ISSN 2395-1621

Video Image Processing For Moving Object Detection And Segmentation

Vishwanath Murlidhar Ohal, Prof. Priti J. Rajput

vishuohal358@gmail.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

The history of segmentation of digital images using computers could be traced back 40 years. Since then, this field has evolved very quickly and has undergone great change. In this chapter, the position of image segmentation in the general scope of image techniques is first introduced; the formal definition and extension of image segmentation as well as three layers of research on image segmentation are then explained. Based on the introduction and explanations, statistics for a number of developed algorithms is provided, the scheme for classifying different segmentation algorithms is discussed and a summary of existing survey papers for image segmentation is presented. These discussions provide a general rendering of research and development of image segmentation in the last 40 years.

Video/Moving objects often contain almost important information for surveillance videos, traffic monitoring, human motion capture etc. Background subtraction methods are widely exploited for moving object detection in videos in many applications. Moving object segmentation is the application in video processing. Segmentation helps in detecting various features of moving objects for further video/image processing.

Keywords: ***Video Processing, Video Segmentation, Deep Learning, Object Detection, Object Tracking, Tensorflow, COCO Dataset***

ARTICLE INFO

Article History

Received: 27th July 2020

Received in revised form : 27th July 2020

Accepted: 29th July 2020

Published online :

4th August 2020

I. INTRODUCTION

In this paper we will have an overview of Detection and segmentation of moving objects in video streams. It is an essential process for information extraction in many computer vision applications, including video surveillance, human tracking, and traffic monitoring and semantic annotation of videos. Surveillance system uses video cameras to monitor the activities of targets (human, vehicle, etc.) in a scene. In order to obtain an automatic motion segmentation algorithm that can work with real images there are several issues that need to be solved, particularly important are: noise, missing data and lack of a priori knowledge. One of the main problems is the presence of noise. For some applications the noise level can become critical. There are three conventional approaches to moving object detection: background s subtraction, temporal differencing and optical flow.

Recently, background modeling and subtraction became the most popular technique for moving object detection in computer vision, such as object recognition and traffic surveillance.

Compared to optical flow and interface difference algorithms, background subtraction algorithm needs less computation and performs better, and it is more flexible and effective. The idea of background subtraction is to differentiate the current image from a reference background model. These algorithms initialize a background model at first to represent the scene with no moving objects and then detect the moving objects by computing the difference between the current frame and the background model. Dynamic background is a challenge for background subtraction, such as waving tree leaves and ripples on river. In the past several years, many background subtraction algorithms have been proposed, and most of them focus on building more effective background model to handle dynamic background as follows:

(1) Features: texture and color

(2) Combining methods: combining two or more background models as the new model

(3) Updating the background model

In this paper, a new pixel wise and nonparametric moving object detection method is proposed. Background model is built by the first frames and sampling times in 3 × 3 neighborhood region randomly. On the one hand, spatiotemporal model represents dynamic background scenes well. On the other hand, a new update strategy makes the background model fit the dynamic background. In addition, the proposed method can deal with ghost well. Experimental results show that the proposed method can efficiently and correctly detect the moving objects from the dynamic background.

II. LITERATURE SURVEY

[1] Detection, identification and tracking of objects during the motion
Author Name: - Lirie Koraqi, Florim Idrizi

This paper intends to introduce discovery of various objects, object classification, object tracking algorithms including analysis and a comparison of different techniques used for different stages of tracking. The purpose of tracking objects is segmenting a region of interest from a video scene and continuing tracking movement, positioning, and match. Detection of the object and classification are the steps preceding the object tracking in a sequence of images. Object detection is performed to control the existence of objects in the video and find that object, and then the detected object can be classified into many categories such as people, vehicles, floating trees, trees and other moving objects. In this paper is elaborated the autonomous tracking system with the aim of opposing movement in the event of a Demonstration of human tracking, as such may enable the former to be distributed as it comprises more than one control center, key component parts of the system include: The camera and its 'interface, Arduino viewing boards. PC - Controlled Center.

[2] Intelligent Vision with TensorFlow using Neural Network Algorithms
Author Name: - Ms. A.Ramya Visalatchi, Ms. T.Navasri, Ms. P.Ranjanipriy, Ms. R.Yogamathi
Computer vision and video analytics are the torrid research area in Machine learning and their establishment process traditionally starts with object detection and eventually tracking. In recent years, there is a tremendous growth in performing comprehensive study based on the field of object detection and Pattern Analysis. In our system we have improvised and experimented with detection method based on machine learning and deep learning approach in object recognition and pattern analysis. We assume object detection as a retrogression problem to spatially separated corresponding class probabilities and bounding boxes. Many prominent algorithms have been designed for object detection, Pattern Analysis and tracking, which also includes edge tracking, color segmentation and pattern matching. A single neural network is capable of predicting class probabilities and bounding boxes directly from the full image per cycle. Therefore we have used various neural network algorithms such as YOLOv3, Single Shot Multiple detection algorithm to carry out video analysis using object detection and drowsiness detection using pattern or behavior analysis with the help of Tensorflow. The framework will recognize object continuously, from the input perceived through camera where it can apparently capture a required frames to predict the object and also to match the pattern. It has been accomplished using real-time video processing and a single camera. The proposed system is versatile to operate in complex, real time, non-plain environment.

[3] Video Instance Segmentation 2019: A winning approach for combined Detection, Segmentation, Classification and Tracking.
Author Name: - Jonathon Luiten, Philip H.S. Torr, Bastian Leibe

Video Instance Segmentation (VIS) is the task of localizing all objects in a video, segmenting them, tracking them throughout the video and classifying them into a set of predefined classes. In this work, divide VIS into these four parts: detection, segmentation, tracking and classification. We then develop algorithms for performing each of these four sub tasks individually, and combine these into a complete solution for VIS. Our solution is an adaptation of UnOVOST, the current best performing algorithm for Unsupervised Video Object Segmentation, to this VIS task. We benchmark our algorithm on the 2019 YouTube-VIS Challenge, where we obtain first place with an mAP score of 46.7%.

[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 video using deep learning algorithms.

Object detection is the task of detecting instances of objects of a certain class within a video and image.

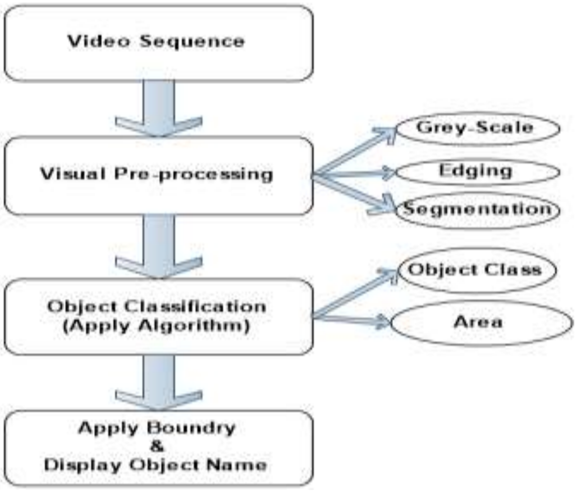

Fig.1:- System Architecture

IV. METHODOLOGIES

[4.1] Video Segmentation: Video (temporal) segmentation is the process of partitioning a video sequence into disjoint sets of consecutive frames that are homogeneous according to some defined criteria. In the most common types of segmentation, video is partitioned into shots, camera-takes, or scenes. A camera take is a sequence of frames captured by a video camera from the moment it starts capturing to the moment it stops. During montage, camera takes are trimmed, split, and inserted one after the other to compose an edited version of a video. The basic element of an edited video is called shot. A shot is a contiguous sequence of frames belonging to a single camera take in an edited video. Content-wise, shots usually possess some degree of visual uniformity. A sceneis a group of contiguous shots that form a semantically.

[4.2] Tensorflow Object detection: - The TensorFlow object detection API is the framework for creating a deep learning network that solves object detection problems. There are already pretrained models in their framework which they refer to as Model Zoo. This includes a collection of pretrained models trained on the COCO dataset, the KITTI dataset, and the Open Images Dataset. These models can be used for inference if we are interested in categories only in this dataset. They are also useful for initializing your models when training on the novel dataset. The various architectures used in the pretrained model are described in this table.

Model name	Speed	COCO mAP	Outputs
ssd_mobilenet_v1_coco	fast	21	Boxes
ssd_inception_v2_coco	fast	24	Boxes
rfcn_resnet101_coco	medium	30	Boxes
faster_rcnn_resnet101_coco	medium	32	Boxes
faster_rcnn_inception_resnet_v2_atrous_coco	slow	37	Boxes

Table 1. Various Model Comparison

[4.3] 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.

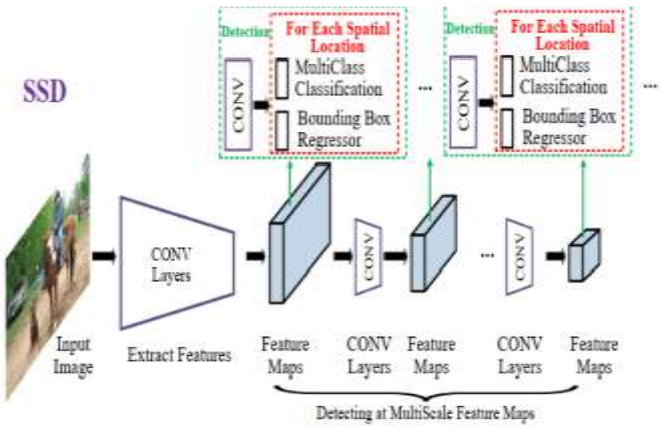

Fig. 2:- Generic SSD Flow

[4.4] 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 = (Live video)
Identify the output applications as O1, O2,,On
O = (Object detection, Object tracking)
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 with video processing and deep learning algorithm. The moving object detection is performed using TensorFlow object detection MobileNet-SSD. The object detection module robustly detects the object from live camera video. Proposed system detect multiple objects like laptop, person, pen, chair, water bottle and etc. The detected object is tracked using MobileNet-SSD algorithm. The proposed approach achieves the accuracy of 96% for standard COCO dataset.

REFERENCE

[1] Lirie Koraqi, Florim Idrizi"Detection, identification and tracking of objects during the motion" 2019

[2] Ms. A.Ramya Visalatchi, Ms. T.Navasri, Ms. P.Ranjanipriy, Ms. R.Yogamathi "Intelligent Vision with TensorFlow using Neural Network Algorithms" 2019

[3] Jonathon Luiten, Philip H.S. Torr, Bastian Leibe "Video Instance Segmentation 2019: A winning approach for combined Detection, Segmentation, Classification and Tracking" 2019

[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)

