International Journal of Innovative Research in Engineering
Volume 5, Issue 3 (May-June 2024), PP: 180-185.
www.theijire.com **ISSN No: 2582-8746**

Exploring Image Processing Based Object Detection & Tracking Techniques in Videos: A Concise Overview

Priya Jain[1], Meenakshi Arora[2], Rohini Sharma[3]
[1]PG. Student, Department of CSE, Sat Kabir Institute of Technology and Management, Haryana, India.
[2]Assistant Professor, Department of CSE, Sat Kabir Institute of Technology and Management, Haryana, India.
[3]Assistant Professor, Government P.G. College for Women, Rohtak, Haryana, India.

How to cite this paper:
Priya Jain[1], Meenakshi Arora[2], Rohini Sharma[3] "Exploring Image Processing Based Object Detection & Tracking Techniques in Videos: A Concise Overview", IJIRE-V5I03-180-185.

Abstract: *In computer vision, object tracking in videos is an essential problem with implications ranging from autonomous driving to espionage. The methods, benefits, and drawbacks of object tracking approaches based on image processing are thoroughly examined in this work. A variety of traditional and modern techniques are covered in the paper, such as feature-enabled, model-enabled, contour-enabled, point-enabled, kernel-enabled, and deep learning-enabled tracking. The effectiveness of point tracking approaches in simulating linear and non-linear motion, including the particle filter and Kalman filter, is explored. The resilience of kernel-based tracking, especially the mean-shift technique, in real-time applications is analysed. The accuracy of contour-based tracking techniques, such as active contours, in detecting boundaries is examined. The paper also delves into model-based tracking, which utilizes 3D models to enhance tracking accuracy, and hybrid methods that combine the strengths of multiple approaches. Additionally, the advent of deep learning has introduced powerful methods which have significantly improved tracking performance through advanced feature representation and temporal dependency modelling. This work provides a concise yet comprehensive overview of the paper's content, outlining the various object tracking techniques covered and emphasizing their significance and applications.*

Keywords: *Object Tracking, Feature Matching, Morphological Operation, Optical Flow*

I.INTRODUCTION

One of the most crucial tasks of devices that communicate with the workings of the real world is object tracking via computer vision. Examples of these devices include independent ground vehicles [1], autonomous aerial drones [2], robotics, and missile tracking systems. It is necessary to keep an eye out for changes in order for machines to function and adjust to the dynamics of the real world. These alterations are often the motions that need to be detected by various sensors, after which the machines must react in accordance with these alterations. Computer vision simulates how a human might see these changes. Because they have multiple senses, humans are able to perceive changes in their surroundings intuitively, which aids in their world navigation. One of the main senses that people use to navigate their surroundings is vision. Computer vision can aid in increasing productivity by helping to build autonomous machines that carry out human duties including driving, fishing, agricultural work, and medical diagnosis. Humans now have greater tools for accomplishing jobs quickly and making more informed decisions because to the application of CV in robotics, medical diagnosis, and human–computer interaction. Therefore, in order to increase productivity and create a self-organizing system that collaborates effectively with people, it is crucial to research various approaches, instruments, and possible applications in order to assess their shortcomings and future possibilities for solving object tracking issues in computer vision.

II.RESEARCH BACKGROUND

The field of computer vision-based object tracking has advanced significantly. Prior reviews and surveys concentrate on a specific aspect of the object tracking challenge. A review with a narrow focus on a particular aspect of the research field is frequently helpful in identifying specific gaps in the literature. Expanding the scope of the literature review, however, aids in determining whether one strategy is superior to another. In addition, a review of the topic of study offers engineers and researchers a path forward for further investigation of the issue based on the demands of the application. Authors in [3] examined the research on depiction using their feature-construction mechanism in their study of appearance frameworks. They came to the conclusion that it was crucial to accurately characterize the 2D appearance of tracked objects for successful visual tracking since object tracking algorithms struggle to handle complicated object appearance changes caused by lighting, obstruction, shape deformations, and camera movement.

Deep learning (DL) based online multiple-object tracking was examined by [4], who also graded the networks using various publicly available datasets as benchmarks. By classifying the current techniques into several DL methods, authors in [5] offered an empirical assessment of the cutting-edge deep learning methods for change detection. Moreover, they provide an empirical examination of the assessment configurations used by current deep learning techniques. Deep learning techniques for multiple-object tracking in autonomous driving were examined by Guo et al. [6]. Based on transformer-based tracking, joint detection and tracking, and tracking by detection, their assessment categorized the algorithms. They found datasets for multiple-object tracking, offered an experimental analysis, and suggested directions for future deep learning research. An outline of the paper is provided by Figure 1:

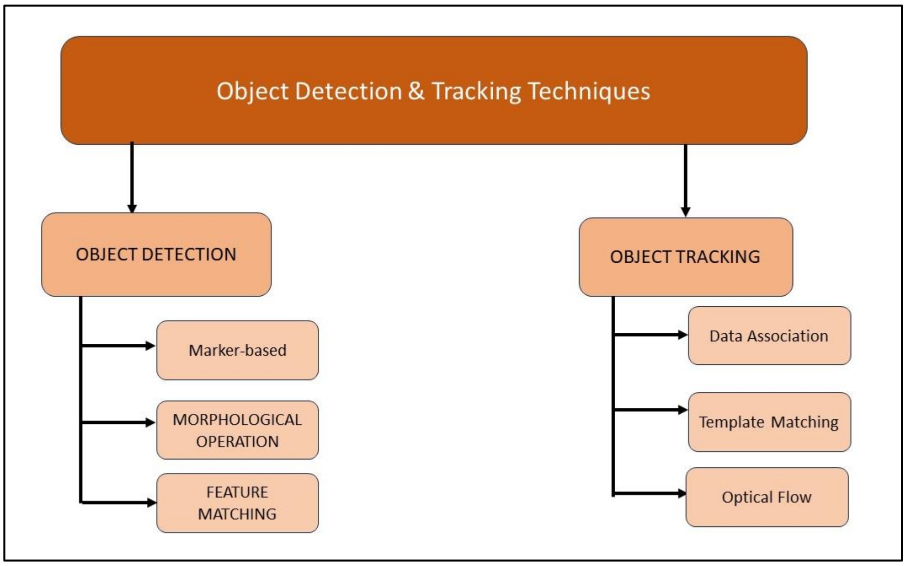

Figure 1: Classification of Object Detection & Tracking in a Video

III.CLASSIFICATION OF VIDEO OBJECT DETECTION METHODS

Detection and Localization Methods

The majority of tracking issues begin with the object's detection and localization. Image processing has long been used in various research investigations as a method for feature detection and tracking. However, because of their increased accuracy and ability to localize and classify objects using end-to-end networks, deep learning techniques are gaining popularity.

feature matching

The process of recognizing features in an image and comparing them with equivalent characteristics on other images is known as image matching.

Reference	Technique	Key Features	Advantages	Disadvantages
[7]	Stereo Visual Odometry	employ feature matching and a stereo camera system to estimate depth	Precise depth determination, appropriate for dynamic settings	heavy computing and prone to noise in water effects
[8]	Dense 3D Reconstruction	Deep 3D modeling in real time using stereo pictures	High precision, real-time processing	considerable processing overhead, potentially problematic in areas lacking texture
[9]	Kalman Filter	Linear filtering and object state estimation	Minimal computational cost, effective in linear, Gaussian noise situations	restricted to models of linear motion; less successful with non-Gaussian noise
[10]	Dense RGB-D Visual Odometry	integrates depth and RGB data, rich matching of features	dependable in a range of lighting situations and has excellent 3D motion tracking precision	High computational cost, needs depth sensor
[11]	Fast Compressive Tracking with Weighted Multi-frame TM	Multiple frame template matching and compressive detection for choosing features	Fast processing, operative for fast-moving objects	May have trouble with occlusions and visual alterations; original template precision may be a limitation.

Table 1: Comparison of feature matching-based object detection in videos

Morphological operation

A group of image processing techniques known as morphological operations apply a structuring element that modifies the features in the image's shape. Erosion, in which an item loses size, and dilatation, in which an object gains size, are two typical morphological operations. Tracking by detection is a generic approach to solving object tracking issues. The goal of tracking via detection is to identify an event in each picture frame in a sequence of videos.

Reference	Technique	Key Features	Advantages	Disadvantages
[12]	Morphological operations, Stereo Vision	mixes stereo vision with morphological procedures to track fish in low-contrast, low-frame-rate films.	Robust against fish movement and obstructions, efficient in low-contrast and low-frame-rate situations	CPU-intensive because of the processing of stereo vision
[13]	Morphological operations, Binocular Camera	employs binocular cameras and morphological methods to recognize 3D characters in applications in medicine.	High recognition accuracy for 3D characters, resistant to noise and variations in lighting	Necessitates precise calibration of binocular cameras, high computational cost
[14]	Canny Edge Detection	Employing minimal suppression and a gradient-based technique, edge detection	Extremely accurate edge detection that works well in a range of lighting situations	sensitive to noise, expensive to compute in edge linking and gradient computing
[15]	SMDWT (Stationary Wavelet Transform), Dense Disparity-Variance	integrates the dense disparity-variance approach with the wavelet transform to detect and track movement.	Efficient at following and identifying motion in a variety of lighting conditions, brightness invariant	intense differential computation and the wavelet transform make it extremely difficult.

Table 2: Comparison of morphological operation-based object detection in videos

Marker-based

Certain detecting techniques make use of preset markers. Markers are physically recognized objects that the visual system is already familiar with. Compared to markerless detection, which combines feature acquisition and analysis of the characteristics of the target item, these markers are comparatively simpler to find.

Reference	Technique	Key Features	Advantages	Disadvantages
[16]	Marker-based tracking using Fiducial Markers	Tracks surface deformations of soft tissue using fiducial markers placed on the breast during surgery	High precision in tracking soft tissue deformations, real-time monitoring	Invasive as it requires placing markers on the tissue, potential for marker displacement
[17]	Marker-based tracking for 3D Reconstruction	use markers to accurately rebuild 3D in situations including autonomous vehicles	high 3D position estimation precision and resilience under different driving circumstances	needs a precise marker configuration and is computationally demanding for real-time applications.
[18]	Feature-based tracking with KAZE Features	employs non-linear scale space for tracking and identifying features.	Strong resistance to noise and scale, excellent feature detection and tracking efficiency	costly to compute because of non-linear scale space calculations, and maybe less precise than marker-based techniques

Table 3: Comparison of Marker-based object detection in videos

IV.CLASSIFICATION OF VIDEO OBJECT TRACKING METHODS

Data association

Data association is the act of comparing newly detected items with prior knowledge about an object's position, movement, and appearance changes while also tracking the object's motions. One of the most popular tracking techniques is data association, which is frequently adjusted to meet application requirements.

Reference	Technique	Key Features	Advantages	Disadvantages
[19]	Stereo Visual-Inertial Navigation, Data Association	integrates inertial data with stereo vision to provide reliable multi-object tracking	Superior precision in ever-changing settings, strong integration of optical and inertial data	computationally demanding; synchronisation of inertial and visual information is necessary
[12]	Stereo Vision, Data Association	employ data association and stereo vision to track fish in low-contrast, low-frame-rate videos.	Resilient against fish movement and obstructions, efficient in low-contrast and low-frame-rate situations	Highly computational because of the processing of stereo vision
[20]	Viterbi Algorithm, Data Association	uses dynamic programming to identify the best route for completing data association activities.	determines the optimal path for solving data association tasks using dynamic programming.	Substantial computational cost and complexity that rises as the sequence's length
[21]	Monocular 3D Detection, Data Association	uses object keypoints for tracking and real-time 3D identification in monocular pictures.	Real-time efficiency that works well for automated driving applications and doesn't require stereo vision	Low depth accuracy relative to stereo or LiDAR-based techniques necessitates excellent keypoint identification
[22]	Part-based Person Retrieval, Data Association	Person recovery in video sequences using enhanced pooling and part-based models	Outstanding person retrieval accuracy that is resilient to changing stances and partial occlusions	substantial training data is necessary, and convolutional neural networks make it computationally costly.

Table 4: A Comparison of Data association -based object tracking in videos

Template matching

Template matching is the procedure of examining the target image to find small areas that, using cross-correlation techniques, match the features to a template image of the object.

Reference	Technique	Key Features	Advantages	Disadvantages
[11]	Fast Compressive Tracking with Multi-frame Template Matching	For quick motion tracking, compressive sensing and balanced multi-frame template matching are used.	Quick processing that works well for items moving quickly	May have trouble with occlusions and appearance modifications; preliminary template precision may be a limitation.
[23]	Matched Spatial Filters for Template Matching	implements spatial filters that match in order to match templates.	High precision in regulated settings and resilience to little changes in appearance	cognitively costly, less efficient in cluttered or dynamic situations
[24]	Online Template Matching	assesses several online template matching techniques for object monitoring	gives thorough benchmark data and makes comparing various tracking techniques easier.	Outcomes vary depending on how the assessed template matching techniques are implemented.
[25]	Munkres Assignment Algorithm	uses the Munkres (Hungarian) algorithm to solve assignment problems; it can also be used for jobs involving template matching.	ensures the best possible assignment, effective for small- to medium-sized issues	Because computational complexity rises with problem size, high-speed tracking jobs might not be a good fit for it.

Table 5: A Comparison of Template matching-based object tracking in videos

Optical flow

The study of flowing patterns in an image caused by the relative motion of the objects or the observer is known as optical flow.

Reference	Technique	Key Features	Advantages	Disadvantages
[26]	Lucas-Kanade Optical Flow	Iterative method for image registration, tracks points between frames using intensity differences	High precision for small motions, computationally efficient for real-time applications	Sensitive to large motions, requires good initialization
[27]	Scale-Space Optical Flow	Uses scale-space filtering for detecting features at multiple scales, tracks objects using optical flow	Vigorous to noise and variations in scale, effective in multi-scale object detection	Computationally expensive, requires general parameter tuning
[28]	Vision-Based Optical Flow	Combines optical flow with vision-based target detection, used for tracking objects from a quadcopter	Effective in dynamic and outdoor environments, robust to varying lighting conditions	Sensitive to swift variations in object appearance, may skirmish with occlusions

Table 5: A Comparison of Optical Flow-based object tracking in videos

V.CONCLUSION

Since each object detection and tracking technique has distinct advantages and disadvantages, it can be used in a variety of ways.

Optical Flow is highly effective in dynamic environments and real-time applications but requires good initialization and struggles with large motions. **Template Matching** is accurate and fast for controlled environments but less effective in cluttered or dynamic settings. **Data Association** methods excel in dynamic environments and provide robust integration with other data sources but can be computationally intensive. **Marker-Based** tracking offers high precision and robustness to occlusions but is invasive and impractical for some scenarios. **Morphological Operation-Based** tracking is effective in low-contrast conditions but computationally expensive and sensitive to object shape variations. **Feature Matching** provides high robustness and accuracy in detecting and tracking features but requires significant computational resources and high-quality features.

References

1. *X. Li et al., "DyStSLAM: an efficient stereo vision SLAM system in dynamic environment," Meas. Sci. Technol., vol. 34, no. 2, p. 25105, 2022.*
2. *C. A. M. Busch, K. A. Stol, and W. van der Mark, "Dynamic tree branch tracking for aerial canopy sampling using stereo vision," Comput. Electron. Agric., vol. 182, p. 106007, 2021.*
3. *X. Li, W. Hu, C. Shen, Z. Zhang, A. Dick, and A. Van Den Hengel, "A survey of appearance models in visual object tracking," ACM Trans. Intell. Syst. Technol., vol. 4, no. 4, pp. 1–48, 2013.*
4. *L. Kalake, W. Wan, and L. Hou, "Analysis based on recent deep learning approaches applied in real-time multi-object tracking: a review," IEEE Access, vol. 9, pp. 32650–32671, 2021.*
5. *M. Mandal and S. K. Vipparthi, "An empirical review of deep learning frameworks for change detection: Model design, experimental frameworks, challenges and research needs," IEEE Trans. Intell. Transp. Syst., vol. 23, no. 7, pp. 6101–6122, 2021.*
6. *S. Guo et al., "A review of deep learning-based visual multi-object tracking algorithms for autonomous driving," Appl. Sci., vol. 12, no. 21, p. 10741, 2022.*
7. *T. Kriechbaumer, K. Blackburn, T. P. Breckon, O. Hamilton, and M. Rivas Casado, "Quantitative evaluation of stereo visual odometry for autonomous vessel localisation in inland waterway sensing applications," Sensors, vol. 15, no. 12, pp. 31869–31887, 2015.*
8. *A. Geiger, J. Ziegler, and C. Stiller, "Stereoscan: Dense 3d reconstruction in real-time," in 2011 IEEE intelligent vehicles symposium (IV), 2011, pp. 963–968.*
9. *R. E. Kalman, "A new approach to linear filtering and prediction problems," 1960.*
10. *F. Steinbrücker, J. Sturm, and D. Cremers, "Real-time visual odometry from dense RGB-D images," in 2011 IEEE international conference on computer vision workshops (ICCV Workshops), 2011, pp. 719–722.*
11. *M. D. Jenkins, P. Barrie, T. Buggy, and G. Morison, "Extended fast compressive tracking with weighted multi-frame template matching for fast motion tracking," Pattern Recognit. Lett., vol. 69, pp. 82–87, 2016.*
12. *M.-C. Chuang, J.-N. Hwang, K. Williams, and R. Towler, "Tracking live fish from low-contrast and low-frame-rate stereo videos," IEEE Trans. Circuits Syst. Video Technol., vol. 25, no. 1, pp. 167–179, 2014.*
13. *J. Yang, R. Xu, Z. Ding, and H. Lv, "3D character recognition using binocular camera for medical assist," Neurocomputing, vol. 220, pp. 17–22, 2017.*
14. *J. Canny, "A computational approach to edge detection," IEEE Trans. Pattern Anal. Mach. Intell., no. 6, pp. 679–698, 1986.*
15. *V. A. Deepambika and M. A. Rahman, "Illumination Invariant Motion Detection and Tracking Using SMDWT and a Dense Disparity-Variance Method," J. Sensors, vol. 2018, no. 1, p. 1354316, 2018.*
16. *W. L. Richey, J. S. Heiselman, M. J. Ringel, I. M. Meszoely, and M. I. Miga, "Soft tissue monitoring of the surgical field: detection and tracking of breast surface deformations," IEEE Trans. Biomed. Eng., 2023.*
17. *A. Geiger, P. Lenz, C. Stiller, and R. Urtasun, "Vision meets robotics: The kitti dataset," Int. J. Rob. Res., vol. 32, no. 11, pp. 1231–1237, 2013.*
18. *P. F. Alcantarilla, A. Bartoli, and A. J. Davison, "KAZE features," in Computer Vision–ECCV 2012: 12th European Conference on Computer Vision, Florence, Italy, October 7-13, 2012, Proceedings, Part VI 12, 2012, pp. 214–227.*
19. *S. Feng et al., "VIMOT: A Tightly-Coupled Estimator for Stereo Visual-Inertial Navigation and Multi-Object Tracking," IEEE Trans. Instrum. Meas., 2023.*
20. *G. D. Forney, "The viterbi algorithm," Proc. IEEE, vol. 61, no. 3, pp. 268–278, 1973.*
21. *P. Li, H. Zhao, P. Liu, and F. Cao, "Rtm3d: Real-time monocular 3d detection from object keypoints for autonomous driving," in*

European Conference on Computer Vision, 2020, pp. 644–660.

22. *Y. Sun, L. Zheng, Y. Yang, Q. Tian, and S. Wang, "Beyond part models: Person retrieval with refined part pooling (and a strong convolutional baseline)," in Proceedings of the European conference on computer vision (ECCV), 2018, pp. 480–496.*
23. *R. Brunelli and T. Poggiot, "Template matching: Matched spatial filters and beyond," Pattern Recognit., vol. 30, no. 5, pp. 751–768, 1997.*
24. *Y. Wu, J. Lim, and M.-H. Yang, "Online object tracking: A benchmark," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2013, pp. 2411–2418.*
25. *J. Munkres, "Algorithms for the assignment and transportation problems," J. Soc. Ind. Appl. Math., vol. 5, no. 1, pp. 32–38, 1957.*
26. *B. D. Lucas and T. Kanade, "An iterative image registration technique with an application to stereo vision," in IJCAI'81: 7th international joint conference on Artificial intelligence, 1981, vol. 2, pp. 674–679.*
27. *A. P. Witkin, "Scale-space filtering," in Readings in computer vision, Elsevier, 1987, pp. 329–332.*
28. *Y. Cheng, F. Lu, and X. Zhang, "Appearance-based gaze estimation via evaluation-guided asymmetric regression," in Proceedings of the European Conference on Computer Vision (ECCV), 2018, pp. 100–115.*