background Modelling, Motion Estimation*

I. INTRODUCTION

Object tracking is a very important aspect of artificial intelligence, digital image processing and computer vision [2] [4]. Object tracking is a core of several computer vision applications such as traffic monitoring, video surveillance, video animation, robot vision and many more [1]. In recent years, a number of single-object tracking algorithms were introduced and implemented robustly, but in case of multiple object tracking scenario is quite different and challenging too because in case of multiple object tracking, targets needs to be matched from frame to frame in video sequence [5]. To deal with this problem, researchers did lot of works to provide better solutions to such applications based on multiple object tracking. The basic flow diagram for multiple object tracking is as following.



This paper is structured in this way: section 1 gives introduction to multiple object tracking. Section 2 provides literature discussion about work related to multiple object tracking. In section 3 we discussed about object detection methods from video. Section 4 briefly describes object tracking methods. At last, section 5 concludes this paper.

II. OBJECTIVE OF THE PROJECT

The main goal of this project is to design, implement and evaluate a model that will detect multiple objects in an image, localize the objects and classify them according to their classes. The aim is to train the model by labelling the dataset and the trained model will be used to predict the output. Our concept includes knowledge about the convolutional neural networks, digital image processing and classification of images. Furthermore, we propose how knowledge of neural networks helps in classifying images and how they are represented in a computer.

III. PROBLEM STATEMENT

To retrieve information from an image using object detection techniques. Our aim is to create an object detection model which will detect multiple objects From an image. The model will detect the object from an image and identify the types of objects Recorded in an image. The aim is to train the model by labeling the dataset and the trained model will be used to predict the output.

IV. RELATED WORK

In literature [5], feature-based algorithm using Kalman filter motion to track multiple objects is proposed by authors. They had used background subtraction method to detect and extract moving object. Algorithm is validated on human as well as vehicle image sequence and also under confusing situation; it achieves efficient tracking of objects. Literature [6] focuses on tracking players in football match. Authors have used modified Hungarian algorithm and Kalman filter. As the result concludes that for tracking multiple objects in football match, the linking process is achieved by modified Hungarian algorithm and motion model building and prediction is achieved by Kalman filter successfully. In literature [7], parallel Kalman filter is used for moving object detection and tracking. The precision and recall value proves that, proposed method is effective for detecting and tracking multiple objects. In literature [8], authors have proposed unique method named as Dual Layer Particle Filtering (DLPF), which simultaneously detect and track multiple target objects. It uses parent-particles (PP) in first layer to detect multiple objects and child-particles (CP) in second layer to track that detected object. In literature [9], the proposed method is based on particle filter, which overcomes some challenges of object tracking. It can track the target object robustly when target object is occluded by the background object or other object. Authors had written that, particle filter has higher flexibility than Kalman filter, also they modifies particle filter to overcome different challenges. In literature [10], proposed method is based on multiple hypothesis tracking. The method is for generic object tracking, which means that there are no priori restrictions in type of objects that can be tracked. However, authors confesses that, proposed method is quite complex too. In literature [11], adaptive template matching algorithm is used for tracking human upper body. The proposed method is fast and robust, as they had added only head edge detection for better tracking. But also they assume that person's upper body and face are visible without any occlusion. In literature [12], authors had written that, in some complex situation like target object have scale changes or similar color with background, traditional meanshift algorithm cannot obtain accurate results. So they suggest new meanshift target tracking algorithm, named as DEPTH & SIFT-Meanshift algorithm. This algorithm is proposed by using a depth camera and SIFT (Scale Invariant Feature Transform) feature metric. The experimental result shows that, proposed method has abilities to overcome described challenges. Literature [13] proposes a robust approach for tracking arbitrary objects. Authors propose a new motion model based on Kernelized Harmonic Means and particle filter, and they introduce their proposed model within a SVM framework.

V. METHODOLOGY AND RESULT ANALYSIS

There are various tracking and image processing algorithms are included in the system. Object detecting is the process of finding instances of real-world objects .The system detects whether object is bicycles, faces, and buildings in images or videos, video tracking is the process of locating a moving object (or multiple objects) over time using a camera. Surveillance means closely and clearly observation of behavior and activities of the objects. Detection of a moving object is necessary for any surveillance system. A static camera can detect and track an object as long as the object is inside the frame of the camera. But as the object goes beyond the boundary of the camera frame, the camera stops tracking it, which is a major limiting factor for the use of a static camera. This limitation can be overcome by using a rotating camera, which will keep continuously rotating and track objects.

5.1 System and Framework

System: Microsoft Windows 10
CPU: Intel Core i5
Programming Language: Python 3.5.2

5.2 The entire methodology is divided into 6 basic parts.

- Setting up the object detection directory and virtual environment
- Gathering and labeling dataset
- Creating a label map
- Training &testing the dataset
- Predicting output of new dataset using trained data

5.2.1 Setting up the object detection directory and virtual environment

First, we import our required packages — as long as OpenCV and NumPy are installed, your interpreter will breeze past these lines.We must parse four command line arguments. Command line arguments are processed at runtime and allow us to change the inputs to our script from the terminal. Our command line arguments include:

- image : The path to the input image. We'll detect objects in this image using YOLO.

- yolo : The base path to the YOLO directory. Our script will then load the required YOLO files in order to perform object detection on the image.

- confidence : Minimum probability to filter weak detections. I've given this a default value of 50% (0.5)

- threshold : This is our non-maxima suppression threshold with a default value of 0.3

5.2.2 Gathering and labeling dataset

In the first phase we have collected the dataset of four items and labelled them. We have classified the four items and using the r-cnn model we have trained the data and tested with the ratio 7:3.After training and testing of the images we have run the python code upon random images and got the output by detecting the object and classifying it with its class name. But, it was less efficient so we developed a new python script with pre-trained data and using a large dataset.

5.2.3 Creating a label map

Using the tool labelmg we labelled the dataset which helps in labelling the dataset and converting the xml file of each image which consists of width, boundaries and many more attributes of an image. The label map tells the trainer what each object is by defining a mapping of class names to class ID numbers. Use a text editor to create a new file and save it as labelmap.pbtxt

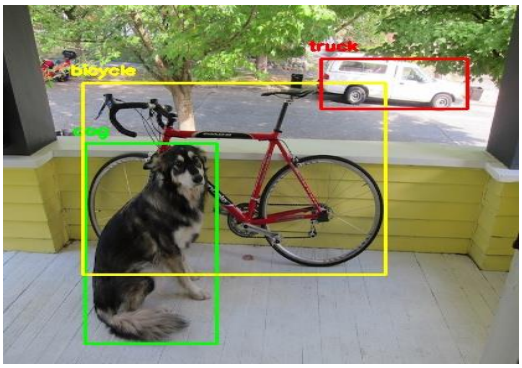
5.2.4 Training and testing the dataset

With the images labeled, it's time to generate the TFRecords that serve as input data to the TensorFlow training model. This tutorial uses the xml_to_csv.py and generate_tfrecord.py. First, the image .xml data will be used to create .csv files containing all the data for the train and test images. The training routine periodically saves checkpoints about every five minutes. You can terminate the training by pressing Ctrl+C while in the command prompt window. Wait until just after a checkpoint has been saved to terminate the training. You can terminate training and start it later, and it will restart from the last saved checkpoint. The checkpoint at the highest number of steps will be used to generate the frozen inference graph.

5.2.5Predicting output

the object goes beyond the boundary of the camera frame, the camera stops tracking it, which is a major limiting factor for the use of a static camera. This limitation can be overcome by using a rotating camera, which will keep continuously rotating and track objects Muti object detection

VI. RESULT



VII. CONCLUSION

After this survey we conclude that lot of work has been done in field of the wireless sensor network and IoT. In order to achieve real-time reactive behaviour we must apply the appropriate image processing and control techniques. This approach used towards increase the object detection with new ideas. We have identified object detection and object tracking application and its implementation using different designs with rotating camera.

VIII. REFERENCE

[1]. M. B. Blaschko and C. H. Lampert. Learning to localize objects with structured output regression. In Computer Vision– ECCV 2008, pages 2–15. Springer, 2008. 4

[2]. L. Bourdev and J. Malik. Poselets: Body part detectors trained using 3d human pose annotations. In International Conference on Computer Vision (ICCV), 2009. 8

[3]. H. Cai, Q. Wu, T. Corradi, and P. Hall. The crossdepiction problem: Computer vision algorithms for recognising objects in artwork and in photographs. arXiv preprint arXiv:1505.00110, 2015. 7

[4]. N. Dalaland B.Triggs. Histogramsoforientedgradientsfor human detection. In Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on, volume 1, pages 886–893. IEEE, 2005. 4, 8

[5]. T. Dean, M. Ruzon, M. Segal, J. Shlens, S. Vijayanarasimhan, J. Yagnik, et al. Fast, accurate detection of 100,000 object classes on a single machine. In Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on, pages 1814–1821. IEEE, 2013. 5

[6]. J. Donahue, Y. Jia, O. Vinyals, J. Hoffman, N. Zhang, E. Tzeng, and T. Darrell. Decaf: A deep convolutional activation feature for generic visual recognition. arXiv preprint arXiv:1310.1531, 2013. 4

[7]. J. Dong, Q. Chen, S. Yan, and A. Yuille. Towards unified object detection and semantic segmentation. In Computer Vision–ECCV 2014, pages 299–314. Springer, 2014. 7

[8]. D.Erhan, C.Szegedy, A.Toshev, and D.Anguelov. Scalable object detection using deep neural networks. In Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on, pages 2155–2162. IEEE, 2014. 5, 6

[9]. M. Everingham, S. M. A. Eslami, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman. The pascal visual objectclasseschallenge: Aretrospective. InternationalJournal of Computer Vision, 111(1):98–136, Jan. 2015. 2

[10]. P.F.Felzenszwalb,R.B.Girshick,D.McAllester,andD.Ramanan. Object detection with discriminatively trained part based models. IEEE Transactions on Pattern Analysis and Machine Intelligence, 32(9):1627–1645, 2010. 1, 4