ISSN 2395-1621

Object Tracking Using Video Processing

Ravish Kumar Raj, Prof. Priti J. Rajput

kumarrajravish@hotmail.com,
preeti.rajput@dyptc.edu.in

Department of Electronics & Telecommunication (Signal Processing),

D Y Patil School of Engineering Academy, Ambi Pune
Savitribai Phule Pune University, Pune, India

ABSTRACT

Object detection is a computer technology related to computer vision and image processing that deals with detecting instances of semantic objects of a certain class (such as humans, buildings, or cars) in digital images and videos. This paper discuss about the method or techniques on how to detect the mango from a mango tree. The techniques using are such as colour processing which are used as primary filtering to eliminate the unrelated colour or object in the image. Besides that, shape detection are been used where it will use the edge detection. This technique will determine the candidates of mango and find the circular pattern with the given radius within an image by collecting the maximum voting. The program should automatically detect the desire object and count the total number of object.

ARTICLE INFO

Article History

Received: 27th July 2020

Received in revised form : 27th July 2020

Accepted: 29th July 2020

Published online : 4th August 2020

Keywords: Image Processing, Deep Learning, Object Detection, Tensorflow

I. INTRODUCTION

In this paper we will have an overview of Single Shot Object detection (SSD) using deep learning for object tracking using image processing, MobileNet and OpenCV. Object detection is hottest topic of the computer vision field. Object detection is breaking into a wide range of industries, with use cases ranging from personal safety to productivity in the workplace. Object detection and recognition is applied in many areas of computer vision, including image retrieval, security, surveillance, automated license plate recognition, optical character recognition, traffic control, medical field, agricultural field and many more.

In Object Detection, we categorize an image and identify where does an object resides in an image. For this, we have to obtain the bounding box i.e (x, y)-coordinates of an image. Frame detection is considered as a regression problem. That makes it easy to understand. Single Shot detection is one of the methods of Object Detection.

By using SSD, we only need to take one single shot to detect multiple objects within the image, while regional proposal network (RPN) based approaches such as R-CNN series that need two shots, one for generating region proposals, one for detecting the object of each proposal. Thus, SSD is much faster compared with two-shot RPN-based approaches.

II. LITERATURE SURVEY

[1] Multiple Object Tracking with Dynamic Fuzzy Cognitive Maps Using Deep Learning
Author Name: - Turan Goktug Altundogan, Mehmet Karakose

Object tracking is the process of matching objects detected on image sequences onto image frames. There are different types of object tracking applications used for different scenarios. For example, if a single object is being traced on an image, this is a single object tracking application. Tracking multiple objects on an image is called multiple object tracking. Fuzzy cognitive maps, on the other hand, form the model of a system by using the features of a system and the relationships between these features. Here, the single object tracking process is a matching problem, so FCM assumes a classifier role. In conventional operations, FCMs use the same weight matrix for all initial concept values. This can reduce the performance of the solution that the FCM produces for the problem it tackles. The FCM structure we use here takes advantage of the dynamic learning of FCM weights with deep learning. The study was tested on different image sequences and the performance of the proposed method were very satisfactory.

[2] Object Tracking Based on Deep CNN Features Together with Color Features and Sparse Representation
Author Name: - Yuchi Liu, Yujuan Qi, Yanjiang Wang

To acquire the better object tracking performance in this paper, the object of interest is modeled by its RGB-color histogram feature together with deep convolutional neural network feature (deep CNN feature) and sparse representation. Because the information of the objects of interest cannot be obtained ahead in practical use, the deep CNN features should be abstracted through the pre-trained VGG network. Then the deep feature and RGB-color histogram are combined to model the object in adaptive mode. And then sparse representation is used to express the model. Above all, an adaptive particle filter algorithm based on deep CNN feature together with RGB-color feature and sparse representation is proposed to track the object of interest. Extensive experiment results demonstrate the effectiveness of our proposed method under the serious object occlusions and appearance changes.

[3] VIRTUALOT - a framework enabling real-time coordinate Transformation & occlusion sensitive tracking using us products, Deep learning object detection & traditional object tracking Techniques.
Author Name: - Bradley J. Koskowich, Maryam Rahnemoonfar, Michael Starek

In this work we explore a combination of methods that allow us to analyze and study hyper-local environmental phenomena. Developing a unique application of monoplotting enables visualization of the results of deep-learning object detection and traditional object tracking processes applied to a perspective view of a parking lot on aerial imagery in realtime. Additionally, we propose a general algorithm to extract some scene understanding by inverting the monoplotting process and applying it to digital elevation models. This allows us to derive estimations of perspective image areas causing object occlusions. Connecting the real world and perspective spaces, we can create a resilient object tracking environment using both coordinate spaces to adapt tracking methods when objects encounter occlusions. We submit that this novel composite of techniques opens avenues for more intelligent, robust object tracking and detailed environment analysis using GIS in complex spatial domains provided video footage and UAS products.

[4] Moving object detection and tracking Using Convolutional Neural Networks
Author Name: - Shraddha Mane, Prof. Supriya Mangale

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.

[5] Objects Talk - Object detection and Pattern Tracking using TensorFlow
Author Name: - Rasika Phadnis, Jaya Mishra, Shruti Bendale

Objects in household that are frequently in use often follow certain patterns with respect to time and geographical movement. Analysing these patterns can help us keep better track of our objects and maximise efficiency by minimizing time wasted in forgetting or searching for them. In our project, we used TensorFlow, a relatively new library from Google, to model our neural network. The TensorFlow Object Detection API is used to detect multiple objects in real-time video streams. We then introduce an algorithm to detect patterns and alert the user if an anomaly is found. We consider the research presented by Laube et al., Finding REMO—detecting relative motion patterns in geospatial lifelines, 201–214.

III.PROPOSED SYSTEM

The main aim of the proposed system is to detect multiple real-time objects from image using deep learning algorithms. Learning. Object detection is the task of detecting instances of objects of a certain class within an image. The state-of-the-art methods can be categorized into two main types: one-stage methods and two stage-methods. One-stage methods prioritize inference speed, and example models include YOLO, SSD and RetinaNet. Two-stage methods prioritize detection accuracy, and example models include Faster R-CNN, Mask R-CNN and Cascade R-CNN.

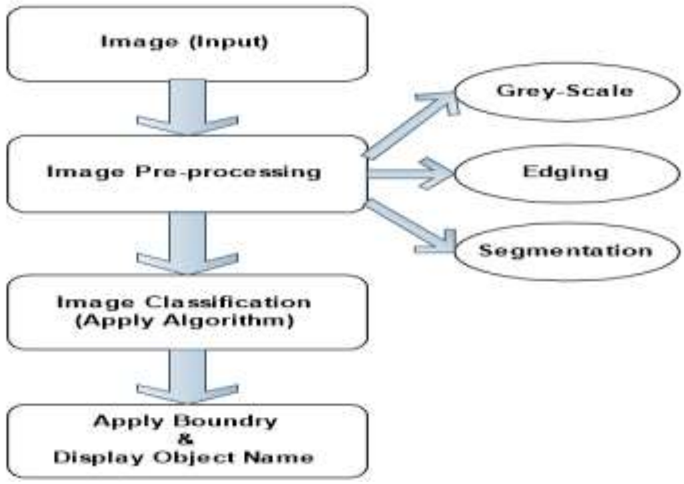

Fig.1:- System Architecture

IV. METHODOLOGIES

[4.1] Deep Learning: - Deep learning (also known as deep structured learning) is part of a broader family of machine learning methods based on artificial neural networks with representation learning. Learning can be supervised, semi-supervised or unsupervised.
Deep learning architectures such as deep neural networks, deep belief networks, recurrent neural networks and convolutional neural networks have been applied to fields including computer vision, machine vision, speech recognition, natural language processing, audio recognition, social network filtering, machine translation, bioinformatics, drug design, medical image analysis, material inspection and board game programs, where they have produced results comparable to and in some cases surpassing human expert performance.

Deep Learning is a subfield of machine learning concerned with algorithms inspired by the structure and function of the brain called artificial neural networks.

[4.2] Object detection: - Object detection is a computer technology related to computer vision and image processing that deals with detecting instances of semantic objects of a certain class (such as humans, buildings, or cars) in digital images and videos.

[4.3] Tensorflow: - TensorFlow is a free and open-source software library for dataflow and differentiable programming across a range of tasks. It is a symbolic math library, and is also used for machine learning applications such as neural networks. It is used for both research and production at Google. TensorFlow was developed by the Google Brain team for internal Google use. It was released under the Apache License 2.0 on November 9, 2015

[4.4] Single Shot Object detection (SSD):- The SSD architecture is a single convolution network that learns to predict bounding box locations and classify these locations in one pass. Hence, SSD can be trained end-to-end. The SSD network consists of base architecture followed by several convolution layers: SSD operates on feature maps to detect the location of bounding boxes. Remember – a feature map is of the size Df * Df * M. For each feature map location, k bounding boxes are predicted. Each bounding box carries with it the following information:

- 4 corner bounding box offset locations (cx, cy, w, h)
- C class probabilities (c1, c2, …cp)

SSD does not predict the shape of the box, rather just where the box is.

The k bounding boxes each have a predetermined shape. The shapes are set prior to actual training. For example, in the figure above, there are 4 boxes, meaning k=4.

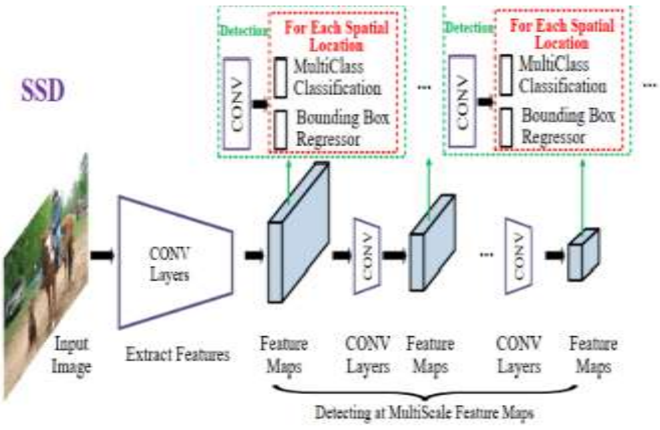

Fig. 2:- Generic SSD Flow

[4.5] Loss in MobileNet-SSD: - With the final set of matched boxes, we can compute the loss like this:

$$L = 1/N\ (L\ class + L\ box)$$

Here, N is the total number of matched boxes. L class is the softmax loss for classification and 'L box' is the L1 smooth loss representing the error of matched boxes. L1 smooth loss is a modification of L1 loss which is more robust to outliers. In the event that N is 0, the loss is set to 0 as well.

[4.6] MobileNet: - The MobileNet model is based on depthwise separable convolutions which are a form of factorized convolutions. These factorize a standard convolution into a depthwise convolution and a 1 × 1 convolution called a pointwise convolution. For MobileNets, the depthwise convolution applies a single filter to each input channel. The pointwise convolution then applies a 1 × 1 convolution to combine the outputs of the depthwise convolution. A standard convolution both filters and combines inputs into a new set of outputs in one step. The depthwise separable convolution splits this into two layers – a separate layer for filtering and a separate layer for combining. This factorization has the effect of drastically reducing computation and model size.

V. MATHEMATICAL MODEL

Set theory is the mathematical theory of well-determined collections, called sets, of objects that are called members, or elements, of the set. In set theory, however, as is usual in mathematics, sets are given axiomatically, so their existence and basic properties are postulated by the appropriate formal axioms.
Let 'S' be the system Where
S= I, O, P, Fs, Ss
Where,

I = Set of input,
O = Set of output
P = Set of technical processes
Fs = Set Of Failure State
Ss = Set Of Success State

Identify the input data I1, I2, , In
I = (Image)
Identify the output applications as O1, O2,,On
O = (Object detection)
Identify the Process as P
P = (image pre-processing, grey scale conversion, edging, object classification, boundary generation, tracking)
Identify the Failure state as Fs
Fs = (if Object not detected, more time required for detection)
Identify the Success state as Ss
Ss = (object detected properly, tracking work properly, if all task executed in less time and space)

VI. DATASET

We have use COCO dataset for proposed system. COCO is a large-scale object detection, segmentation, and captioning dataset. The COCO dataset is an excellent object detection dataset with 80 classes, 80,000 training images and 40,000 validation images. COCO has several features:

- Object segmentation
- Recognition in context
- Super pixel stuff segmentation
- 330K images (>200K labeled)
- 1.5 million object instances
- 80 object categories

VII. ACKNOWLEDGEMENT

I wish to express my profound thanks to all who helped us directly or indirectly in making this paper. Finally I wish to thank to all our friends and well-wishers who supported us in completing this paper successfully. I am especially grateful to our guide ***Prof. Priti J. Rajput*** for her time to time, very much needed, and valuable guidance. Without the full support and cheerful encouragement of my guide, the paper would not have been completed on time.

VIII. CONCLUSION

In this paper, novel approach for object detection and tracking has been presented using single shot object detection. The moving object detection is performed using TensorFlow object detection MobileNet-SSD. The object detection module robustly detects the object. Proposed system detect multiple objects like laptop, person, pen, chair, water bottle and etc. The detected object is tracked using MobileNet-SSD 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 approach and combined the global features and local characteristics in a natural way. The proposed approach achieves the accuracy of 96%.

REFERENCE

[1] Turan Goktug Altundogan, Mehmet Karakose "Multiple Object Tracking with Dynamic Fuzzy Cognitive Maps Using Deep Learning" 2019

[2] Yuchi Liu, Yujuan Qi, Yanjiang Wang "Object Tracking Based on Deep CNN Features Together with Color Features and Sparse Representation" 2019

[3] Bradley J. Koskowich, Maryam Rahnemoonfar, Michael Starek "VIRTUALOT - a framework enabling real-time coordinate Transformation & occlusion sensitive tracking using us products, Deep learning object detection & traditional object tracking Techniques.-" 2018

[4] Shraddha Mane, Prof. Supriya Mangale "Moving object detection and tracking Using Convolutional Neural Networks" 2018

[5] Rasika Phadnis, Jaya Mishra, Shruti Bendale "Objects Talk - Object detection and Pattern Tracking using TensorFlow"

[6] Song, Y. M. and Jeon, M., 2016, Online multiple object tracking with the hierarchically adopted gm-phd filter using motion and appearance, IEEE International Conference on Consumer Electronics-Asia, 1-4.

[7] Wei, L., and Xingwei, L, 2016, Multiple object tracking based on modified algorithm of GMMCP tracker, 2016 IEEE International Conference on Signal and Image Processing, 11-15.

[8] Zhang, Y., Huang, Y., and Wang, L., 2017, Multi-task Deep Learning for Fast Online Multiple Object Tracking, 2017 4th IAPR Asian Conference on Pattern Recognition (ACPR), 138-143.

[9] Jahanshahi, P., Masoud, A. and Moghadam, E., 2015, Multi-view tracking using Kalman filter and graph cut., 2015 AI & Robotics (IRANOPEN), 1-5.

[10] Cuan, B., Idrissi, K. and Garcia, C., 2018, Deep Siamese Network for Multiple Object Tracking, 2018 IEEE 20th International Workshop on Multimedia Signal Processing 1-6.

[11] Altundoğan, T. G., and Karaköse, M. (2018, September). An Approach for Online Weight Update Using Particle Swarm Optimization in Dynamic Fuzzy Cognitive Maps. In 2018 3rd International Co.

