e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:05/Conference:01/March-2023 Impact Factor- 7.868 www.irjmets.com
National Conference on Trending Technology for Achieving Sustainable Development Goals
NCTTASDG 2023 Organized by Shri Shankarprasad Agnihotri College of Engineering, Wardha

OBJECT DETECTION IN REAL TIME A COMPARATIVE STUDY OF TRADITIONAL AND DEEP LEARNING APPROACHES

Prof. M. A. Ramteke[*1], Rahul Jani[*2], Rushikesh Yerawar[*3], Shalini Ambade[*4], Saurabh Dhumne[*5], Rupali Dhakite[*6], Prof Vaishali V. Jikar[*7], Prof A. R. Ghongade[*8]

[*1,7,8]Prof, Department of Computer Science &Engineering, SSPACE, Ramnagar, Wardha, Maharashtra, India.

[*2,3,4,5,6]Student, Department of Computer Science & Engineering, SPACE, Wardha, Maharashtra, India.

ABSTRACT

Object recognition systems aim to identify objects in the real world from an image by leveraging pre-existing object models. Despite its ubiquity in humans, object recognition is a difficult task to implement algorithmically. In this chapter, we will outline the various steps involved in object recognition and discuss techniques that have been employed across a range of applications. We will explore the different types of recognition tasks that vision systems may need to perform, assessing the complexity of each and presenting approaches that are useful at different phases of the recognition process. At its core, the object recognition problem involves labelling regions in an image as belonging to one or more known objects, based on a set of models that the system has access to. This task is closely related to segmentation, as without at least some degree of object recognition, it is impossible to accurately segment an image, and conversely, without effective segmentation, object recognition becomes much more challenging. we will discuss the various steps involved in object recognition and explore the techniques used in a variety of applications. Throughout the article, we will present approaches that are useful at different stages of the recognition process.

Keywords: CNN, Dataset, YOLO, object detection, tracking, surveillance.

I. INTRODUCTION

Surveillance cameras are ubiquitous in modern society and are used for more than just security. They can help identify areas of interest, aid in completing tasks, and play a crucial role in machine vision for target detection, recognition, positioning, tracking, and navigation. Object detection systems currently repurpose classifiers to identify objects by evaluating them at various locations and scales in an image. This method, such as deformable parts models (DPM), employs a sliding window approach where the classifier runs at evenly spaced locations over the entire image. While many researchers have proposed methods for detecting humans in video images, there is a need for real-time and accurate detection, positioning, and motion analysis of human bodies in various scenarios. However, such video detection and recognition often encounter various problems in complex natural conditions, such as differences in lighting, environment, and shooting angle, as well as gaps in semantic understanding, computational complexity, and adaptability. Furthermore, in many cases, motion recognition is necessary, requiring the detection and analysis of the detected people's motion.Object detection is an essential task in computer vision, and various techniques have been proposed, including feature-based and template-based approaches, as well as background subtraction. However, selecting the best technique for a specific application depends on the available hardware resources and the scope of the application. Feature-based detection searches for corresponding features in successive frames, such as Harris corner, edges, SIFT, contours, or colour pixels. Background subtraction is a popular method that uses a static background and calculates the difference between the hypothesized background and the current image. This approach is fast and suitable for fixed backgrounds, but it cannot handle dynamic environments with differentillumination and motions of small objects. Tracking aims to establish a correspondence between the detected target object of images over frames. Tracking using mean shift kernel is also introduced, which performs well when there is occlusion, which can be solved using templates. Camshift (Continuously Adaptive Meanshift) can track a single object fast and robust using colour features, but it is ineffective for occlusion. Appearance-based object detection is also a research area that uses whole 2-D images to perform tracking for navigation in faster time. However, this approach requires several templates and does not work when the target object, colour, or perspective view is changed. The main challenge in object detection and tracking is the temporal variation of

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:05/Conference:01/March-2023 Impact Factor- 7.868 www.irjmets.com
National Conference on Trending Technology for Achieving Sustainable Development Goals
NCTTASDG 2023 Organized by Shri Shankarprasad Agnihotri College of Engineering, Wardha

objects due to perspective, occlusion, interaction between objects, and appearance or disappearance of objects. This causes the appearance of a target to change during long tracking. The background in a long image sequence is also dynamic, even if it is taken by a stationary camera. Detecting and tracking multiple objects simultaneously is an important issue for real-time performance. Comprehensive search in multiple tracking is computationally expensive and incapable of being a realtime system. Another issue arises when using a moving camera instead of a fixed location camera, which requires the analysis of the camera platform coordinate system. In this review, we analyse four different real-time object detection and tracking techniques in terms of accuracy, computational time, and memory consumption, and propose the best technique for real-time implementation in mobile robots. These techniques are:

(1) Object tracking by image differencing

(2) Object tracking by using local transformations

(3) Object tracking by using morphological-based object detection

(4) YOLO

The development of hardware technology also affects the real-time performance of object detection and tracking. In real-world object tracking systems, the system must be robust to handle changing environments with real-time constraints and limited processing resources and memory. Thus, handling complex tracking using only software solutions is not flexible as it is limited by the processing capability under real-time constraints. A real-time application requires that the tracking system must be fast enough, power-efficient, and have managed memory to meet hard real-time constraints. Therefore, in thispaper, wepresent a detailed study of each algorithm with the steps involved for implementation in Section I. Section II provides a brief comparison of each algorithm and proposes the most suitable algorithm for real-time object detection and tracking. Finally, in Section III, we summarize the important aspects of the studied algorithms and provide future recommendations.

II. TRADITIONALMETHODS OF OBJECT DETECTION

Various methods for object detection and tracking have been proposed in the literature.

1) One algorithm discussed by A.J. Lipton, H. Fujiyoshi and R.S. Patil is based on frame differencing. While it is effective for objects that remain consistent in size and colour, the error between frames increases exponentially throughout the video sequence.

2) Stein, Rosenberg, and Werman proposed the use of non-parametric local transforms for object tracking. Ramin Zabih and John Woodfill applied this idea and developed a transform called the Census Transform. Their approach works well in noisy and variable lighting environments, but it is computationally expensive.

3) Owensa, Hunterb, and Eric developed an algorithm for tracking moving objects based on morphological characteristics. This method solves the problem of object merging when tracking multiple objects. However, recognition through morphological methods can be complex and must be repeated continuously.

4) Carlo Tomasi and Takeo Kanade proposed a simple object tracking method that minimizes the sum of squared intensities between consecutive frames. This method is fast and robust and is recommended for real-time object tracking.

2.1 Census Transform Method- The Census Transform Method is an advanced video sensing technique that offers a better approach to object tracking. This technique is based on non-parametric local transforms, which rely on the local order of intensity values surrounding a central pixel, rather than on the actual intensity values of the pixels. To use this method for object tracking, we first apply a local transform to consecutive frames, and then compute the correspondence of similar pixels between the two frames using correlation. To achieve this, we form consecutive difference images by subtracting consecutive frames of a video, and then replace each pixel's local surrounding with either bit 1 if it is greater than the central pixel or 0 if it is not. This process generates a bit-string, called a signature vector, for each pixel. We then create separate lists for each image, containing signature vectors for all pixels, along with their corresponding pixel positions.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:05/Conference:01/March-2023 Impact Factor- 7.868 www.irjmets.com
National Conference on Trending Technology for Achieving Sustainable Development Goals
NCTTASDG 2023 Organized by Shri Shankarprasad Agnihotri College of Engineering, Wardha

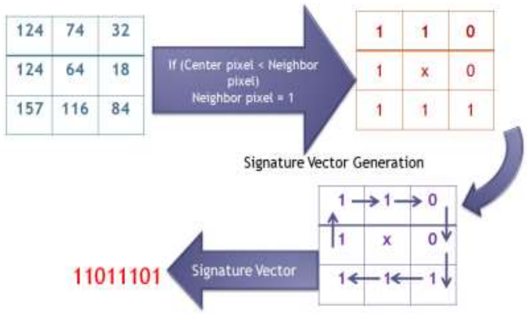

Fig-1:Census Method: Signature Vector Generation

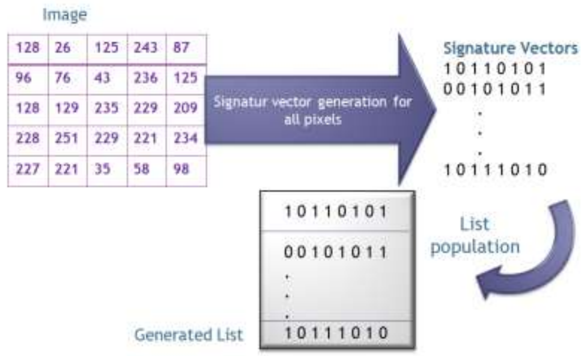

Fig-2: Census Method Steps

This algorithm is robust to noise and illumination changes because the value of any pixel depends on the values of its surrounding pixels, and changes in one pixel do not affect the output significantly. Moreover, this method is faster and more accurate than previous algorithms. However, it is not suitable for hardware implementation in mobile robots. For hardware implementation, further improvements to these parameters are required

2.2 Absolute Difference-Jaewon Shin proposed the use of the absolute difference method for motion detection, which involves comparing two image frames and computing the absolute difference between the gray level intensities of neighbouring pixels.as shown in the Figure 1, 2. The mathematicalrepresentation of the absolute differences is given by the equation:

$D(t) = |I(ti) - I(tj)|$ where I(ti) and I(tj) denote the images at times i and j, respectively, and D(t) represents the absolute difference at that specific time instance. When there is no motion, the two images will be the same, and thus, the absolute difference will be zero.

Fig-1: a) Input image with objects. b) Background model

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:05/Conference:01/March-2023 Impact Factor- 7.868 www.irjmets.com
National Conference on Trending Technology for Achieving Sustainable Development Goals
NCTTASDG 2023 Organized by Shri Shankarprasad Agnihotri College of Engineering, Wardha

Fig-2: Detected Objects.

There are two primary methods for motion detection using absolute differences, namely

1) frame difference
2) background subtraction.

Although these techniques are relatively straightforward to implement and produce satisfactory results, they suffer from several limitations. Firstly, both absolute differencing methods involve numerous computations to calculate the grey level pixel intensities differences, resulting in a large instruction set that is stored in the hardware memory, thereby making the algorithm time-consuming. Secondly, the large instruction set required for absolute differencing makes it unsuitable for implementation on a digital signal processor (DSP). Thus, this method is not feasible for use in DSP based motion detection applications.

2.3 YOLO- Object detection is a fundamental task in computer vision and has seen significant advancements in recent years. Traditional methods involve generating potential bounding boxes in an image using region proposal methods, followed by classification and post-processing to refine the bounding boxes and eliminate duplicate detections. However, these complex pipelines are slow and hard to optimize, as each component must be trained separately. In contrast, YOLO (You Only Look Once) reframes object detection as a single regression problem, from image pixels to bounding box coordinates and class probabilities.

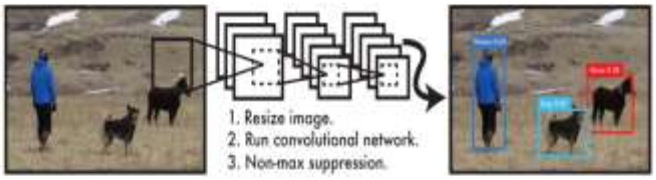

Figure 1: The YOLO Detection System. Processing images with YOLO is simple and straightforward. Our system (1) resizes the input image to 448 × 448, (2) runs a single convolutional network on the image, and (3) thresholds the resulting detections by the model's confidence. The YOLO model consists of a single convolutional network that simultaneously predicts multiple bounding boxes and class probabilities. Unlike traditional methods, YOLO reasons globally about the image during training and testing, which allows it to encode contextual information about classes as well as their appearance. Additionally, YOLO learns generalizable representations of objects, making it less likely to break down when applied to new domains or unexpected inputs. One of the main benefits of YOLO is its speed. Since it frames detection as a regression problem, it does not require a complex pipeline, and the base network runs at 45 frames per second with no batch processing on a Titan X GPU. A faster version runs at more than 150 fps, making it possible to process streaming video in real-time with less than 25 milliseconds of latency. Another advantage of YOLO is its superior performance in terms of mean average precision (mAP) compared to other real-time systems. YOLO achieves more than twice the mAP of other systems and makes less than half the number of background errors compared to Fast R-CNN, a top detection method. In summary, YOLO is a highly efficient and accurate object detection system that learns generalizable representations of objects and reasons globally about the image during prediction. These features make it an ideal choice for real-time object detection in various domains.

2.4 Kanade-Lucas Technique- The Kanade-Lucas-Tomasi (KLT) feature tracker is a popular method for object tracking in computer vision. It was first introduced by Tomasi and Kanade in 1991 and later refined by Lucas and Kanade in 1994. The technique is based on finding sparse correspondences between image frames by tracking a set of key feature points. The KLT method has been widely used in various applications such as surveillance, autonomous vehicles, and augmented reality. The algorithm works by first detecting feature points in an image frame using a corner detector such as the Harris corner detector. Next, the algorithm tracks the movement of these feature points by finding their correspondences in the subsequent frames. One of the advantages of the KLT method is its efficiency in tracking objects in real-time. The algorithm can track objects even in the presence of significant changes in illumination, scale, and rotation. The KLT method also provides high accuracy in object tracking and can handle occlusion and object deformation. However, the KLT method has some limitations. Firstly, it relies on finding sparse correspondences between frames, which can be challenging in situations where there are few distinct feature points. Secondly, the algorithm can be sensitive to changes in lighting conditions, which can affect the accuracy of the feature detection process. To address these limitations, researchers have proposed several modifications to the KLT method. One such modification is the use of dense optical flow methods to estimate the motion of all pixels in the image, rather than just a sparse set of feature points. Another modification is the use of deep learning techniques to improve the accuracy and robustness of feature detection. In conclusion, the KLT method is a widely used technique for object tracking in computer vision due to its efficiency, accuracy, and ability to handle various challenges. While the method has some limitations, ongoing research is focused on addressing these limitations and improving the performance of the algorithm.

III. APPLICATIONS

Real-time object detection and tracking have a diverse range of applications, including the detection and tracking of cars passing on a highway. An image processing system counts the tracked cars to provide traffic information. Another significant and popular application is for security surveillance. A real-time surveillance system can detect and track suspicious movements by individuals. The industries in this field are experiencing rapid development to provide better performance. Object tracking is also a crucial issue in mobile robots that use vision-based systems. The robot can track objects and use feature information to build a map for localization. In complex environments, a combination of traditional and modern methods is applied, such as the combination of background subtraction with KLT/meanshift. These advanced techniques allow for more accurate tracking and detection of objects in challenging environments.

IV. CONCLUSSION

In this literature review, we have examined the performance of five methods for real-time object detection and tracking in terms of their accuracy, computational time, and memory consumption. The effectiveness of a video tracking algorithm is dependent on its response quality to high frame-rate input videos. With a higher frame-rate, the accuracy of the algorithm decreases, posing a challenge for its performance. Based on the simulation results, we propose the utilization of the Kanade-Lucas algorithm for real-time object detection and tracking in mobile robots. This algorithm is the most efficient in terms of speed and memory usage, with no implementation complexities. It performs well in scenarios with high distortion and provides exceptional support for video sequences with high frame rates due to its iterative nature. However, the Kanade-Lucas algorithm imposes an additional requirement that the image intensities be constant between consecutive frames. Additionally, the YOLO (You Only Look Once) algorithm is another popular and effective method for real-time object detection and tracking. YOLO has demonstrated promising results in a variety of challenging scenarios, detecting multiple objects in a single frame, making it a popular choice in object detection and tracking applications. Evaluating the fundamental nature of each algorithm is a challenging task, as some algorithms perform well in specific conditions, but not in others. As a result, a combination of algorithms, such as YOLO and Kanade-Lucas, can be utilized to enhance the overall performance of the system in complex scenarios with varying conditions. In future research, hybrid algorithms can be developed that integrate the strengths of multiple algorithms to achieve better performance and accuracy in real-time object detection and tracking.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:05/Conference:01/March-2023 Impact Factor- 7.868 www.irjmets.com
National Conference on Trending Technology for Achieving Sustainable Development Goals
NCTTASDG 2023 Organized by Shri Shankarprasad Agnihotri College of Engineering, Wardha

TABLE : Result Comparison of Different Methods

Technique	**Strength**	**Weaknesses**
Absolute Differences Technique	1)Fast and simple 2)Effective for detecting motion in station	1)Can generate false positives due to noise and lighting changes 2) Limited to simple scenarios
Census Transform technique	1)Robust to lighting changes and noise 2) Can detect small changes in the scene	1)Can be computationally expensive 2) May miss objects with complex textures or patterns
Kanade Lucas Technique	1)Fast and accurate for tracking objects with large motion 2)Iterative nature provides support for high frame-rate videos	1)Assumes constant image intensities between consecutive frames 2)May not work well in scenarios with small object motion
YOLO (You Only Look Once)	1)Fast and efficient for real-time object detection 2)Can detect multiple objects in a single frame	1)May struggle with small object detection 2)Can generate false positives in complex scenes with many objects.

V. REFERENCES

[1] P. F. Felzenszwalb, R. B. Girshick, D. McAllester, and D. Ramanan. Object detection with discriminatively trained part-based models. IEEE Transactions on Pattern Analysis and Machine Intelligence, 32(9):1627–1645, 2010.

[2] A. J. Lipton, H. Fujiyoshi, and R. S. Patil, “Moving target classification and tracking from real-time video,” in Applications of Computer Vision, 1998. WACV’98. Proceedings., Fourth IEEE Workshop on. IEEE, 1998, pp. 8–14.

[3] W. Junqiu and Y. Yagi, “Integrating color and shapetexture features for adaptive real-time object tracking,” IEEE Trans on Image Processing, vol. 17, no. 2, pp. 235–240, 2008.

[4] J. Owens, A. Hunter, E. Fletcher et al., “A fast model-free morphologybased object tracking algorithm,” 2002.

[5] C. Tomasi and T. Kanade, “Shape and motion from image streams under orthography: a factorization method,” International Journal of Computer Vision, vol. 9, no. 2, pp. 137–154, 1992.

[6] J. Shin, “Initialization of visual object tracker using frame absolute difference,” 2000.

[7] Q. Wang and Z. Gao, “Study on a Real-Time Image Object Tracking System,” in Computer Science and Computational Technology, 2008. ISCSCT’08. International Symposium on, vol. 2, 2008.

[8] J. Shin, “Initialization of visual object tracker using frame absolute difference,” 2000.

[9] Q. Wang and Z. Gao, “Study on a Real-Time Image Object Tracking System,” in Computer Science and Computational Technology, 2008. ISCSCT’08. International Symposium on, vol. 2, 2008.

[10] R. Girshick, J. Donahue, T. Darrell, and J. Malik. Rich feature hierarchies for accurate object detection and semantic segmentation. In Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on, pages 580–587. IEEE, 2014.

[11] R. B. Girshick. Fast R-CNN. CoRR, abs/1504.08083, 2015.

[12] P. F. Felzenszwalb, R. B. Girshick, D. McAllester, and D. Ramanan. Object detection with discriminatively trained part-based models. IEEE Transactions on Pattern Analysis and Machine Intelligence, 32(9):1627–1645, 2010.

[13] Redmon, Joseph, et al. "You Only Look Once: Unified, Real-Time Object Detection." IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 779-788.