



Visual Object Tracking Using Machine Learning

Ammar Odeh¹(✉), Ismail Keshta², and Mustafa Al-Fayoumi¹

¹ Department of Computer Science, King Hussein School of Computing Sciences, Princess Sumaya University for Technology Amman, Amman, Jordan
a.odeh@psut.edu.jo

² Department of Computer Science, College of Applied Sciences, AlMaarefa University Riyadh, Riyadh, Kingdom of Saudi Arabia

Abstract. Visual object tracking has become a very active research area in recent years. Each year, a growing number of tracking algorithms are proposed. Object detection and tracking is a critical and challenging task in many critical computer vision applications, including automated video surveillance, traffic monitoring, autonomous robot navigation, and intelligent environments. Object tracking is segmenting an object of interest and tracking its velocity, orientation, and occlusion in a video scene to extract useful information. Over the last two decades, several object tracking approaches have been developed to design a robust object tracker that covers all practical obstacles in real-world operations. This paper reviews recent trends and advances in tracking and assesses the reliability of various trackers based on feature extraction techniques. In video processing, visual tracking has a wide range of applications. When a target is identified in one video frame, it is frequently advantageous to track that object in subsequent frames. Every successful frame in which the target is tracked yields more information about the target's identity and activity. Because tracking is more straightforward than detection, tracking algorithms can require fewer computational resources than object detectors.

Keywords: Visual object tracking (VOT) · Object Representation · Object recognition · Object Classification · Object tracking

1 Introduction

A visual object tracker's goal is to estimate the location of a target in all frames of a video sequence based on the target's initial location (or bounding rectangle) [1, 2]. The computer vision field has studied the object tracking problem for decades. However, creating a reliable and efficient visual object tracking system for all realistic real-world applications remains a challenge. Furthermore, various factors influence the object tracker's performance, including lighting fluctuations, size variations, occlusions, deformations, motion blur, rotations, and low resolutions [3–5].

Visual object tracking VOT is the method of recognizing an object of interest in a sequential manner. It comprises four implementations of this model: target preparation, body shape model, movement forecasting, and target locating [6, 7]. The process of

annotating an object position or region of interest with any of the following representations: object bounding box, ellipse, centroid, object skeleton, object contour, or object silhouette is known as target initialization [8, 9].

In most cases, an object bounding box is provided in the first frame of a sequence, and the tracking algorithm estimates the target position in the subsequent frames. Appearance modeling identifies visual object characteristics to better demonstrate a region of interest and effectively builds a mathematical model for detecting objects using learning techniques [10, 11]. The target location is predicted in successive frames in motion prediction. The target positioning operation entails maximum posterior prediction, also known as greedy search. Limitations set on the appearance and motion models can help to simplify tracking problems [8, 12]. New target appearances are integrated during tracking by updating the appearance and motion models [13, 14].

The basic concept underlying visual object tracking is to follow an object in a sequence of frames, with the first frame containing the center point and surrounding box. It's worth noting that there's presently no viable tracking mechanism that works in all situations where the object's appearance changes.

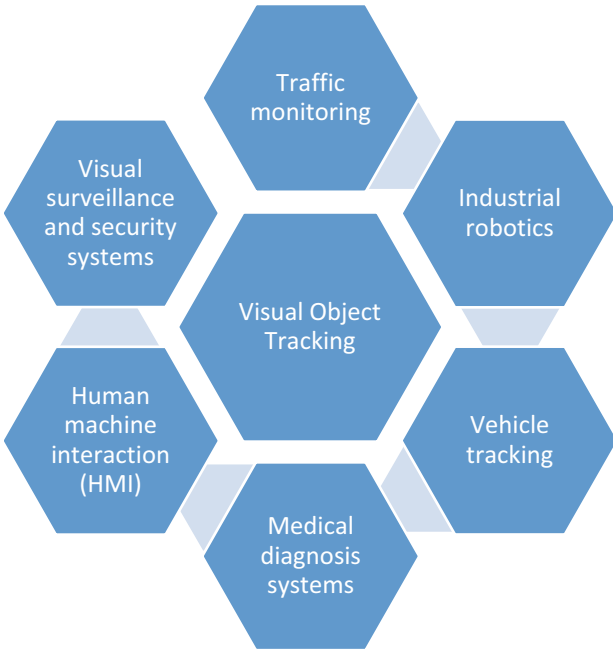


Fig. 1. Visual Object Tracking applications

Figure 1 shows some visual object tracking applications; for example, Interaction between human machines (HMI) in this emerging field, VOT can significantly improve community life by easily interacting with machinery. These systems are omnipresent visual monitoring and security systems, and VOT is an essential part of intelligent visual monitoring. Monitoring of public and defense sites and buildings to detect intruders

[15, 16]. VOT may be employed for engine and road traffic management, such as traffic monitoring, traffic accident detection, etc. [17].

The findings are based on only a few sequences with distinct attributes or factors, which is a prevalent difficulty when evaluating tracking approaches. As a result, the findings do not offer a complete picture of these methods. Conduct a thorough and fair performance evaluation [18].

2 Components of the Object Recognition System

Figure 2 shows a block diagram that shows interactions and data flows among various system components. The model database contains all known models for the system. The data in the model database is determined by the recognition method. It could be anything from a qualitative or functional description to precise geometry data on the surface. Object models are often abstract feature vectors, as mentioned later in this section. A feature is a characteristic that helps characterize and identify a thing related to other objects. Size, color, and shape are the most commonly used characteristics [19–21].

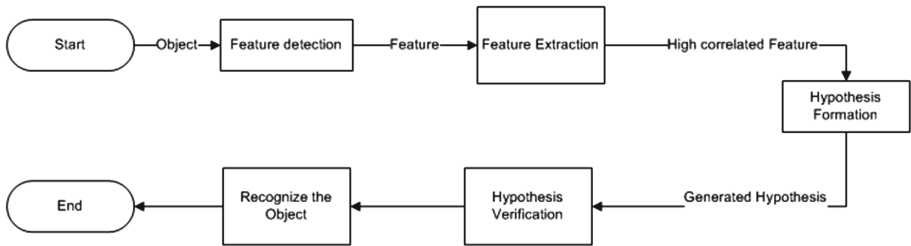


Fig. 2. Components of an object recognition

Operators use the detector on images to discover where features can be used to form hypotheses about the object. The type and organization of the model database of items to be recognized determine the system’s components. The hypothesizer assigns possibilities to items in the picture based on the image’s properties. With some features, this step decreases the recognizer’s search space. The model base is arranged using some sort of indexing technique to make it easier to eliminate implausible item choices from consideration. The verifier then uses object models to verify the hypothesis and enhance the likelihood. The system will select the object as the correct object based on all evidence [22, 23].

All object recognition systems use templates and feature detectors based on these object models, whether explicitly or implicitly. The formulation and verification components of hypotheses vary in importance in different approaches to object recognition. Some systems rely solely on hypothesis development and select the object with the highest probability as the appropriate object. Pattern classification approaches are a good illustration of this strategy. On the other hand, many artificial intelligence systems place less emphasis on hypothesis creation and place more emphasis on checks. The

typical approach of matching the template entirely skips the hypothesis generation stage [24, 25].

An object recognition system must select appropriate tools and technologies for the aforementioned phases. Many considerations must be considered while choosing the best procedures for a particular application. The following are the most important factors to consider while creating an object recognition system:

Object or representation of model: How should objects be displayed in the model database? What are the essential characteristics or characteristics of things in these templates? Geometric descriptions may be available for particular objects and can also be effective, while generic or functional characteristics may require one in another class. An object's image should capture all relevant information without any redundancies and arrange it to allow easy access to the various components of the object recognition system [26].

Extraction feature: What features must be identified, and how can they be reliably detected? Most components can be calculated in two-dimensional images but have to do with three-dimensional object characteristics. Because of the nature of the imaging process, some features can be reliably estimated, while others are very complicated [26, 27].

They are matching feature models: how are image functions compared to database models? Many functionalities and numerous objects exist in most object recognition tasks. A comprehensive combined approach solves the problem of recognition but can be too slow to work. When developing a matching system, the efficiency and the effectiveness of a matching technique must be considered [28].

Formation of hypotheses: How to select a set of likely objects based on the feature matching, and how can each possible object possibility be allocated? The construction of the hypothesis is essentially a heuristic step that reduces the space for search. This step uses application domain expertise to give different objects in the domain some probability or confidence measure. This measure reflects the likelihood that things are presently based on the detected characteristics [29].

Object confirmation: How can object designs be used to select the subject most likely in a given image from the set of likely objects? Each object can be checked for its presence by its models. Every plausible hypothesis must be limited to determine or ignore the object's presence. If models are geometric, the camera location and other scene parameters can quickly check things accurately. In other cases, a hypothesis may not be checkable [30, 31].

3 Dimension Object Classification

Multiple factors affect the object recognition task. We classify the problem of object recognition into the classes below.

3.1 Two-Dimensional

In many applications, images from a distance are obtained sufficiently for the orthographic projection. If the objects are always stable in the scene, they can be considered

double-dimensional. A two-dimensional model base can be used in these applications. Two possibilities exist:

- Objects can be occluded by other items of interest or be partially visible as in a parts issue bin.
- Objects are not occluded as in remote sensing or many industrial applications.

While the objects may be remote, in some cases, they can appear in several stable views in several different positions. In those cases, the problem can also inherently be regarded as recognizing two-dimensional objects.[32].

3.2 Three-Dimensional

If images of objects from arbitrary viewpoints can be obtained, an object can appear in its two views very distinctly. The perspective effect and viewpoint of the image must be considered to recognize the object using three-dimensional models. The three-dimensionality of the models and the images containing only two-dimensional information affect the recognition of objects. Again, whether or not objects are separated from other objects is the two factors to consider [33].

The information used in the object recognition task should be considered in 3-dimensional cases. Two cases are different:

- Intensity: no surface information is explicitly available in images of intensity. Intensity values should be used to recognize characteristics that match the three-dimensional structure of objects.
- 2.5-dimensional images: In many applications, images are available or can be calculated from surface representations with viewer-centric coordinates. In object recognition, this information may be used. Pictures are 2.5 dimensional as well. These images distinguish from a particular point of view to different points in an image.

4 Object Representation

An object may be defined for further examination in a tracking scenario as anything of interest. The following things may be crucial to track in a particular domain: boats in the water, fish within an aquarium, roads, airplane vehicles, people strolling on the road, or bubbles in the water. By their forms and appearances, objects can be represented. In this part, we first discuss the representations of object shapes often used to track and then address the representations of joint form and appearance [34, 35] (Fig. 3).

4.1 Points. The object consists of one point, the center of the object, or several points. The point representation is generally suitable for the following objects in a picture, which occupy small areas [36].

4.2 Main geometrical forms. Object form is a rectangle, ellipse, etc. Object form. Object motion is usually modeled upon by translation, affinity, or projective transformation of the representations (homograph). Although early geometry forms represent simple rigid objects more appropriately, they are also used to track no rigid objects. They are more appropriate [37].

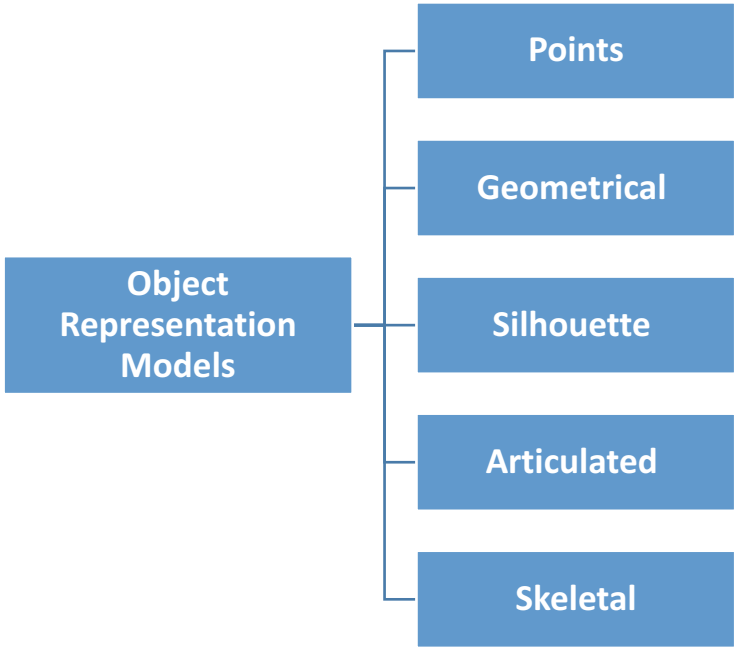


Fig. 3. Object Representation Models

4.3 Silhouette and outline of the object. The depiction of the contour defines the object boundary. The contour region is called the object’s silhouette. The silhouette and contour representations are suitable [38].

4.4 Models of articulated form. Articulated objects consist of body parts held with joints. The human body is a joint object, for example, that has joints connected to its torso, legs, hands, head, and feet. The connection between the components is regulated by films, such as joint angle, etc. The component components can be modeled using cylinders or ellipses to represent an articulated object [39].

4.5 Models of the skeletal. By applying the medial axis to the object’s silhouette, the skeleton can be removed. This model is usually used as a form representation for object recognition. A skeleton representation can be used for modeling articulated and rigid objects [40].

5 Object Recognition

A statistical estimation theory is used to examine the problem of identifying objects subject to the finite transformation of images from a physical angle. In an estimated six-dimensional parameter vector describing an object subject to transformation and generalizing the bundle of the one-dimensional position error previously achieved in the radar and sonar pattern recognition, we focus first on objects that occlude zero-mean scenes with additional noise and thus generalize the band on one-dimensional position error [41–43].

Objects that can be uniquely identified by six affinity criteria and a seventh parameter that specifies the class of objects will be evaluated in complex real-world settings as a problem. The joint probability distribution of our charging coupled (CCD) images for pixel brightness measurements, which are distorted by an additive gauze noise with zero mean additives, is determined using experimental data [44].

This distribution is then used to create the probability function for the refined vector parameter, which defines the object from our image data.

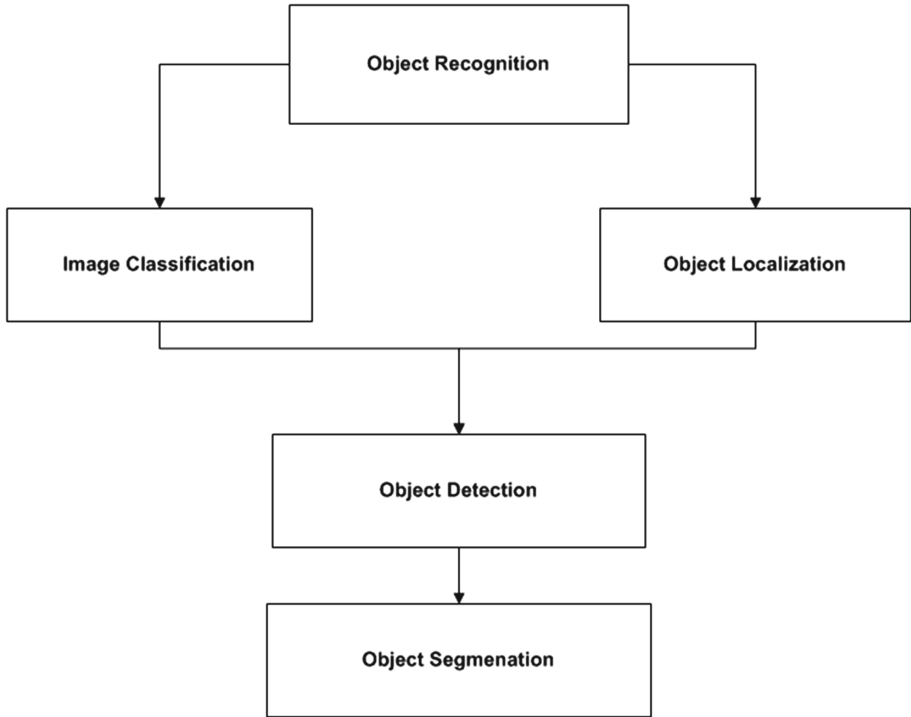


Fig. 4. Object Recognition

Figure 4 shows the object recognition block; the general term object recognition is a collection of related visual tasks involving identifying objects in digital photographs. Classification of images means predicting an object's class in an image. Object localization means identifying and drawing an abundant box around one or more objects in an image. Object detection brings together these two tasks, locates one or more objects, and classifies them in an image.

The object's Fisher information is derived from two practical image descriptors independent of the noise level and is directly calculated from the probability function. The first is a generalized consistency scale, which determines how an object is self-related to an affinity transformation, thereby providing a physical measurement of the extent to which an object can be resolved by affinity [45, 46].

The object's Fisher information is derived from two practical image descriptors independent of the noise level and is directly calculated from the probability function. The first is a generalized consistency scale, which determines how an object is self-related to an affinity transformation, thereby providing a physical measurement of the extent to which an object can be resolved by affinity. The second is the scalar measure of the complexity of an object, which is constantly undergoing affinity with a robust reverse relationship to the ambiguity of recognition [47, 48].

The practical value of this complexity measure is that it can quantitatively describe the level of ambiguity of the problem of recognition. In an estimation of the six-dimensional affine vector parameter, which represents the 2-D position of an object, rotation, dilatation, and skew in a zero-medium scene with added noise, the general Cramer-Rao error is derived. Thus the one-dimensional position error previously associated with radar and sonar pattern recognition is generalized [49].

Authors in [50] develop a recognition method based on the normalized correlation factor to address the problem of recognition of subjects in complex real-world scenes, which usually contain nonzero-mean backgrounds. The coefficient is used to measure the "match" between certain sections of the scene and a "template object." An affine transformation in the corresponding "model image" is used to calculate the template object. Model pictures are collected in advance and represent the classes of recognizable objects.

In the predicted class label, the performance of a model is measured using the average classification error. The model performance is measured using the distance between the predicted and expected bounding box of the predicted class for a unique object location. The results of an object recognition model are evaluated with precise accuracy and recall in each of the best matching boundaries in the image for the known objects [51].

6 Models for Object Tracking

The tracking of objects is now a demanding application in video sequences of the camera. The identification and tracking of objects in video sequences are much more challenging. There are many existing object tracking methods, but there are all inconveniences. Some of the existing Object Tracking models include contour, regional, and dot-based models (Fig. 5).

6.1 Contour-Based Object Tracking Model

An active contour model is used to locate an image's object contour. The objects are plotted as border contours in the contour-based tracking algorithm [52, 53].

These contours are subsequently mistakenly updated in the following frames. This approach is presented in a different version of the active contour model. The discrete approach utilizes the point distribution model to limit the shape. This algorithm, however, is highly sensitive to tracking initialization, which makes automatic tracking complicated to begin.

An object contour tracking algorithm for tracking video sequence contours of objects. The active contour was segmented using their algorithm's graph-cut image segmentation

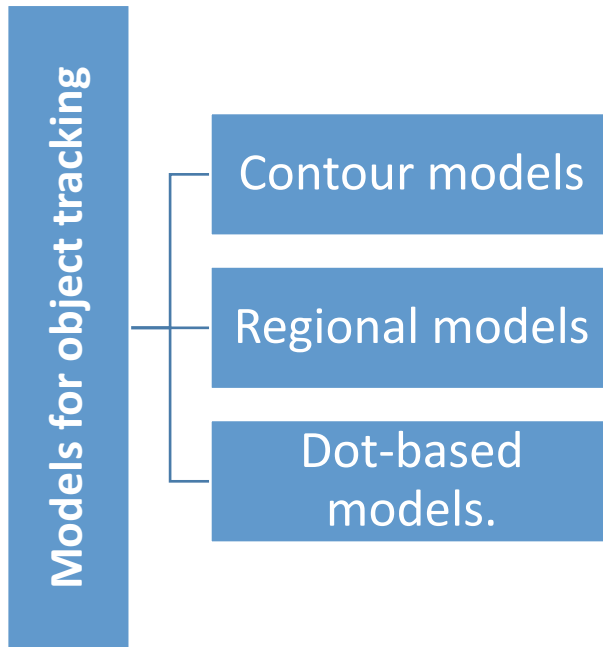


Fig. 5. Models for object tracking

method. Every frame has the resulting contour of the previous frame. The intensity of data from the current and differential frame and the previous frame is used to detect a new object contour [54].

They used the combination of the weighted gradient and contours of the object for the driver-face tracking problem. They calculated the gradient of an image in the segmentation step. They proposed an object tracking gradient-based attraction field [54].

Neural Fuzzy tracking model of an active contour-based object. The shaped model is used for the object's characteristic vector extraction. Their approach uses the self-construction neural fluids inference network to train and recognize moving objects. In this paper, they took horizontal and vertical projections of the histograms of the figurative of the human body, transforming them through a Discrete Fourier Transformation (DFT) [55].

The two-stage object tracking method. The kernel method is first used to find an object in a complex environment, such as partial occlusions, conflicts, etc. Again, they used a contour-based method to improve tracking results and precisely tracked the object contour after the target location. The initial target position is predicted and evaluated in the target localization step with the Kalman filter and Bhattacharyya coefficient [56].

The multi-hypothesis algorithm uses color and contour integration information based on the particle filter. The Sobel operator is used to detect contours. The shape similitude is assessed by corresponding points in the two contour images between the observing and sample positions [57].

Through the multi-feature fusion approach, the rough position of the object is found. They have extracted contours for accurate and robust object contour tracking using regionally-based object contour removal. A color histogram and the corner Harris feature a fusion method to provide the object's rough location in their model. They used the Harris corner fusion method for the particle filter method [58].

The region-based tempo difference model is used in the object contour detection step, which results in the rough location tracking result. A practical contour tracking framework for objects. The framework they proposed included different models, such as initialization tracking algorithms, algorithms of color-based development of contours, and the evolution of adaptive form contours and dynamic form model based on the Markov model [59].

The automatic and fast-tracking initialization algorithm uses optical flow detection. In the color-based algorithm for contour evolution, the correlations between the values of adjoining pixels to estimate the probability are measured using the Markov random field (MRF) theory [60].

Their algorithm for adaptive shape evolution combines the color feature alone and the shape priors to achieve the final outline. A new PCA technique is being implemented to update the form model and make it flexible to refresh it. The dominant set clustering is used in the dynamical model based on Markov to achieve the typical form modes of periodic movement [61].

Multiple object tracking algorithms modified contour-based by point processing. Multiple objects have the benefit of this approach. Their system is capable of detecting and tracking people in indoor videos. Their background estimation method was the Gaussian mixture model (GMM) [62].

6.2 Region-Based Object Tracking Model

The object model is based on a region based on the color distribution of the tracked object for tracking of objects. It shows the color-based object. It is therefore computer-efficient. However, its efficiency is declined when several objects move in the image sequences together. Accurate tracking is not possible when several objects move as a result of occlusion [63].

Furthermore, the object tracking depends mainly on the background model used in extracting the object outlines if no object-form information exists.

An Adaptive Kalman Filter Corner-based Method for tracking objects. The moving object corner function is used first. Then, to set up the estimate parameters automatically for the Kalman Filter, you use the number of corner points in consecutive frames. The discriminatory features were chosen using the object/background separation voting strategy. They introduced an improved mid-size algorithm for object tracking using discriminative features [64].

The FLIR Object Tracking Framework is based on a mean shifting algorithm and features matching for forward-looking infrared imagery. In the corresponding feature step, the Harris detector was used to extract template object and candidate area feature points. To measure the resemblance of feature points, they further developed a better Hausdorff distance [65].

Self-adaptive tracking panel views based on the location and NMI function of the target center. The standard inertia moment (NMI) characteristics are combined to locate the tracking object center in real-time. A mean algorithm for shifting is available to track the object [66].

The enhanced tracking method tracks both single objects and several objects in video sequences that can move the object quickly or slowly. The method proposed is based on the subtraction of the background and the matching of SIFT features. The object is detected with the aid of background subtraction. Combining motion characteristics and SIFT characteristics helps to detect and track the object [67].

The new object tracking framework combines the sift feature and the color and particle filter combinations. For target reproduction and localization, SIFT features are used. The transformation of an image produces a local feature vector. The scaling, translation rotation, and illumination changes are all subjects of the feature vectors. Approximation of the solution to the sequential estimate is found with the particulate filter (PF) [68].

Algorithm for object tracking based on Mean Shift and selection of online functionality. The objective object is defined in 4-D state space. Function space in R, G, and B channels are created according to color pixel values. The best space to track the objects and background scenes during tracking is chosen. The state estimation of the objects is done using a Kalman filter in their algorithm. The robust online tracking method applies adaptive classifiers in consecutive frames to match the detected vital points. The approach proposed shows that integrating the robust local feature and the adaptive online boosting algorithm can contribute to the adaptation of different frames [69].

Real-time picture processing on mobile devices. They use a holistic hair-like feature to track exciting objects. The robustness of their method was achieved with the help of an online feature update system. A color filtering feature detection method for tracking recovery: an algorithm combines motion detection, function extraction, and block matching background information. The four adjacent directions are detected by a series of features called Shape Control Points (SCPs). Using an adaptive background generation method, the weakness of the block matching algorithm was reduced [70].

The representative object appearances have been stored as candidate templates during tracking, and the best template is selected to match new frames. This template addition process is updated with further object appearances and changed via the online strategy. They showed that feature-based methods could be extended to objects that are not planned or undergo significant modifications. Extended their object tracking feature-based method using sparse shape points. Possible data association events with the particulate filter are sampled. The filter also helps to estimate the global position and velocity of the object. They have used time, together with partial regression of the least square, to improve the performance of the tracker [71].

A multi-part SIFT rotating object tracking feature model for tracking objects. The reference and target object are represented to extract the possible significant similarity points measurement. The filter solves the state space estimate when the state equation is non-linear and the subsequent density is non-Gaussian. A tracking particle filter that is useful for non-linear or non-Gaussian problems. They use Bhattacharyya object distance and the predicted position of the object obtained by the particle filter to find the posterior

probability of the particle filter. The rear chance is used to update the filter's status. Their experiment has shown HSV to be the optimal color room for changes in scale, occlusion, and lighting [72].

New Distance Metric Learning (DML) tracking framework for Object Tracking combined with Nearest Neighbor (NN) grades. They used a canny edge detector to detect the object; the object can be distinguished from other objects by using the Nearest Neighbor Classifier. The background can be removed from the Nearest Neighbor (NN) algorithm framework. The algorithm of the closest neighbor uses the distance between the object and the background to remove it. The object is then determined based on skin color utilizing a blob detector. An abounding box is created for the identified object [73].

Enhanced Monte Carlo (MCMC) Markov chain was known as the MCMC (OF-MCMC) vehicle flow sampling algorithm. The automatic movement model has been applied to achieve the moving direction of the vehicle in the initial frames using the optical flow method, which resolves the scale change problem and the moving speed of the object. They have produced a more accurate feature template with different weighted features to handle vehicle tracking in low resolution of video data and obtain better follow-up results [74].

7 Feature Point-Based Tracking Algorithm

Feature-point models are used to describe objects. The feature-point tracking algorithm has three basic steps. The first step is to detect and track the object with elements extracted. The second step is to group them into higher levels. The last step is to match the extracted features in successive frames between images. The essential steps in function-driven object tracking are feature extraction and feature correspondence. The challenge of tracking functions is the correspondence of a feature because a point in one picture may have many similar points in a different image, which leads to ambiguity of feature correspondence [75].

New video sequence segmenting method for objects supervised. The user entry object outline is considered to be a video object in the proposed method. The model included the segmentation of the area and the motion estimation of the object in moving object tracking. The active contour model is also used [76]. The backward region-based classification video object tracking system. Their system comprises five phases, region pre-processing, region extraction, regionally-based motion estimation, area classification, and regional post-processing [77].

A combination of morphological segmentation tools and human support can be used to locate a semantic video object boundary. Motion estimates, video object compensation, and iframes border information are taken in the remaining frames to identify other video objects [78]. The object partition is initialized in the initial frame can the object tracking algorithm avoid segmentation. The tracking process is carried out with object limit forecast using block motion vectors and then updating the object contour by occlusion/discussion detection method. They used an adaptive block-based approach to estimate the motion between frames. Changes to the algorithm for dis-occlusion detection help to develop algorithms for occlusion detection by considering the duality principle [79].

Descriptors are derived from regions for segmentation and tracking. Partitioning an image into a series yields the image's homogeneous regions. As a result, the problem of object extraction shifts from pixel-based to database-based analysis. Two trackers are essentially composed of an object extraction algorithm. The pixel-wise tracker retrieves an object using the Adaboost-based global color selection function. The region-specific tracker is done at the start to regionalize each frame K-means clustering. The region tracking is performed using a two-way labeling system [80].

A backdrop image updating approach is utilized to ensure accurate object detection in a confined setting. The filter has been used to create a robust object tracking framework under challenging situations and considerably enhance the accuracy of estimates for complicated tracking challenges [81].

Automatic background modeling detection and tracking of moving objects. Instead of geometrical limits, their proposed region level-set approach was employed to detect and follow motion via statistics on image intensity within each subset. Background modeling is completed before turning to object segmentation and tracking [82].

A generic object tracking and segmentation region-based particle filter. Its approach combines a particle filter based on color and a particular filter based on region. The program reliably tracks objects and delivers a precise segmentation during the sequence. Particle filters use Multi-hypotheses to monitor things [83]. A robust 3D tracking model can extract independent item motion paths in an uncontrolled environment. Two new algorithms, including motion segmentation and Mean Shift tracking approaches based in the region, have been developed. A Kalman filter is used to integrate their tracking results from both algorithms [84].

8 Conclusion

This paper gives a literature classification and a quick survey of related subjects on visual object tracking approaches. The components of the object recognition system are made up of a series of phases that begin with feature detection and continue with feature extraction to locate highly correlated features before constructing the hypothesis system to recognize the object. Points, Geometrical, Silhouette, Articulated, and Skeletal, are the five kinds of Object Representation Models. There are three types of tracking techniques: contour, region, and function. With rich theoretical details of the tracking algorithms and bibliographic contents, we intend to contribute to research on object tracking in images and promote future studies.

Acknowledgments. This research was supported by Princess Sumaya University for Technology (PSUT) and Researchers Supporting Program (TUMA-Project-2021-14), AlMaarefa University.

Conflict of Interest. The authors declare no conflict of interest.

References

1. Xuan, S., et al.: Siamese networks with distractor-reduction method for long-term visual object tracking. *Pattern Recogn.* **112**, 107698 (2021)

2. Jiang, M., et al.: High speed long-term visual object tracking algorithm for real robot systems. *Neurocomputing* **434**, 268–284 (2021)
3. Mehmood, K., et al.: Context-aware and occlusion handling mechanism for online visual object tracking. *Electronics* **10**, 43 (2021)
4. Wang, Y., Ma, J.: Visual object tracking using surface fitting for scale and rotation estimation. *KSII Trans. Internet Inf. Syst.* **15** (2021)
5. Wu, J., et al.: Towards accurate estimation for visual object tracking with multi-hierarchy feature aggregation. *Neurocomputing* **451**, 252–264 (2021)
6. Rinnert, P., Nieder, A.: Neural code of motor planning and execution during goal-directed movements in crows. *J. Neurosci.* **41**, 4060–4072 (2021)
7. Clarence, A., et al.: Unscripted retargeting: reach prediction for haptic retargeting in virtual reality. In: 2021 IEEE Virtual Reality and 3D User Interfaces (VR), pp. 150–159 (2021)
8. Zhao, H., et al.: Deep mutual learning for visual object tracking. *Pattern Recogn.* **112**, 107796 (2021)
9. Guo, Q., et al.: Exploring the effects of blur and deblurring to visual object tracking. *IEEE Trans. Image Process.* **30**, 1812–1824 (2021)
10. Jia, S., et al.: IoU attack: towards temporally coherent black-box adversarial attack for visual object tracking. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 6709–6718 (2021)
11. Zhu, J., Zhang, G., Zhou, S., Li, K.: Relation-aware Siamese region proposal network for visual object tracking. *Multimedia Tools Appl.* **80**(10), 15469–15485 (2021). <https://doi.org/10.1007/s11042-021-10574-z>
12. Chen, Y., Wang, J., Xia, R., Zhang, Q., Cao, Z., Yang, K.: The visual object tracking algorithm research based on adaptive combination kernel. *J. Ambient. Intell. Humaniz. Comput.* **10**(12), 4855–4867 (2019). <https://doi.org/10.1007/s12652-018-01171-4>
13. Lee, S.-H., et al.: Learning discriminative appearance models for online multi-object tracking with appearance discriminability measures. *IEEE Access* **6**, 67316–67328 (2018)
14. He, M., et al.: Fast online multi-pedestrian tracking via integrating motion model and deep appearance model. *IEEE Access* **7**, 89475–89486 (2019)
15. Franzoni, V., et al.: Emotional machines: the next revolution. In: *Web Intelligence*, pp. 1–7 (2019)
16. Li, S., Yeung, D.-Y.: Visual object tracking for unmanned aerial vehicles: a benchmark and new motion models. In: *Thirty-First AAAI Conference on Artificial Intelligence* (2017)
17. Cakir, S., Cetin, A.E.: Visual object tracking using Fourier domain phase information. *Signal Image Video Process.* 1–8 (2021)
18. Yuan, D., Zhang, X., Liu, J., Li, D.: A multiple feature fused model for visual object tracking via correlation filters. *Multimedia Tools Appl.* **78**(19), 27271–27290 (2019). <https://doi.org/10.1007/s11042-019-07828-2>
19. Chowdhury, P.R., et al.: Brain Inspired Object Recognition System. arXiv preprint [arXiv:2105.07237](https://arxiv.org/abs/2105.07237) (2021)
20. Dawod, M., Hanna, S.: BIM-assisted object recognition for the on-site autonomous robotic assembly of discrete structures. *Constr. Robot.* **3**(1–4), 69–81 (2019). <https://doi.org/10.1007/s41693-019-00021-9>
21. Poza-Lujan, J.-L., et al.: Distributed architecture to integrate sensor information: object recognition for smart cities. *Sensors* **20**, 112 (2020)
22. Girish, S., et al.: The lottery ticket hypothesis for object recognition. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 762–771 (2021)
23. Fu, J., et al.: A multi-hypothesis approach to pose ambiguity in object-based SLAM. arXiv preprint [arXiv:2108.01225](https://arxiv.org/abs/2108.01225) (2021)

24. Kutschbach, T., et al.: Sequential sensor fusion combining probability hypothesis density and kernelized correlation filters for multi-object tracking in video data. In: 2017 14th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), pp. 1–5 (2017)
25. Wang, Q., et al.: HypoML: visual analysis for hypothesis-based evaluation of machine learning models. *IEEE Trans. Visual Comput. Graphics* **27**, 1417–1426 (2020)
26. Long, L., et al.: Object-level representation learning for few-shot image classification. arXiv preprint [arXiv:1805.10777](https://arxiv.org/abs/1805.10777) (2018)
27. Hubert, C.: More on the model: building on the ruins of representation. *Archit. Des.* **91**, 14–21 (2021)
28. Li, Z., et al.: Self-guided adaptation: progressive representation alignment for domain adaptive object detection. arXiv preprint [arXiv:2003.08777](https://arxiv.org/abs/2003.08777) (2020)
29. Huang, J.: Auto-attentional mechanism in multi-domain convolutional neural networks for improving object tracking. *Int. J. Intell. Comput. Cybern.* (2021)
30. Bekiroglu, Y., et al.: Learning tactile characterizations of object-and pose-specific grasps. In: 2011 IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 1554–1560 (2011)
31. La Porta, F., et al.: Unified Balance Scale: an activity-based, bed to community, and aetiology-independent measure of balance calibrated with Rasch analysis. *J. Rehabil. Med.* **43**, 435–444 (2011)
32. Du, B., et al.: A discriminative manifold learning based dimension reduction method for hyperspectral classification. *Int. J. Fuzzy Syst.* **14**, 272–277 (2012)
33. Hsiao, E., Hebert, M.: Occlusion reasoning for object detection under arbitrary viewpoint. *IEEE Trans. Pattern Anal. Mach. Intell.* **36**, 1803–1815 (2014)
34. Bajcsy, R.: Three-dimensional object representation. In: Kittler, J., Fu, K.S., Pau, L.F. (eds) *Pattern Recognition Theory and Applications*, pp. 283–295. Springer, New York (1982). https://doi.org/10.1007/978-94-009-7772-3_17
35. Moghaddam, B., Pentland, A.: Probabilistic visual learning for object representation. *IEEE Trans. Pattern Anal. Mach. Intell.* **19**, 696–710 (1997)
36. Laurentini, A.: The visual hull concept for Silhouette-based image understanding. *IEEE Trans. Pattern Anal. Mach. Intell.* **16**, 150–162 (1994)
37. Ashok, V., Ganapathy, D.: A geometrical method to classify face forms. *J. Oral Biol. Craniofac. Res.* **9**, 232–235 (2019)
38. Wagemans, J., et al.: Identification of everyday objects on the basis of Silhouette and outline versions. *Perception* **37**, 207–244 (2008)
39. Sapp, B., et al.: Cascaded models for articulated pose estimation. In: Daniilidis, K., Maragos, P., Paragios, N. (eds.) *Computer Vision – ECCV 2010*. LNCS, vol. 6312, pp. 406–420. Springer, Heidelberg (2010). https://doi.org/10.1007/978-3-642-15552-9_30
40. Geldof, A.A.: Models for cancer skeletal metastasis: a reappraisal of Batson’s plexus. *Anticancer Res.* **17**, 1535–1539 (1997)
41. Jarrett, K., et al.: What is the best multi-stage architecture for object recognition? In: 2009 IEEE 12th International Conference on Computer Vision, pp. 2146–2153 (2009)
42. Xu, G., Zhang, Z.: *Epipolar Geometry in Stereo, Motion and Object Recognition: A Unified Approach*, vol. 6. Springer, New York (2013). <https://doi.org/10.1007/978-94-015-8668-9>
43. Riesenhuber, M., Poggio, T.: Models of object recognition. *Nat. Neurosci.* **3**, 1199–1204 (2000)
44. Grasselli, G., et al.: Quantitative three-dimensional description of a rough surface and parameter evolution with shearing. *Int. J. Rock Mech. Min. Sci.* **39**, 789–800 (2002)
45. Dutton, Z., et al.: Attaining the quantum limit of superresolution in imaging an object’s length via predetection spatial-mode sorting. *Phys. Rev. A* **99**, 033847 (2019)

46. Betke, M., Makris, N.C.: Information-conserving object recognition. In: Sixth International Conference on Computer Vision (IEEE Cat. No. 98CH36271), pp. 145–152 (1998)
47. Barrett, H.H., et al.: Objective assessment of image quality. II. Fisher information, Fourier crosstalk, and figures of merit for task performance. *JOSA A* **12**, 834–852 (1995)
48. Betke, M., Makris, N.C.: Recognition, resolution, and complexity of objects subject to affine transformations. *Int. J. Comput. Vision* **44**, 5–40 (2001)
49. Tian, T., et al.: Cramer-Rao bounds of localization estimation for integrated radar and communication system. *IEEE Access* **8**, 105852–105863 (2020)
50. Zheng, Y., et al.: A new precision evaluation method for signals of opportunity based on Cramer-Rao lower bound in finite error. In: 2019 Chinese Control Conference (CCC), pp. 3934–3939 (2019)
51. Lee, S., et al.: Estimation error bound of battery electrode parameters with limited data window. *IEEE Trans. Industr. Inf.* **16**, 3376–3386 (2019)
52. Li, X., Yang, F., Cheng, H., Liu, W., Shen, D.: Contour knowledge transfer for salient object detection. In: Ferrari, V., Hebert, M., Sminchisescu, C., Weiss, Y. (eds.) ECCV 2018. LNCS, vol. 11219, pp. 370–385. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-01267-0_22
53. Gong, X.-Y., et al.: An overview of contour detection approaches. *Int. J. Autom. Comput.* **15**, 656–672 (2018)
54. Philbrick, K.A., et al.: RIL-contour: a medical imaging dataset annotation tool for and with deep learning. *J. Digit. Imaging* **32**, 571–581 (2019)
55. Dai, Y., et al.: Trajectory tracking control for seafloor tracked vehicle by adaptive neural-fuzzy inference system algorithm. *Int. J. Comput. Commun. Control* **13**, 465–476 (2018)
56. Guan, W., et al.: Visible light dynamic positioning method using improved Camshift-Kalman algorithm. *IEEE Photonics J.* **11**, 1–22 (2019)
57. Hu, B., Niebur, E.: A recurrent neural model for proto-object based contour integration and figure-ground segregation. *J. Comput. Neurosci.* **43**(3), 227–242 (2017). <https://doi.org/10.1007/s10827-017-0659-3>
58. Qin, J., et al.: An encrypted image retrieval method based on Harris corner optimization and LSH in cloud computing. *IEEE Access* **7**, 24626–24633 (2019)
59. Cheng, J., et al.: Hidden Markov model-based nonfragile state estimation of switched neural network with probabilistic quantized outputs. *IEEE Trans. Cybern.* **50**, 1900–1909 (2019)
60. Cai, D., et al.: A moving target detecting and tracking system based on DSP. In: 2017 International Conference on Optical Instruments and Technology: Optoelectronic Imaging/Spectroscopy and Signal Processing Technology, p. 106200Z (2018)
61. Dimeas, F., Doulgeri, Z.: Progressive automation of periodic tasks on planar surfaces of unknown pose with hybrid force/position control. In: 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 5246–5252 (2020)
62. Zong, B., et al.: Deep autoencoding Gaussian mixture model for unsupervised anomaly detection. In: International Conference on Learning Representations (2018)
63. Lee, H., Kim, D.: Salient region-based online object tracking. In: 2018 IEEE Winter Conference on Applications of Computer Vision (WACV), pp. 1170–1177 (2018)
64. Yu, T.T., War, N.: Condensed object representation with corner HOG features for object classification in outdoor scenes. In: 2017 18th IEEE/ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD), pp. 77–82 (2017)
65. Wang, X., et al.: Aerial infrared object tracking via an improved long-term correlation filter with optical flow estimation and SURF matching. *Infrared Phys. Technol.* **116**, 103790 (2021)
66. Sadegh, A.M., Worek, W.M.: Marks' Standard Handbook for Mechanical Engineers: McGraw-Hill Education (2018)

67. Chhabra, P., Garg, N.K., Kumar, M.: Content-based image retrieval system using ORB and SIFT features. *Neural Comput. Appl.* **32**(7), 2725–2733 (2018). <https://doi.org/10.1007/s00521-018-3677-9>
68. Amaya, M., et al.: Adaptive sequential Monte Carlo for posterior inference and model selection among complex geological priors. *Geophys. J. Int.* **226**, 1220–1238 (2021)
69. Bae, S.-H., Yoon, K.-J.: Confidence-based data association and discriminative deep appearance learning for robust online multi-object tracking. *IEEE Trans. Pattern Anal. Mach. Intell.* **40**, 595–610 (2017)
70. Mbakop, S., et al.: Inverse dynamics model-based shape control of soft continuum finger robot using parametric curve. *IEEE Robot. Autom. Lett.* **6**, 8053–8060 (2021)
71. Liu, F., et al.: Robust visual tracking revisited: From correlation filter to template matching. *IEEE Trans. Image Process.* **27**, 2777–2790 (2018)
72. Kaskman, R., et al.: HomebrewedDB: RGB-D dataset for 6D pose estimation of 3D objects. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops* (2019)
73. Xu, J., et al.: Bilevel distance metric learning for robust image recognition. *Adv. Neural. Inf. Process. Syst.* **31**, 4198–4207 (2018)
74. Cabeza de Vaca, I., et al.: Enhanced Monte Carlo methods for modeling proteins including computation of absolute free energies of binding. *J. Chem. Theory Comput.* **14**, 3279–3288 (2018)
75. Guler, Z., et al.: A new object tracking framework for interest point based feature extraction algorithms. *Elektronika ir Elektrotechnika* **26**, 63–71 (2020)
76. Pareek, A., et al.: A robust surf-based online human tracking algorithm using adaptive object model. In: *Proceedings of International Conference on Artificial Intelligence and Applications*, pp. 543–551 (2021)
77. Rejeesh, M.: Interest point based face recognition using adaptive neuro fuzzy inference system. *Multimedia Tools Appl.* **78**, 22691–22710 (2019)
78. Kann, K., et al.: Fortification of neural morphological segmentation models for polysynthetic minimal-resource languages. *arXiv preprint arXiv:1804.06024* (2018)
79. Noyel, G., et al.: Morphological segmentation of hyperspectral images. *arXiv preprint arXiv:2010.00853* (2020)
80. Yang, X., et al.: A face detection method based on skin color model and improved AdaBoost algorithm. *Traitement du Signal*, vol. 37 (2020)
81. Hameed, K., et al.: A sample weight and AdaBoost CNN-based coarse to fine classification of fruit and vegetables at a supermarket self-checkout. *Appl. Sci.* **10**, 8667 (2020)
82. Sun, Y., et al.: Active perception for foreground segmentation: an RGB-D data-based background modeling method. *IEEE Trans. Autom. Sci. Eng.* **16**, 1596–1609 (2019)
83. Voigtlaender, P., et al.: MOTs: multi-object tracking and segmentation. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 7942–7951 (2019)
84. Gunjal, P.R., et al.: Moving object tracking using Kalman filter. In: *2018 International Conference on Advances in Communication and Computing Technology (ICACCT)*, pp. 544–547 (2018)