



Taxonomy of Object Detection methods: A Survey

MIFTAH Rajaa¹, HANOUNE Mostafa², BENTAIB Mohssine³

¹Laboratory of Information Technology and Modeling, Hassan II University, Faculty of science Ben M'sik, Casablanca, Morocco, miftah.rajaa@gmail.com

²Laboratory of Information Technology and Modeling, Hassan II University, Faculty of science Ben M'sik, Casablanca, Morocco, mhanoune@gmail.com

³Laboratory of Information Technology and Modeling, Hassan II University, Faculty of science Ben M'sik, Casablanca, Morocco, m.bentaib@gmail.com

Received Date December 05, 2021 Accepted Date : December 30, 2021 Published Date : January 07, 2022

ABSTRACT

Object detection (OD) within a video is one of the relevant and critical research areas in the computer vision field. Due to the widespread of Artificial Intelligence, the basic principle in real life nowadays and its exponential growth predicted in the epochs to come, it will transmute the public. Object Detection has been extensively implemented in several areas, including human-machine Interaction, autonomous vehicles, security with video surveillance, and various fields that will be mentioned further. However, this augmentation of OD tackles different challenges such as occlusion, illumination variation, object motion, without ignoring the real-time aspect that can be quite problematic. This paper also includes some methods of application to take into account these issues. These techniques are divided into five subcategories: Point Detection, segmentation, supervised classifier, optical flow, a background modeling.

This survey decorticates various methods and techniques used in object detection, as well as application domains and the problems faced. Our study discusses the cruciality of deep learning algorithms and their efficiency on future improvement in object detection topics within video sequences.

Keywords: Object detection, Artificial Intelligence, Deep Learning, Machine Learning, Convolutional Neural Networks

1. INTRODUCTION

Deep learning, which is a subfield of machine learning that leverages neural networks and big data, inspired by the structure and function of the brain to deal with complex algorithms, then being able to make observations about interesting patterns in the data. It has made major inroads and attracted much research in recent years, due to its wide

application to various and complex tasks in computer vision, such as semantic segmentation [1], object detection [3], fraud detection, human pose estimation [1]-[5], Healthcare, etc.(Figure 1).

Object detection is one of the most fundamental and longstanding computer vision problems, which includes image processing and video analysis to detect objects in images and videos. It usually consists of different subtasks such as face detection [2], pedestrian detection [3], and skeleton detection [4].

Particularly, significant attention has been directed to object detection in images, because of the record-beating performance of deep networks in the field. However, video object detection draws less attention, due to the heavy computational complexity required. Meanwhile, the complexity of this task prompted researchers to focus on two important research topics, which are the detection or the extraction of visual objects and their categorization.

In this paper we restrict our attention to the task set for the detection of objects by presenting a detailed description of its wide-spreading applications, moreover, the challenges faced, and the techniques used to tackle these obstacles.

2. OBJECT DETECTION APPLICATIONS

Object detection within deep learning has become a popular topic nowadays, which prompts researchers to focus on its ubiquitous use cases ranging from personal to productive areas, including security, surveillance, autonomous systems, image retrieval, and machine inspection. In this section we will be introducing the application of the object detection task in some narrow domains of computer vision :

- Human-computer interaction: New challenges are posed due to the rise of computer vision, especially deep learning which opens up opportunities in this multidisciplinary field, by bridging the gap between humans and machines. Thus, we

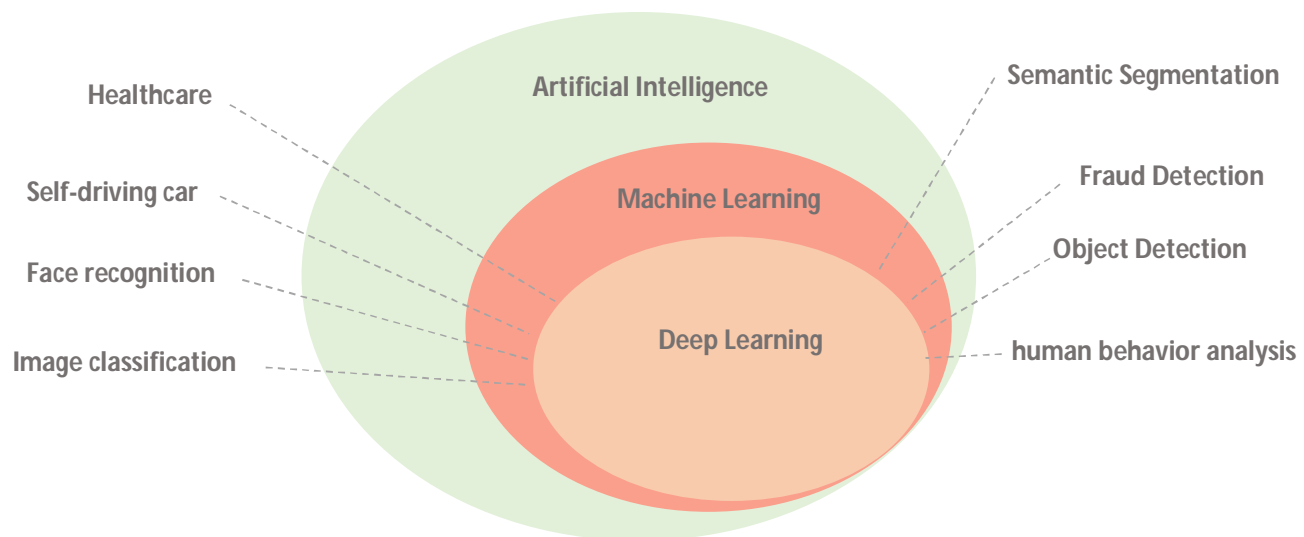


Figure 1: Object detection applications

gradually moved from mainframe computers to PCs, laptops, smartphones, and beyond. There is now practically a computer in everything, from coffee makers to vehicles, airplane engines, and the human body.[6], it focuses on the interaction between machines and humans in many ways.

- **Healthcare:** deep learning and object detection gain much importance in the healthcare industry, due to their great potential in analyzing complex images and patterns determination. Images are identified, evaluated, and interpreted by deep learning algorithms, which helps in detecting anomalies and patterns and obtaining a diagnosis.

- **Self-driving car:** Object detection methods based on deep learning are giving the best performances in identifying objects on the road, such as vehicles, pedestrians, traffic lights, etc[5]. Deep learning models identify the distance between the car and the moving objects to create alerts, and guide the self-driving car, the main idea is to detect objects and capture images from different angles, then send them to the image processing unit. Based on deep learning algorithms the model will be trained to give the car instructions or alerts as an output.

- **Image retrieval:** an image retrieval system is one of the significant research areas which can be used for browsing, searching, and retrieving images from a large database of digital images[2]. This system indexes and retrieves images from the database of collected images, based on matching the similarity between the query image and the digital images stocked on the database.

- **Surveillance systems:** Computer-aided video surveillance has been receiving significant attention as it enables authorities to trace potential threats to public safety[3][4]. Deep learning algorithms have been triggered in this domain, as it shows overwhelming results in detecting, localizing, and reporting suspicious objects in public spaces as well as weapons, knives, and guns.

Despite the tremendous progress in the subject of object detection, this field remains significantly more primitive than human vision and still cannot yet satisfactorily address the real-world challenges

3. OBJECT DETECTION CHALLENGES

Although the deep learning technology greatly improves the accuracy of object detection, it still tackles many challenges such as :

- **Occlusion:** this is the major cause of failure in detection and tracking, heavily occluded objects are more difficult for classification algorithms to identify correctly than un-occluded objects [1]. Occlusion happens when an object partially occludes another one. The detector or the tracker gets confused and fails to detect the object robustly.
- **Deformation:** this limits the performance of many object detection algorithms, multiple objects are moving bodies, which make them subject to be deformed in many ways.
- **Illumination conditions:** this condition leads objects to the exhibition of different colors due the different lightning, which affects the capability of the detector.
- **Cluttering:** known as a texture background problem, it may blend the object into the background, which makes the detection a puzzling task.
- **Viewpoint variation:** the angle of view here may create an issue for the detectors. An object may look different whenever the angle of view changes.
- **Speed for real-time detection:** detecting and localizing significant objects is not sufficient (to judge a model accuracy) for a model to be efficient for a detection task, the speed is a major factor in determining the accuracy of a model. Object detection algorithms must be incredibly fast at

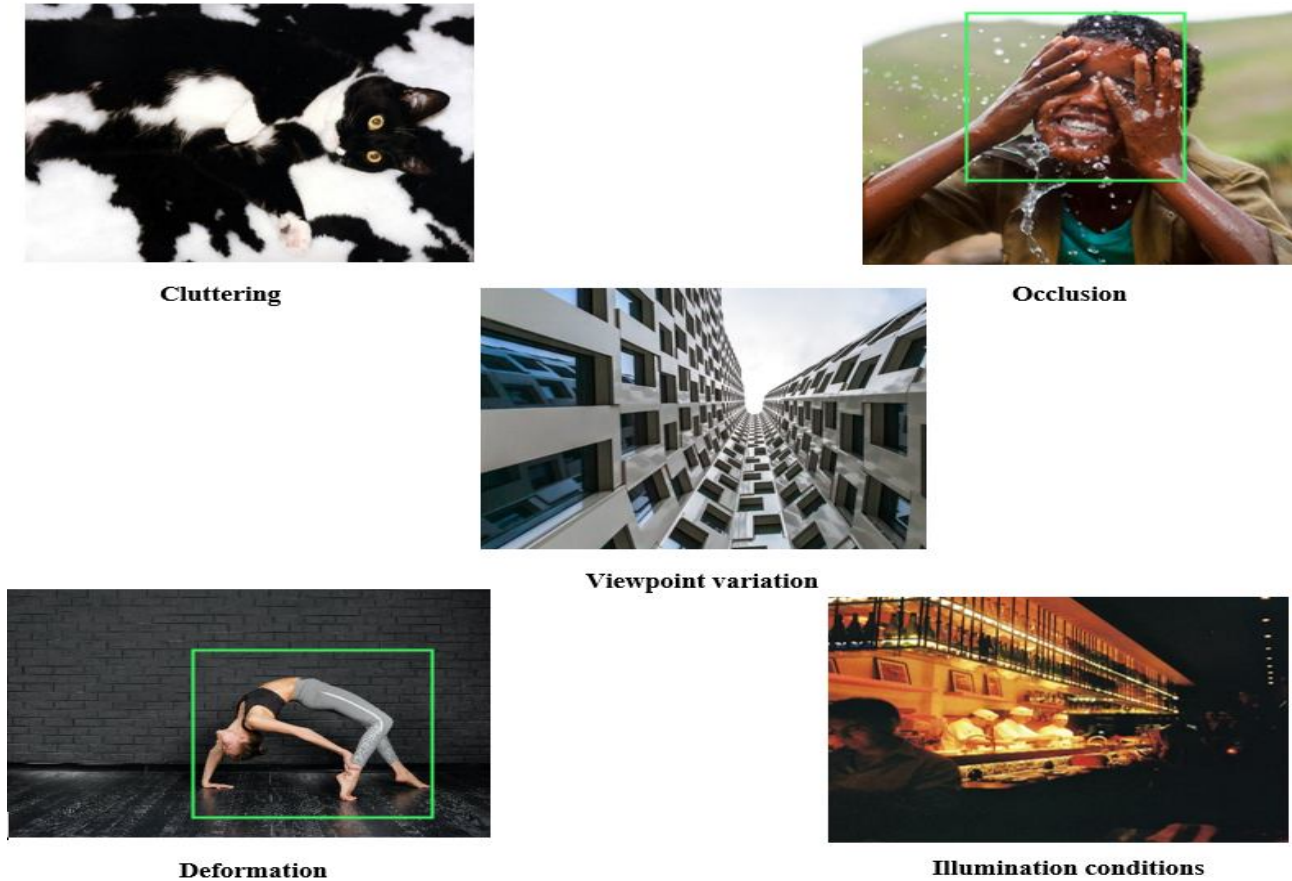


Figure 2: Object detection obstacles

prediction time to meet the real-time demands of video processing [7] (Figure 2)

4. OBJECT DETECTION TECHNIQUES

In this section, we will introduce the task setting for object detection from video and give a detailed description of the various techniques building a branched functional-structural schema of object detection. The schema describing the methodologies of video object detection system is shown in Figure 4

There are several typical methods such as Point Detector, Supervised Classifier, Background modelling, Segmentation

and Optical flow, as presented in the schema of Figure 4.

In this part, we aim to assess the techniques setting for object detection and give a detailed description of each one of them.

- **Point Detection**

- **Moravec’s Detector:** The main idea of this detector is to consider the neighborhood of a pixel and determine the average changes in intensity in the considered neighborhood when the pixel moves in various directions.

- **Harri’s Detector:** in 1988, Chris Harris and Mike Stephens introduced An algorithm dedicated to extract image corners and infer features. Compared



Figure 3: Generic Object Categories

to the Moravec’s Detector, Harris’ corner detector takes the differential of the corner score into account with reference to direction directly, instead of using shifting patches for every 45-degree angles [8]

- **Scale-invariant feature Transform Detector:** The SIFT detector is an algorithm for image description, employed in order to identify and detect local features in different numerical images. It was published in 1999 by David Lowe. The SIFT descriptor is invariant to translations, rotations, and scaling transformations in the image domain and robust to moderate perspective transformations and illumination variations [5].

Chervonenkis in 1963, classified as a supervised learning algorithm, this linear model can be used for both classification and regression issues, it draws much importance in ML because it covers many applications in real life such as : face detection, handwriting recognition, image classification, Bioinformatics, etc.

- **NN-based detector:** The neural network forms the basis of deep learning, where the algorithms are inspired by the function of the human brain, NN takes in data train themselves to recognize the patterns for this input data and then predict the output for a new set of similar data. Thus, it plays a

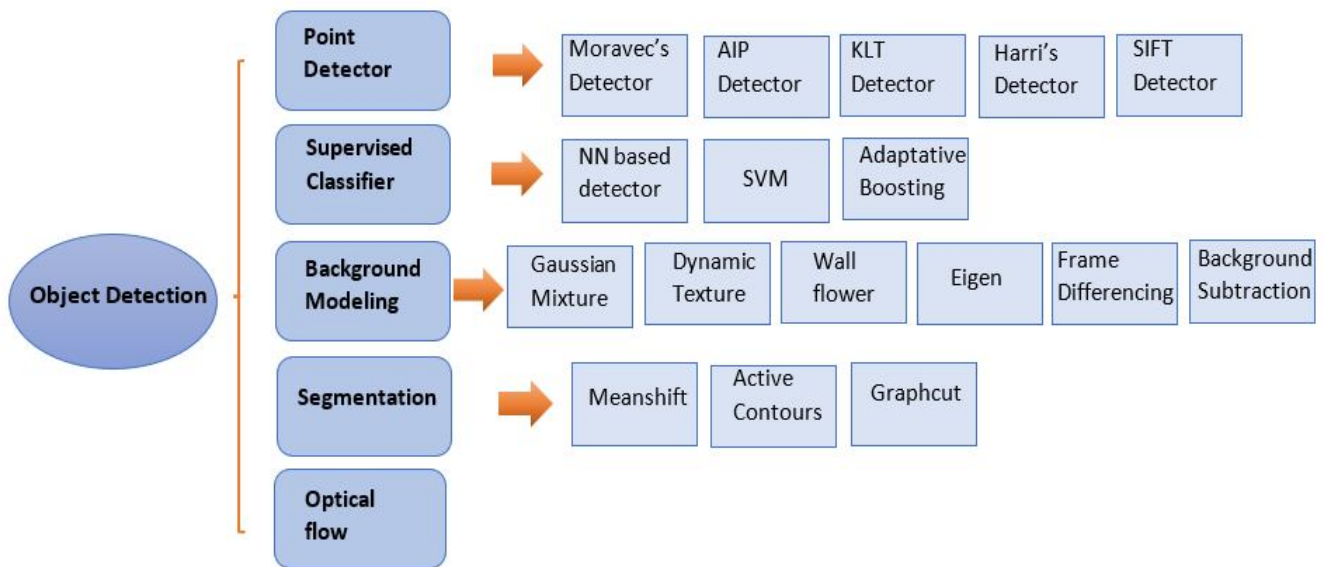


Figure 4: Object detection methodologies

- **Affine Invariant Point Detector:** This method is applied for corner detection; it generalizes the scale-invariant detector’s application due to the notable scaling differences between directions. The non-uniform scaling has an influence on the localization, the scale, and the shape of a local structure. Therefore, the scale-invariant detectors fail in the case of significant affine transformations.[9]
- **KLT Detector:** Kanade–Lucas–Tomasi feature tracker, is a feature extraction approach that relies on the detection and the tracking of point features as the main objective, besides dealing with the image registration problem efficiently and faster than traditional techniques that cost much.

• **Supervised Classifier**

- **SVM:** Support vector machine is an ML method invented by Vladimir N. Vapnik and Alexey Ya.

major role in this respect. In detection, recognition, or tracking.

- **Ada BOOSTING:** Short for Adaptive Boosting, as its name indicates, it is the most boosting technique to be adapted in solving machine learning practices.

The main idea here is to create sequential algorithms, with the objective that each model minimizes the error of the previous one.

• **Background modeling**

The goal of background modeling is to build and maintain a representation of the background so as to carry out further tasks like detecting foreground objects. Among various approaches so far of this technique, we find the following important approaches:

- **Gaussian Mixture models (GMM):** are very useful clustering models, expressed based on a

mixture density. GMMs are commonly used as a parametric model of the probability distribution of continuous measurements or features in a biometric system, such as vocal-tract-related spectral features in a speaker recognition system[10].

- **Eigen:** this method is strongly relied upon in detecting moving objects, visual surveillance. The rationale in this approach is that of generating events and tracking their influences. The Eigen plays a critical role in background modeling.
- **Wall flower:** The perfect system does not exist. However, we could improve the accuracy of a system through the application of adequate algorithms. The Wallflower algorithm is dedicated for background maintenance and attempts to address some of the detection challenges enumerated previously. Wallflower was intended to solve as many of the canonical problems as possible.[11]



Figure 5: Image captured at time t



Figure 6: Image captured at time t-1



Figure 7: Subtracted background

- **Dynamic Texture:** dynamic texture segmentation essentially refers to the extraction of regions with a similar kind of motion[12] refers to moving textures (localization / segmentation)
- **Frame differencing:** As the name suggests, it is a simple, yet effective technique applied to detect moving objects in a video. The main motive is to differentiate between two video frames based on the change of the pixels values that indicates that the objects have moved.
- **Background subtraction:** a widely used approach for detecting moving objects in videos (Vehicles, humans, etc.), it aims to detect differences by comparing captured images at time t and t1 from video sequences, then separate the background from the foreground based on those differences.(Figure 5, 6, 7)

• Segmentation

The image segmentation is a critical part of the image semantic understanding, it refers to pixels' identification. The basic idea of the method is to separates the images into several classes based on objects boundaries in image. We discuss in this part some of segmentation techniques :

- **Meanshift:** It can be considered as an advanced version of K-means, where we do not need to identify the number of cluster K in advance, another advantage of this algorithm is its independency of the shape of cluster.
- **Graphcut:** a classic technique of segmentation that consists of dividing the image into small parts and try to expect the region of interest based on graph theory, the purpose of this theory is to combine similar pixels together and promote different pixels to become apart: foreground and background.(Figure 8)

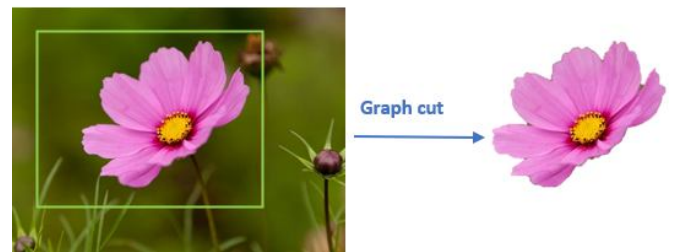


Figure 8: Graph cut segmentation algorithm

- **Optical flow:** This technique reflects the flow or the change in image, in other words it is based on the change in optical intensity or the brightness values to track the moving objects. The advantage of this method is that it can represents

a 3D motion of object points across a 2D image, as well as it could focus on the interested moving objects while ignoring motion of the unimportant ones.

- **Active Contour:** more commonly known as “Snakes model”, it is a dynamic structure (framework) used in image analysis applications such as, segmentation, parsing and medical image recognition.

The approach seeks to find contours of an object in image by solving an energy-minimizing issue. It has been introduced by Kass, Witkin, and Terzopoulos in 1987.(Figure 9)



Figure 9: Active contours model

It was a quick review about some methods and techniques of the object detection task which aim to cope several detecting challenges, and each model has its own strength and weakness in detection task.

5. DISCUSSION AND CONCLUSION

In this conclusion section, we will present and discuss some degree of the taxonomy of object detection. The goal is to provide a comprehensive survey around the subject by sitting The main axes: Applications domain of the field, the longstanding challenges, and the various techniques to overcome it. Object detection, this technology has been a research hotspot in recent years. The survey starts by presenting the field and its pipelines. Then, briefly review the domain of its application. The paper also provides a comprehensive overview of different subproblems, namely, occlusion, deformation, illumination conditions, cluttering, viewpoint variation, speed for real-time. Finally, we propose several effective methodologies to gain a thorough understanding of the object detection system. This survey is also meaningful for the developments in the mentioned techniques related to the object detection landscape, which provides valuable insights and guidelines for future progress.

REFERENCES

1. B. Chandler and E. Mingolla, “**Mitigation of Effects of Occlusion on Object Recognition with Deep Neural Networks through Low-Level Image Completion,**” *Comput. Intell. Neurosci.*, vol. 2016, 2016, doi: 10.1155/2016/6425257.
2. K. Lalitha and S. Murugavalli, “**A survey on image retrieval techniques,**” *Adv. Parallel Comput.*, vol. 37, no. 1, pp. 396–400, 2020, doi: 10.3233/APC200174.
3. M. Kmiec and A. Glowacz, “**Object detection in security applications using dominant edge directions,**” *Pattern Recognit. Lett.*, vol. 52, pp. 72–79, 2015, doi: 10.1016/j.patrec.2014.09.018.
4. Y. Nam, S. Rho, and J. H. Park, “**Intelligent video surveillance system: 3-tier context-aware surveillance system with metadata,**” *Multimed. Tools Appl.*, vol. 57, no. 2, pp. 315–334, 2012, doi: 10.1007/s11042-010-0677-x.
5. T. Lindeberg, “**Scale Invariant Feature Transform,**” *Scholarpedia*, vol. 7, no. 5, p. 10491, 2012, doi: 10.4249/scholarpedia.10491.
6. P. PATHAK. *Human-computer interaction: An AI revolution*, May, 2020
7. D. Tyagi. *Introduction to Harris Corner Detector*, Mar. 2019.
8. K. Fessel. *Reviewing unique challenges presented by object detection problems and researchers’ efforts to overcome these obstacles*, Sep. 2019.
9. K. Mikolajczyk and C. Schmid, “**Scale & Affine Invariant Interest Point Detectors,**” vol. 60, no. 1, pp. 63–86, 2004.
10. D. Reynolds, “**Gaussian Mixture Models * Main Body Text,**” no. 2, pp. 1–5.
11. B. M. Kentaro Toyama, John Krumm, Barry Brumitt, “**Wallflower Principles and Practice of Background Maintenance,**” vol. 1, p. 29, 1999, [Online]. Available: <http://www.mm.media.kyoto-u.ac.jp/research/thesis/2008/m/mori/mori.pdf>.
12. V. Kaltsa, K. Avgerinakis, A. Briassouli, I. Kompatsiaris, and M. G. Strintzis, “**Dynamic texture recognition and localization in machine vision for outdoor environments,**” *Comput. Ind.*, vol. 98, pp. 1–13, 2018, doi: 10.1016/j.compind.2018.02.007