An Analysis on Entity Recognition and Existence of Images in Video

Maheswari M[a,1], Josephine M S[b] and JeyabalarajaV[c]

[a]*Research Scholar, Deptof CSE, Dr.M.G.R. Educational and Research Institute*

[b]*Professor, Dept of Computer Applications, Dr.M.G.R.Educational and Research Institute*

[c]*Professor, Dept. of CSE,Velammal Engineering College, Chennai, India*

Abstract: Humans can effortlessly notice and perceive objects found in a photograph and video. Human visual system is speedy, correct and performs complicated tasks like identifying multiple entities and spot obstacles. An Entity is referred as an object. Object detection in videos entails verifying the presence of an item in image sequences. It is related to many real time appliances like car belief, video surveillance and so forth. The aim of this paper is to investigate and assess the distinctive strategies that are used in item tracking and detection through the usage of video sequences.

Keywords:. Object Detection, Classification, Background Subtraction, Deep Learning

1 Introduction

Identifying Objects in a video is a challenging task. Object Detection involves identifying an object in an image or a video. Detection of an object can be done in three main stages, Image Classification, Object Localization and Object Detection. Classification identifies the type or class of an object in a frame. Localization locates the object in a frame with a bounding box. Video is a sequence of frame. A frame can be considered as an image. From a frame background model will be constructed and when background has been constructed then objects and its features can be classify and then it can be identified. Frame differencing finds the difference between two successive images to decide the presence of moving objects. Background subtraction constructs background model and the changes that are identified in a frame are considered as foreground. Optical flow uses distribution characteristics of pixels of object. Huge calculations are involved and it is sensitivity to noise. [9] Applications of Object identification involve Visual surveillance, retrieval and autonomous and assisted driving, detection of faces, detection of humans, vehicles, etc.

---

[1]Maheswari M, Research Scholar, Dept of CSE, Dr.M.G.R. Educational and Research Institute, Chennai;
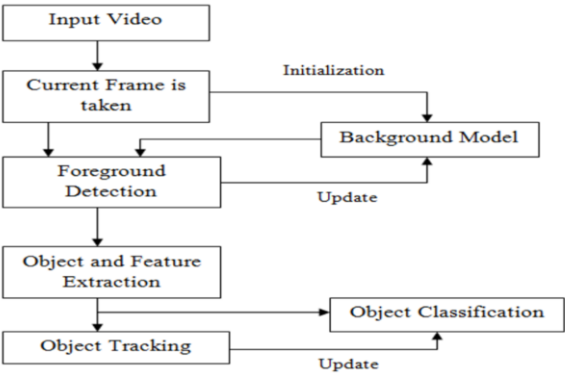E-mail: m.mahe05@gmail.com.
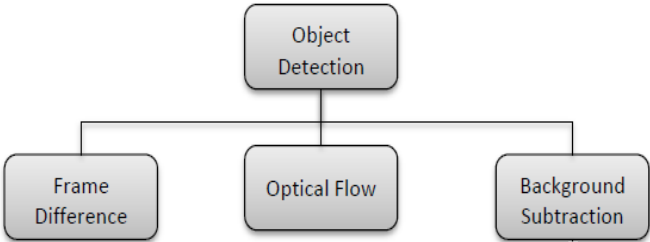
Figure 1. Steps involved in identifying Objects

Figure 2. Traditional classification of Object Detection

This is a study about conventional methods and CNN that are used in identifying objects.

2   Literature Review

[1][11] Image Space reduction is done to reduce spatial noise which occurs because of unsteady input i.e input received from handheld devices. Color noise is produced digital camera and the flicker of lights.This color noise is eliminated by Color Space reduction method.Image- and Color-space reduction is used for background subtraction which is used for detecting objects from dynamic background. Image Space Reduction is performed by, P×Q→ R× S, Where P & Q are width and height of image frame, R & S are width and height of reduced image space.While using digital cameras color noises are produced because of flicker of lights.Pixel values are taken from RGB values. The color space is mapped from Red * Green * Blueto Red' * Green' * Blue'. The limitation is it occupies more memory for dynamic backgrounds.[2] Performs Motion Mask i.e., performs removal of motionless video frames and thenGraphs(

$$\rho_t(i, j) = \frac{\sum_{k=1}^{p} m_{i,j,k}(d_{i,k} - \bar{d}_i)(d_{j,k} - \bar{d}_j)}{\sqrt{\sum_{k=1}^{p} m_{i,j,k}[(d_{i,k} - \bar{d}_i)^2 + (d_{j,k} - \bar{d}_j)^2]}}$$

Spatial and temporal) are formed by correlation coefficient.Spatial graph is constructed for a single image.Temporal graph is constructed for series of images at different time Fused Lasso is used for enhancing Foreground Detection.Sparse Subspace Clustering is used for completing incomplete information when the information is represented in a low dimensional space. The algorithm used by the authors is Motion-assisted Spatiotemporal Clustering of Low-rank (MSCL). Spatial and Temporal graph are constructed using Motion-aware CorrelationCoefficient to handle background modelling. Temporal Graph is constructed for the sequence of frames as

The main advantage of this work is it Generates an precise background model even in the presence of occlusions, clutter, jitter, and abrupt intensity variations. This algorithm does not work well for videos captured from moving and PTZ cameras(pan–tilt–zoom camera: security camera) and in more crowded area.

[3] Tensor Principal component analysis (TPCA) converts correlated measurements into components. This method combines many of the low rank component of block tensor. Background and foreground are detached using a novel method called IBTVST. The IBTSVT improves the efficiency by parallel processing.Authors have employed the face detection algorithm, Viola-Jones algorithm to detect faces and the eyes. This algorithm Specifies the entire face must point towards the camera. The main drawback is reduces the algorithms utility.

[4] Robust Principal Component Analysis (RPCA)is used for Moving Object Detection. Background and foreground is modelled using tensor framework. To distinguish from foreground and background, Saliently Fused Sparsity (SFS) Regularization is used and to have smooth solutions for the moving object 3 Dimensional Locally Adaptive total variation (3D-LATV) is used. To get regularization for the foreground, three modules (3DLATV, l1 norm, and saliency)called as "saliently fused sparsity". Background is regularized by using Tensor nuclear norm (TNN).This does not concentrate on camouflage, bootstrap and moving cameras

[5] Authors have used De-Noising to remove noises and Low Rank approximation for object detection. Background works well for static camera, constant illumination and static background.Low Rank Approximation,is used for mathematical modelling and compression. DNLRl1TV method combines low rank minimization and l1, l2 and TV-norm regularizations for background separation. l1-norm is used for identifying spikes at sparse points. Spatial Total Variation (TV) norm conserves edge information. The TV regularization is used for enhancing   spatio-temporal continuity. The authors have tested with different types of noises like Gaussian, Poisson, salt & pepper, speckle noises. This is also tested for a background which is dynamic, when the weather is bad and camera jitter.Comparative Study of Object Detection

Table 1. Comparison of Conventional Methods

Method	Accuracy	Time Efficiency
Background Subtraction	Medium	Medium
Optical Flow	Moderate	High
Frame Differencing	High	Low to Medium

2.1.   *Deep Learning Models*

[6] Deep Learning models are referred as neural networks with deep structure. Deep learning based algorithms are used for object detection and classification like Convolutional Neural Network(CNN), Region based Convolutional neural network(R-CNN), Fast Region based Convolutional neural network(Fast R-CNN), Faster Region based Convolutional neural network(Faster R-CNN), and You Only Look Once(YOLO).

[7] Datasets plays a vital role of object identification. To train deep neural networks, large amount of data's such as (images and videos) and time are required. The main datasets that are used for object identification are PASCAL VOC, ImageNet, MS COCO and Open Images.

[8] Convolutional Neural Network(CNN): CNN performs image classification. The layers that are present in CNN are convolutional layer, pooling layer and a Fully Connected Layer. Convolution performs blur of image, edge detection, blur and sharpen of images by applying filters. When size of image is large, Pooling layer, reduces the size of the information. Fully Connected Layer flattens the matrix by converting into single vector which can be given as the input for the next stage.
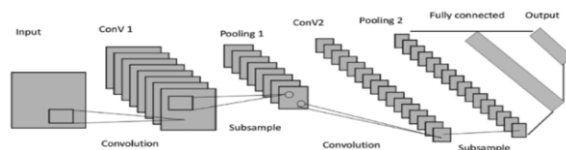
Figure 3. Convolutional Neural Network

Input is sent to the various layers that is convolutional and pooling layers. The image is given to CNN which divides the image into various regions. Each of these regions is considered as an image. These images are passed to CNN to classify into various classes. This does not work well when object occupies a large portion in the image or when object occupies a less portion in the image. The main drawback of CNN is huge number of regions are selected for identifying objects. This results in high computational cost.

Region- based Convolutional Neural Network (R-CNN)

RCNN uses selective search algorithm for extracting 2000 regions from a image called Region Proposals. These regions are passed to CNN which acts as a feature extractor. The extracted features are passed to SVM to classify the objects, and forms bounding boxes for the objects. The main drawback of are it is time taken to classify all the 2000 regions are very high. Selective search is Fixed algorithm and so learning does not takes place here.
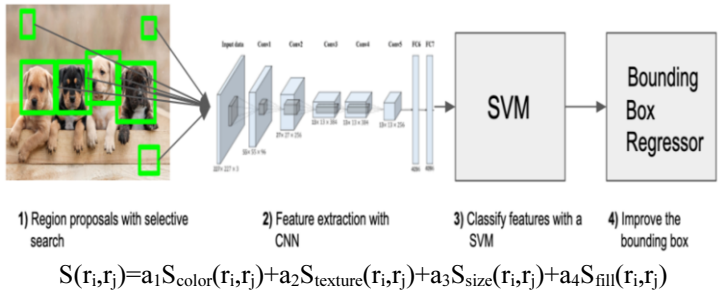
$$S(r_i,r_j)=a_1 S_{color}(r_i,r_j)+a_2 S_{texture}(r_i,r_j)+a_3 S_{size}(r_i,r_j)+a_4 S_{fill}(r_i,r_j)$$

Figure 4. Region based Convolutional Neural Network

Fast R-CNN

Rather than passing the regions proposals to CNN, the input is passed to the CNN for generating features. And features are passed to softmax layer for predicting the class of an object. When an object has been cut in an region then it can be adjusted using offset values for the bounding boxes. In R-CNN convolution operation is done once per image. Fast R-CNN" is faster than R-CNN because,every time it is not necessary to feed 2000 region proposals to the CNN. As an alternative, the convolution operation is done once per image and from which a feature map is generated.
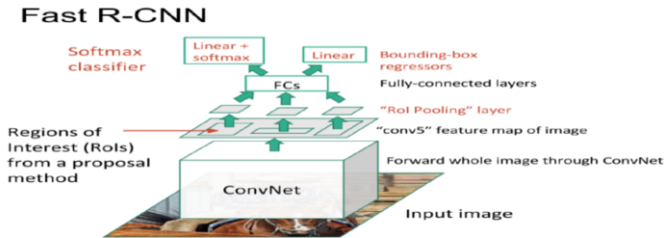

Figure 5. Fast R-CNN

Faster R-CNN

R-CNN and Fast R-CNN uses selective search which is fixed search algorithm to identify regions.Where as Faster R-CNN uses network learning algorithm for identifying region proposals and so this improves the performance of the network.

Figure 6. Faster R C-NN

You Only Look Once (YOLO)

YOLO is a CNN for detecting Objects in real time. The full image is passed to neural network which divides it into various regions for predicting bounding boxes. Bounding boxes are weighted with probabilities. The input given to YOLO is partitioned into a S* S grid of cells. Where each cell is helps in object prediciton.

Figure 7. YOLO

Table 2. Comparison of speeds and performances for models trained with   PASCAL VOC datasets

Model	mAP	FPS	Real Time Speed
Fast R-CNN	70.0%	0.5	No
Faster R-CNN	62.1%	18	No
YOLO	63.4%	45	Yes

3. Conclusion

The goal of this paper is to present a top level view of object classification. The overview starts with conventional methods on item detection and leads to Deep Learning. It defined, briefly, about object classification and how it progresses. Hence, it acted as a manner to introduce the idea of item categorizing as a start point to readers who're unknown with this concept and to provide an inspiration for more advanced

researchers. This evaluation is significant for the advancement in neural networks and associated system studies that affords valued insights and guiding principles for upcoming progress.

References

[1] HasupLee ; HyungSeok Kim ; Jee-In Kim ,Background Subtraction Using Background Sets With Image and Color Space Reduction.IEEE(2016).

[2] SajidJaved, ArifMahmood, Thierry Bouwmans, and Soon Ki Jung, Background–Foreground Modeling Based on Spatiotemporal Sparse Subspace Clustering.IEEE(2017).

[3] LongxiChen,Yipeng Liu, Ce Zhu, Iterative Block Tensor Singular Value Thresholding For Extraction Of Low Rank Component Of Image Data,arXiv.org(2017).

[4] Wenrui Hu, Yehui Yang, Wensheng Zhang, and Yuan Xie, Member, Moving Object Detection Using Tensor-Based Low-Rank and Saliently Fused-Sparse decomposition. IEEE (2017).

[5] Shijila B., Anju Jose Tom, Sudhish N. George,Simultaneousdenoising and moving object detection using low rank approximation, Elsevier,(2019).

[6] Zhong-QiuZhao,PengZheng,Shou-taoXu, and Xindong Wu, Object Detection with Deep Learning: A Review.IEEE Transaction(2019).

[7] Li Liu, WanliOuyang, Xiaogang Wang, Paul Fieguth, Jie Chen, Xinwang Liu & MattiPietikäinen, Deep Learning for Generic Object Detection: A Survey, Springer(2019).

[8] KavithaSubramani and Murugavalli S , Recognizing Ancient Characters from Tamil Palm Leaf Manuscripts using Convolution Based Deep Learning, International Journal of Recent Technology and Engineering,(2019), Volume 10, Issue 10, pp 6873-6880.

[9] ShreyamshKamate,NuriYilmazerApplication of Object Detection and Tracking Techniques for Unmanned Aerial Vehicles,Science Direct,(2015).

[10] AbdulVahab, MarutiSNaik, PrasannaGRaikar,Prasad S R, Applications of Object Detection System.International Research Journal of Engineering and Technology(2019).

[11] ApoorvaRaghunandan, Mohana, PakalaRaghav and H. V. Ravish Aradhya,Object Detection Algorithms for Video Surveillance Applications, International Conference on Communication and Signal Processing,(2018).