TOPIC
OBJECT DETECTION BY TENSORFLOW

Name:- Piyush Mahajan
Reg No:-11803539
Section:-KM073

Introduction

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. Well-researched domains of object detection include face detection and pedestrian detection. Object detection has applications in many areas of computer vision, including image retrieval and video surveillance

Literature review

In the previous study most of them have concentrated towards Object detection (Ben Ayed et al., 2015; Najva and Bijoy, 2016; Ramya and Rajeswari, 2016; Risha and Kumar, 2016; Shen et al., 2013; Soundrapandiyan and Mouli, 2015; Viswanath et al., 2015) ,Object tracking (Bagherpour et al., 2012; Coşkun and Ünal, 2016; Foytik et al., 2011; Lee et al., 2012; Poschmann et al., 2014; Weng et al., 2013; Yilmaz et al., 2006; Zhang et al., 2016) and Object recognition (Chakravarthy et al., 2015; Elhariri et al., 2015; Gang et al., 2010; Ha and Ko, 2015; Nair et al., 2011) for tracking the object using video sequences.

Studies related to object detection

The detection of an object in video sequence plays a significant role in many applications. Specifically as video surveillance applications (Amandeep and Goyal, 2015).
The previous studies related to object detection is discussed as follows: Shen et al. (2013) proposed a novel hierarchical moving target detection method based on spatiotemporal saliency. Further, they achieved the refined detection results using temporal and spatial saliency information. The experimental results show that this approach identifies moving substances in the airborne video with high accuracy and efficiency. Additionally, this method does not have the effect of time delay when compared with HMI technique. However, this method evaluated object locations in all the video frames as self-sufficiently, false alarms which are inevitable. Guo et al. (2012) suggested object detection approach for tracking the objects in video frames.
The simulation result shows this technique was effective and accurateness, robust for generic object classes' detection with good performance. Further needs to focus towards increase classification accurateness in real-time object recognition. Ben Ayed et al. (2015) proposed a method for detection of text data based on a texture in video frames by big data analytics. The video frames are decomposed into various fixed size blocks and these blocks are analyzed using har wavelet transform technique. Further, they used a neural network to classify the text and non-text blocks. However, this study needs to concentrate on extracting the regions towards remove the noisy regions as well as exclude the text like sections. Viswanath et al. (2015) suggested and modeled the approach using non-panoramic background modeling. By the use of this approach, they modeled the entire picture element with one Spatio-temporal Gaussian. The simulations result shows this method able to identify the moving substances with fewer false alarms. However, this method fails once the adequate features are not obtainable from the section. Soundrapandiyan and Mouli (2015) suggested a novel and adaptive method for pedestrian detection. Further, they separated the foreground objects from the background by image pixel intensities. Subsequently, they used high boost filter for enhancing the foreground edges.
The efficacy of the proposed method is evident from the subject evaluation results as well as objective evaluation with around 90% of pedestrian's detection rate compared to the other single image existing methods. In future, they planned to improve the performance of the method with higher detection rate and low false positives on par with sequence image methods. Ramya and Rajeswari (2016) suggested a modified frame difference method which uses the correlation between blocks of current image and background image to categorize the pixels as foreground and background. The blocks in the current image which are highly correlated with the background image are considered as background. For the other block, the pixel-wise comparison is made to categorize it as foreground or background. The experiments conducted proved this approach improves the frame difference method particularly as finding accuracy with speed. However, this study needs to focus towards other information available in the blocks such as shape and edge can be used to improve the detection accuracy. Risha and Kumar (2016) suggested an optic flow with the morphological operation for object detection in video. Further applied morphological operation towards obtaining clear moving target image. This study only concentrated on static camera. So need to focus on moving the camera as well as identify multiple objects in video frames. Najva and Bijoy (2016) proposed a model for detection and classification of objects in videos by combining Tensor features with SIFT approach towards classifying the detected objects using Deep Neural Network(DNN). The DNN capable of handling large higher dimensional data with billions of parameters as like human brain. Simulation results obtained illustrate that the proposed classifier model produces more accurate results than the existing methods, which combines both SIFT and tensor features for feature extraction and DNN for classification.

Studies related to object tracking

The object tracking is the term which used to identify the moving object position as well as tracking them from video sequences (Balasubramanian et al., 2014). The tracking method is classified into three types such as kernel, point and silhouette based tracking (Yilmaz et al., 2006). Compared to silhouette method, existing most of them have focused on kernel-based method due to high accuracy with less computational cost. However, the point tracking method has less computational cost with reduce in accuracy (Weng et al., 2013).
The previous studies related to object tracking method are discussed as follows: Sarkar et al. (2012) developed color information method for identification of skin in image occurred from FERET especially for mouth and eye region detection. This technique requires less computational cost as well as applied to video sequences. However, this method incapable towards identifying the small face images which away from the camera.
Weng et al. (2013) proposed an algorithm for improving the performance of natural feature selection in the real world. Further, they used to speed up robust features (SURF) for features extraction from live mobile camera image and recognition.
These extracted features are calculated using pose matrix through Homography approach. The simulation result shows this algorithm tracked and recognized the object from natural features in easy, speedy with suitable way. However, its speed and accuracy need to be improved. Zhang et al. (2016) proposed an approach by combining frame difference and nonparametric method for video analysis traceability. The simulation result proved this approach performance was better than the traditional frame difference and GMM. Further, it can able to remove the noise from a background which gives us the ability to detect the moving object more precisely in the applications such as food and agriculture related product traceability analysis. However, this study needs to enhance the capability of traceability system and supply the visual supply chain for the common user to ensure the safety. Coşkun and Ünal (2016) suggested camshift technique towards track the object from the video sequences. Further, they demonstrated this approach successful carried out in mobile platform even with the change in object size and illumination. The drawback observed in this proposed technique fails to perform for input video with full occlusion. Houssineau et al. (2016) suggested a parameterization based disparity space for non-rectified camera networks, extended to moving objects, and integrated into a Bayesian multi-object tracking with sensor calibration technique.
The performance of the obtained framework has been demonstrated for camera calibration on simulated with real data, underlying the problems of single-object localization and tracking, as well as for multi-object tracking. Further, they planned to enhance the proposed method towards other multi-object filters and a comparative study of these approaches for camera-based tracking as well as for camera calibration. Oiwa et al. (2016) suggested probabilistic background model towards tracking the object from video sequences. The simulation results show the accuracy and effectiveness of this method high compared to previous technique. However, this study needs to concentrate on higher speed as well as improve the accuracy of object tracking. Mohammed and Morris (2014) suggested a color-based technique which was the combination of accruing and normalizing histograms towards object tracking under different conditions based on a mobile device. This technique was easy to use and robust against varying illuminations. However, this technique fails to identify the entire region of symbols due to the extreme camera view. Aggarwal et al. (2006) suggested a novel technique which was the combination of motion estimation and background subtraction for object tracking using video sequences. The system mainly focused on four scenarios such as interpolation, identify the object, subtract the background and object selection. However, this study needs to focus more towards full occlusions of video sequences, multiple object tracking, fast camera motion and unsupervised object tracking.
Fatima et al. (2013) suggested image segmentation approach for track objects of interests through specifying the color intensities. A minimum distance classifier approach is used for object classification. They attain the object tracking through specifying the object centroids in all the video frames. The simulation results proved this technique was more efficient for contextual approximation. However, this study needs to focus towards implementing segmentation which would work with occluded images with multiple objects and computationally more efficient. Lecumberry and Pardo (2005) proposed an algorithm for semi-automatic object tracking in videos by various features with probabilistic relaxation method. The proposed algorithm performs effectively for object tracking specifically for object borders is smooth and accurate. When combined with sophisticated methods it performs effectively for object tracking. The accuracy of the borders of the tracked objects depends on the power of discrimination of selected features as well as the appearance of new objects and/or background. However, this study needs to focus towards improving the accuracy of this algorithm, particularly as rigid objects.

Studies related to object recognition

Gang et al. (2010) developed a kernel locality preserving projections (KLPP) towards improving the accuracy and separable description of the objects. The simulation results proved, this approach more appropriate for space object recognition mainly considering changes of viewpoints. However, this study needs to concentrate on improving the accuracy level towards recognizing with limited trained objects as reference. Also, test the effectiveness of this approach on images of other objects models. Nair et al. (2011) developed a technique for a combination of face recognition, tracking and detection approach towards identifying the individual faces. The simulation result shows this method increases the level of accuracy in recognizing and tracking faces and applicable to the real-time application. However, it fails in varying illumination conditions as well as the existence of lesser background in face recognition which limits the system performance

Zhang and Jiang (2014) suggested regression based kernel technique on behalf of identifying multiview objects as well as approximating their poses. The simulation results proved obtained results are improved recognition, while comparative analyses with state-of-the-arts. Further, they authenticated the robustness and efficiency of this approach. However, the existing research needs to focus more attention on resolving issues of object space recognition Chakravarthy et al. (2015) suggested and estimated a technique for video sequence stream processing.

Further, they demonstrated the various types of situations which were based on arable and relational illustrations. However, this study fails to express complex situations. So, needs to focus towards accepting feature vectors as well as bounding boxes towards identifying the temporal and spatial computations. Ha and Ko (2015) proposed vision-based shadow gesture recognition method for interactive projection systems. This method only splits the shadow area through merging the binary image with an input image using a learning algorithm that isolates the background from the input image.

The developed approach isolates or differentiates shadow of a hand in based on convex hull, moment and defect in each region. After that, isolated the hand shadows by the convex hull, defect, and moment of each region. However, the robustness of this processing is not always assured. It means that at some conditions the gesture recognition fails, and unexpected results would happen. Elhariri et al. (2015) discussed Random Forests (RF), Linear Discriminant Analysis (LDA) classifiers, Support Vector Machines (SVMs).Furthermore, for edge detection segmentation approach has been utilized in this research. Subsequently extracted the features using morphological operations. The simulation result shows the LDA and SVM provide better performance. However, the input dataset as elderly needs to be involved.

Proposed Architecture

Object detection and image segmentation are two related visual tasks that extend past simple classification and are more complex to architect for than simple classification. In **object detection** tasks, we are to predict the bounding box for objects in an image. For **image segmentation**, we are to divide the image into different aspects of the scene, on a pixel-by-pixel basis.

These are both intrinsically very hard tasks that require interesting model architectures to address.

Result and Analysis

When play low vision frame. Extract the frame. result are compare original frame and enhance frame. Enhance frame are get apply all algorithm for improve quality and enhance the degraded frame

CONCLUSION

Due to its powerful learning ability and advantages in dealing with occlusion, scale transformation and background switches, deep learning based object detection has been a research hotspot in recent years. This paper provides a detailed review on deep learning based object detection frameworks which handle different sub-problems, such as occlusion, clutter and low resolution, with different degrees of modifications on R-CNN. The review starts on generic object detection pipelines which provide base architectures for other related tasks. Then, three other common tasks, namely salient object detection, face detection and pedestrian detection, are also briefly reviewed. Finally, we propose several promising future directions to gain a thorough understanding of the object detection landscape. This review is also meaningful for the developments in neural networks and related learning systems, which provides valuable insights and guidelines for future progress.

Future Scope

Object detection in images and video has received lots of attention in the computer vision and pattern recognition communities over recent years. We have had great progress in the field, processing a single image used to take 20 seconds per image and today it takes less than 20 milliseconds.Of the problems related to these fields, analyzing an image and recognizing all objects remains to be one of the most challenging ones.For humans and many other animals, visual perception is one of the most important senses; we heavily rely on vision whenever we interact with our environment. In order to pick up a glass, we need to first determine which part of our visual impression corresponds to the glass before we can find out where we have to move our hands in order to grasp it.The same code that can be used to recognize Stop signs or pedestrians in a self driving vehicle signs can also be used to find cancer cells in a tissue biopsy.If we want to recognize another human, we first have to find out which part of the image we see represents that individual, as well as any distinguishing factors of their face.Notably, we generally do not actively consider these basic steps, but these steps pose a major challenge for artificial systems dealing with image processing.Most existing algorithms only tackle a small subset of the different tasks necessary for understanding an image and are very expensive computationally. In order to reproduce a fraction of the average person’s ability to detect objects, one would have to combine several different algorithms to make a combined system that runs in real time, an enormous challenge with today’s hardware.Indeed, object detection is a key task for most computer and robot vision systems. Although there has been great progress in the last several years, there will be even bigger improvements in the future with the advent of artificial intelligence in conjunction with existing techniques that are now part of many consumer electronics or have been integrated in assistant driving technologies.

However, we are still far from achieving human-level performance in open-world learning.

Furthermore, object detection has not been applied in many areas where it could be of great help. Consider for example the possibility of applications of object detection systems to robotic excavation when venturing into previously unexplored territory, such as the deep sea or other planets, in which the detection systems will have to learn new object classes on the job. In such cases, a real-time, open-world learning ability will be critical.This fascinating computer technology related to computer vision and image processing that detects and defines objects, such as persons, vehicles, and animals from digital images and videos, will be incredibly important in the near future.We have developed many methods for object detection, but the application of deep learning promises higher accuracy for a wider variety of object classes.

Object detection is breaking into a wide range of industries, including computer vision, image retrieval, security, surveillance, automated vehicle systems, and machine inspection.Although the possibilities are endless when it comes to future use cases for object detection, there are still significant challenges remaining.Herewith are some of the main useful applications of object detection: Vehicle’s Plates recognition, self-driving cars, Tracking objects, face recognition, medical imaging, object counting, object extraction from an image or video, person detection.

The future of object detection technology is in the process of proving itself, and much like the original Industrial Revolution, it has the potential to free people from menial jobs that can be done more efficiently and effectively by machines. It will also open up new avenues of research and operations that will reap additional benefits in the future.

Thus, these challenges circumvent the need for a lot of training requiring a massive number of datasets to serve more nuanced tasks, with its continued evolution, along with the devices and techniques that make it possible, it could soon become the next big thing in the future.

Reference

- International Journal of Computational Intelligence Research ISSN 0973-1873 Volume 13, Number 5 (2017), pp. 745-765

- Object Detection with Deep Learning: A Review Zhong-Qiu Zhao, Member, IEEE, Peng Zheng, Shou-tao Xu, and Xindong Wu, Fellow, IEEE
- blog.rebellionresearch.com/blog/the-future-of-object-detection
- wikipedia.org
- Imperial Journal of Interdisciplinary Research (IJIR) Vol-2, Issue-6, 2016 ISSN: 2454-1362,

GitHub Link:https://github.com/Piyush152/Object-Detection