



Vision Tracking: A Survey of the State-of-the-Art

Anjan Dutta¹ · Atreyee Mondal¹ · Nilanjan Dey¹ · Soumya Sen² · Luminița Moraru³ · Aboul Ella Hassanien⁴

Received: 24 October 2019 / Accepted: 23 December 2019 / Published online: 11 January 2020
© Springer Nature Singapore Pte Ltd 2020

Abstract

Vision tracking is a well-studied framework in vision computing. Developing a robust visual tracking system is challenging because of the sudden change in object motion, cluttered background, partial occlusion and camera motion. In this study, the state-of-the-art visual tracking methods are reviewed and different categories are discussed. The overall visual tracking process is divided into four stages—object initialization, appearance modeling, motion estimation, and object localization. Each of these stages is briefly elaborated and related researches are discussed. A rapid growth of visual tracking algorithms is observed in last few decades. A comprehensive review is reported on different performance metrics to evaluate the efficiency of visual tracking algorithms which might help researchers to identify new avenues in this area. Various application areas of the visual tracking are also discussed at the end of the study.

Keywords Visual tracking · Visual computing · Motion estimation · Object motion · Object localization

Introduction

Visual tracking is one of the significant problems in computer vision having wide range of application domains. A remarkable advancement of the visual tracking algorithm is

observed because of the rapid increase in processing power and availability of high resolution cameras over the last few decades in the field of automated surveillance [1], motion-based recognition [2], video indexing [3], vehicle navigation [4], and human–computer interaction [5, 6]. Visual tracking can be defined as, estimating the trajectory of the moving object around a scene in the image plane [7].

Various computer vision tasks to detect, track and classify the target from image sequences are grouped in visual surveillance to analyze the object behavior [7]. A better surveillance system is developed by integrating the motion detection and visual tracking system in [8]. A content-based video indexing technique is evolved from object motion in [9]. The proposed indexing method is applied to analyze the video surveillance data. Visual tracking is effectively applied in vehicle navigation. A method for object tracking and detection is developed in [10] for maritime surface vehicle navigation using stereo vision system to locate objects as well as calculating the distance from the target object in the harsher maritime environment. A methodology of human computer interaction to compute eye movement by detecting the eye corner and the pupil center using visual digital signal processor camera is invented in [11]. The mentioned novel approach helps the users to move their head freely without wiring any external gadgets.

In visual tracking system, the 3D world is projected on a 2D image that results in loss of information [12]. The problem becomes more challenging due to the presence of

✉ Anjan Dutta
anjan.dutta.edu@gmail.com

Atreyee Mondal
atreyee.shakshi@gmail.com

Nilanjan Dey
neelanjan.dey@gmail.com

Soumya Sen
iamsoumyasen@gmail.com

Luminița Moraru
Luminita.moraru@ugal.ro

Aboul Ella Hassanien
aboitcairo@gmail.com

¹ Department of Information Technology, Techno International New Town (Formerly known as Techno India College of Technology), Kolkata, India

² A.K. Choudhury School of Information Technology, University of Calcutta, Kolkata, India

³ Department of Physics, Faculty of Sciences, Univ. “Dunarea de Jos”, Str. Domneasca, nr. 111, 800201 Galati, Romania

⁴ Information Technology Department, Faculty of Computers and Information, Cairo University, Giza, Egypt

noise in images, unorganized background, random complex target motion, object occlusions, non-rigid object, variation in the number of objects, change in illumination, etc. [13]. These issues need to be handled effectively to prevent the degradation of tracking performance and even failure. Different visual representations and statistical models are used in literature to deal with these challenges. These models use state-of-the-art algorithms and different methodologies for visual tracking. Different metrics are used to effectively measure the performance of the tracker. Motivated by this, different state-of-the-art visual tracking models widely used in literature are discussed in this paper. In each and every year, a substantial number of algorithms for visual tracking are proposed in literature. To efficiently evaluate their performance, different performance metrics for robust evaluation of trackers are elaborated here after vividly describing the tracking models. Several popular application domains of visual tracking are identified and briefly described here. One can have overall overview of visual tracking methods and best practices as well as a vivid idea about the different application domains related to visual tracking from this study.

A visual tracking system consists of four modules, i.e., object initialization, appearance modeling, motion

estimation and object localization. Each of these components and associated tracking methods are briefly described in Sect. 2. Some popular performance measures for visual tracking, for, e.g., center location error, bounding box overlap, tracking length, failure rate, area under the lost-track-ratio curve, etc. are discussed in Sect. 3. Progress in visual tracking methodologies introduced a revolution in health care, space science, education, robotics, sports, marketing, etc. Section 4 highlights some pioneering works related to different application domains of visual tracking. Conclusion section is presented in Sect. 5.

Visual Tracking Methods

In visual tracking system, a trajectory of the target over the time is generated [14] based on the location of the target, positioned in consecutive video frames. The detected objects from the consecutive frames maintained a correspondence [15] using visual tracking mechanism.

The fundamental components of a visual tracking system are object initialization, appearance modeling, motion estimation and object localization [16]. Figure 1 reports the detailed taxonomy of vision tracking.

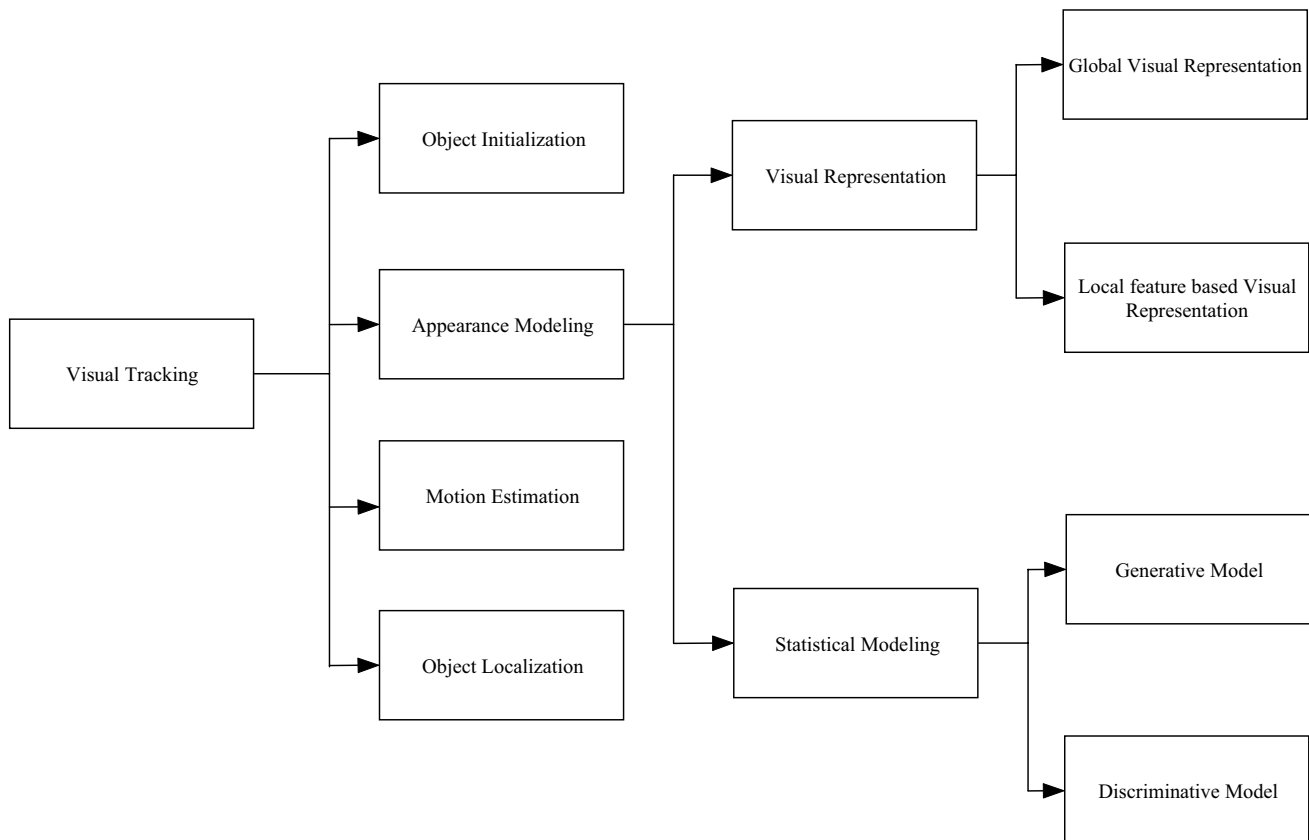


Fig. 1 Visual tracking taxonomy

Object Initialization

Manual or automatic object initialization is the initial step of visual tracking methods. Manual annotation using bounding boxes or ellipses is used to locate the object [17]. Manual annotation is a time-consuming human-biased process, which claims an automated system for easily, efficiently and accurately locating and initializing the target object. In recent decades, automated initialization has wide domains for real-time problem solving (for, e.g. face detection [18], human tracking, robotics, etc. [19–22]). A dynamic framework for automated initialization and updating the face feature tracking process is proposed in [23]. Moreover, a new method to handle self-occlusion is presented in this study. This approach matched each candidate with a set of predefined standard eye templates, by locating the eyes of the candidates. Once the subject's eyes are located accurately, lip control points are located using the standard templates. An automated, integrated model comprising of robust face and hand detection for initializing a 3D body tracker to recover from failure is proposed in [24]. Useful data for initialization and validation are provided to the intended tracker by this system.

Object initialization is the prerequisite for the appearance modeling. A detailed description of appearance modeling is reported in the following section.

Appearance Modeling

The majority of the object properties (appearance, velocity, location, etc.) are described by the appearance or observation model [25]. Various special features are used to differentiate the target and background or different objects in a tracking system [26]. Features like color, gradient, texture, shape, super-pixel, depth, motion, optic flow, etc. or fused features are most commonly used for robust tracking to describe the object appearance model.

Appearance modeling is done by visual representation and statistical modeling. In visual representation, different variants of visual features are used to develop effective object descriptors [27]. Whereas, statistical learning techniques are used in statistical modeling to develop mathematical models that are efficient for object identification [28]. A vivid description of these two techniques is given in the below section.

Visual Representation

Global Visual Representation Global visual representation represents the global statistical properties of object appearance. The same can also be represented by various other representation techniques, namely—(a) raw pixel values (b) optical flow method (c) histogram-based representation (d)

covariance-based representation (e) wavelet filtering-based representation and (f) active contour representation.

(a) Raw pixel values

Values based on raw pixels are the most frequently used features in vision computing [29, 30] for the algorithmic simplicity and efficiency [31]. Raw color or intensity information of the raw pixels is utilized to epitomize the object region [32]. Two basic categories for raw pixel representation are—vector based [33, 34] and matrix based [35–37].

In vector-based representation, an image region is transformed into a higher-dimensional vector. Vector-based representation performed well in color feature-based visual tracking. Color features are robust to object deformation, insensitive to shape variation [38], but suffer from small sample size problem and uneven illumination changes [39].

To overcome the above-mentioned limitations of vector-based representation, matrix-based representation is proposed in [40, 41]. In matrix-based representation, the fundamental data units for object representation are built using the 2D matrices or higher-order tensors because of their low-dimensional property.

Various other visual features (e.g., shape, texture, etc.) are embedded in the raw pixel information for robust and improved visual object tracking. A color histogram-based similarity metric is proposed in [42], where the region color and the special layout (edge of the colors) are fused. A fused texture-based technique is proposed to enrich the color features in [43].

(b) Optical flow representation

The relative motion of the environment with respect to an observer is known as optical flow [44]. The environment is continuously viewed to find the relative movement of the visual features, e.g., points, objects, shapes, etc. Inside an image region, optical flow is represented by dense field displacement vectors of each pixel. The data related to the spatial-temporal motion of an object are captured using the optical flow. From the differential point of view, optical flow could be represented as the change of image pixels with respect to the time and is expressed by the following equation [45].

$$I_i(x_i + \Delta x, y_i + \Delta y, t + \Delta t) = I_i(x_i, y_i, t), \quad (1)$$

where $I_i(x_i, y_i, t)$ is the intensity of the pixel at a point (x_i, y_i) at a given time t . The same is moved by $\Delta x, \Delta y, \Delta t$ in the subsequent image frame.

The Eq. 1 is further expanded by applying the Taylor Series Expansion [23] and the following equation is obtained.

$$I_i(x_i + \Delta x, y_i + \Delta y, t + \Delta t) = I_i(x_i, y_i, t) + \frac{\partial I_i}{\partial x} \Delta x + \frac{\partial I_i}{\partial y} \Delta y + \frac{\partial I_i}{\partial t} \Delta t. \quad (2)$$

From these Eqs. (1 and 2), Eq. 3 is obtained as follows:

$$\frac{\partial I_i}{\partial x} \Delta x + \frac{\partial I_i}{\partial y} \Delta y + \frac{\partial I_i}{\partial t} \Delta t = 0. \quad (3)$$

Dividing both RHS and LHS by Δt , the following equation is obtained

$$\frac{\partial I_i}{\partial x} \left(\frac{\Delta x}{\Delta t} \right) + \frac{\partial I_i}{\partial y} \left(\frac{\Delta y}{\Delta t} \right) + \frac{\partial I_i}{\partial t} \left(\frac{\Delta t}{\Delta t} \right) = 0. \quad (4)$$

A differential point of view is used here to establish the estimation of the optical flow. The variations of the pixels with respect to time is the basis of the explanation. The solution of the problem can be reduced to the following equation:

$$\frac{\partial I_i}{\partial x} v_x + \frac{\partial I_i}{\partial y} v_y + \frac{\partial I_i}{\partial t} = 0, \quad (5)$$

or

$$i_x v_x + i_y v_y + i_t = 0, \quad (6)$$

where v_x and v_y are the x and y components of the velocity or optical flow of

$$I_i(x_i, y_i, t) \quad \text{and} \quad i_x = \frac{\partial I_i}{\partial x}, i_y = \frac{\partial I_i}{\partial y}, i_t = \frac{\partial I_i}{\partial t}.$$

Equation 7 is derived from Eq. 6 as follows:

$$i_x v_x + i_y v_y = -i_t, \quad (7)$$

or

$$\Delta i \cdot \vec{v} = -i_t. \quad (8)$$

This problem is converged into finding the solution of \vec{v} . Optical flow cannot be directly estimated since there are two unknowns in the equation. This problem is known as the aperture problem. Several algorithms for estimating the optical flow have been proposed in literature. In [46], the authors reported four categories of optical flow estimation techniques, namely—differential method, region-based matching, energy-based matching and phase-based techniques.

As mentioned in the previous section, the derivatives of image intensity with respect to both space and time are used in different method. In [45], a method has proposed using the global smoothness concept of discovering the optical flow pattern which results in the Eq. (8). In [33], an image registration technique is proposed, where a good match is found using the spatial intensity gradient of the images. This is an iterative approach to find the optimum disparity vector which is a measure for finding the difference between pixel values in a particular location in two images. In [47], an algorithm is presented to compute the optical flow which

avoids the aperture problem. Here, second-order derivatives of the brightness of images are computed to generate the equations for representing optical flow.

A global method of computing the optical flow was proposed in [45, 48]. Here, an additional constraint, i.e., the smoothness of the flow is introduced as a second constraint to the basic equation (Eq. 8) for calculating the optical flow. Thereafter, the resulting equation was solved using an iterative differential approach. In [49], an integrated classical differential approach is proposed with correlation-based motion detectors. A novel method of computing optical flow using a coupled set of nonlinear diffusion equations is presented here.

In region-based matching, an affinity measure based on region features is used and applied to region tokens [50]. Thereafter, the spatial displacements among centroids of the corresponding regions are used to identify the optical flow. Region-based methods act as an alternative to the differential techniques in those fields where due to the presence of a few number of frames and background noise, differential or numerical methods are not effective [51]. This method reports the velocity, similarity, etc. between the image regions. In [52], Laplacian pyramid is used for region matching; whereas in [53], a sum of squared distance is computed for the same.

In energy-based matching methods, minimization of a global energy function is performed to determine optical flow [54]. The main component of the energy function is a data term which encourages an agreement between a spatial term and frames to enforce the consistency of the flow field. The output energy of the velocity tuned filters is the basis of the methods based on energy [55].

In phase-based techniques, the optical flow is calculated in the frequency domain by applying the local phase correlation to the frames [56]. Unlike energy-based methods, velocity is represented by the outputs of the filter exhibiting the phase behavior. In [57, 58], spatio-temporal filters are used in phase-based techniques.

(c) Histogram representation

In histogram representation, the distribution characteristics of the embedded visual features of object regions are efficiently captured. Intensity histograms are frequently used to represent target objects for visual tracking and object recognition. Mean-shift is a histogram-based methodology for visual tracking which is widely used because it is simple, fast and exhibits superior performance in real time [59]. It adopts a weighted kernel-based color histogram to compute the features of object template and regions [60]. A target candidate is iteratively moved to locate the target object from the present location p_{old}^{\wedge} to the new position p_{new}^{\wedge} based on the following relation:

$$p_{\text{new}}^{\wedge} = \frac{\sum_i k(p_i - p_{\text{old}}^{\wedge})w_s(p_i)p_i}{\sum_i k(p_i - p_{\text{old}}^{\wedge})w_s(p_i)}, \tag{9}$$

where the influence zone is defined by a radically symmetric kernel $k(\cdot)$ and the sample weight is represented by $w_s(p)$. Usually, histogram back projection is used to determine $w_s(p)$.

$$w_s(p) = \sqrt{\frac{d_m(I_c(p))}{d_c(I_c(p))}}, \tag{10}$$

where $I_c(p)$ represents the pixel color and the density estimates of the pixel colors of the target model and target candidate histograms are denoted by d_m and d_c .

Intensity histograms are widely used in the tracking algorithms [61, 62]. In object detection and tracking, efficient algorithms like integral image [63] and integral histogram [64] are effectively applied for rectangular shapes. Intensity histograms are failed to compute efficiently from region bounded by uneven shape [65]. The problem due to shape variation with respect to histogram-based tracking method is minimized using a circular or elliptical kernel [66]. The kernel is used to define a target region and a weighted histogram is computed from this. In other words, kernel brings simplicity in tracking the irregular object by enforcing a regularity constraint on it. In the above-mentioned approaches, the spatial information in histograms is not considered; however, spatial data are highly important to track a target object where significant shape variation is observed [67]. The above-mentioned issue is addressed in [68] by introducing the concept of spatiogram or spatial histogram. A spatiogram is the generalized form of a histogram where spatial means and covariance of the histogram bins are defined. Robustness in visual tracking is increased since spatial information assists in capturing richer description about the target.

In histogram models, the selected target histogram at the starting frame is compared with the candidate histograms in the subsequent frames [69] for finding the closest similar pair. The similarity among the histograms is measured by applying the Bhattacharyya Coefficient [70–72]. The similarity is represented by the following formula

$$\varphi_b(S_t) = \sum_{x,y=1}^n \sqrt{\frac{H_x}{\sum_{x=1}^n H_x} \times \frac{H_y}{\sum_{y=1}^n H_y}}, \tag{11}$$

where the target selected in the initial frame is represented by a bin H_x from the histogram, whereas H_y represents the bin corresponding to the candidate histogram. The target histogram bin index is given by x and the candidate model histogram bin index is given by y . S_t represents the generated

target state and the Bhattacharyya coefficient is represented by φ_b .

(d) Covariance representation

Visual tracking is challenging because there might be a change in appearance of the target due to the illumination changes and variations in view and pose. The above-mentioned appearance models are affected by these variations [73]. Moreover, in histogram approach, there is an exponential growth of the joint representation of various features as the number of features increases [74]. Covariance matrix representation is developed in [74] to record the correlation information of the target appearance. Covariance is used here as a Region Descriptor using the following formula:

$$I_f(p, q) = \varphi(I, p, q), \tag{12}$$

where a three-dimensional color image or a one-dimensional intensity is represented by I . I_f is the extracted feature image from I . The gradients, color, intensity, etc. mappings are represented by φ .

A $m \times m$ covariance matrix is built from the feature points which denotes the predefined rectangular region $R(R \subseteq I_f)$ by the following equation:

$$c_R = \frac{1}{n-1} \sum_{x=1}^n (g_x - \mu)(g_x - \mu)^T, \tag{13}$$

where $\{g_x\}_x = 1 \dots n$ are the m -dimensional feature points inside the region R and μ is the mean of the points.

Using covariance matrices as a region descriptor has several advantages [75]. Multiple features are combined naturally using covariance matrices without normalizing the features. The information inherent within the histogram and the information obtained from the appearance model are both represented by it. The region could be effectively matched with respect to different views and poses by extracting a single covariance matrix from it.

(e) Wavelet filtering-based representation

In wavelet transform, the features can be simultaneously located in both time and the frequency domains. The object regions are filtered out in various directions by this feature [76]. Using Gabor wavelet networks (GWN) [77, 78], a new method is proposed for visual face tracking in [79]. A wavelet representation is formed initially from the face template spanning through a low-dimensional subspace in the image space. Thereafter, the orthogonal projection of the video sequence frames corresponding to the tracked space is done into the image subspace. Thus, a subspace corresponding to the image space is efficiently defined by selectively choosing the Gabor wavelets. 2D Gabor wavelet transform is used in [80] to track an object in a video sequence. The predetermined globally placed selected feature points are used to model the target object by local features. The energy obtained from GWT coefficients of the feature points is

considered for stochastically selecting the feature points. The higher the energy values of the points, the higher is the probability of being selected. Local features are defined by the amplitude of the GWT coefficients of the selected feature point.

(f) Active contour representation

Active contour representation has been widely used in literature for tracking non-rigid objects [81–85]. The object boundary is identified by forming the object contour from a 2D image, having a probability of noisy background [86]. In [87], a signed distance map φ , which is also known as level set representation, is represented as follows:

$$\varphi(x_i, x_j) = \begin{cases} 0(x_i, x_j) & \in C \\ d(x_i, x_j, C) & \in Z_o \\ -d(x_i, x_j, C) & \in Z_{in} \end{cases}, \quad (14)$$

where the inner and outer regions of the contour are represented by Z_{in} and Z_o , respectively. The shortest Euclidian distance from the contour and the point (x_i, x_j) is calculated by the function $d(x_i, x_j, C)$.

The level set representation is widely used to form a stable numerical solution and its capability to handle the topological changes. In the same study, the evaluation of active contour methods is classified into two categories—edge based and region based. Each of these methods is briefly described in the following section.

In edge-based methods, local information about the contours (e.g., gray-level gradient) is mainly considered. In [88], a snake-based model is proposed which is one of the most widely used edge-based models. Snake model is very effective for a number of visual tracking problems for edge and line detection, subjective contour, motion tracking, stereo matching, etc. A geodesic model is proposed in [89], where more intrinsic geometric image measures are presented compared to classical snake model. The relation between the computation of the minimum distance curve or geodesics and active contours is the basis of this proposed model. In [81], an improved geodesic model is proposed. In this study, active contours are described by level sets and gradient descent method is used for contour optimization.

Edge-based algorithms [90–92] are simple and effective to determine the contours having salient gradient, but they have their drawbacks. They are susceptible to boundary leakage problems where the object has weak boundaries and sensitive to inherent image noise.

Region-based methods use statistical quantities (e.g., mean, variance and histograms based on pixel values) to segment an image into objects and background regions [93–96]. Target objects with weak boundaries or without boundaries can be successfully divided despite of the existence of image noise [97]. Region-based model is widely used in Active

contour models. In [98], an active contour model is proposed where no well-defined region boundary is present. Techniques like curve-evolution [99], Mumford–Shah function [100] for segmentation and level set [101] are used here. A region competition algorithm is proposed in [102], which is used as a statistical approach to image segmentation. A variation principle-based minimized version of a generalized Bayes/MDL (minimum description length) is used to derive the competition algorithm. A variation calculus problem for the evolution of the object contour was proposed in [103]. The problem was solved using level sets-based hybrid model combining region-based and boundary-based segmentation of the target object. Particle filter [104] is extended to a region-based image representation for video object segmentation in [105]. The particle filter is reformulated considering image partition for particle filter measurement and that results into enrichment of the existing information.

Visual Representation Based on Local Feature Visual representation using local features encodes the object appearance information using saliency detection and interest points [106]. A brief discussion on the local feature-based visual representation used in several tracking methods is given below.

In local template-based technique, an object template is continuously fitted to a sequence of video frames in template-based object tracking. Establishing a correspondence between the source image and the reference template is the objective of the template-based method [107]. Template-based visual tracking is considered as a kind of nonlinear optimization problem [108, 109]. In the presence of significant inter-frame object motion, tracking method based on nonlinear optimization has its disadvantage of being trapped in local minima. An alternative approach is proposed in [110] where geometric particle filtering is used in template-based visual tracking. A tracking method for human identification and segmentation was proposed in [111]. A hierarchical approach of part-template matching is introduced here, considering the utility of both local part-based and global template human detectors.

In segmentation-based technique, the cues are incorporated in segmentation-based visual representation of object tracking [112]. In a video sequence, segmenting the target region from the background is a challenging task. In computer graphics domain, it is known as video cutout and matting [113–116]. Two closely related problems of visual tracking are mentioned in [117]—(i) localizing the position of a target where the video has low or moderate resolution (ii) segmentation of the image of the target object where the video has moderate to high resolution. In the same study, a nonparametric k-nearest-neighbor (kNN) statistical model is used to model the dynamic changing appearance of the image regions. Both localization and

segmentation problem are solved as a sequential binary classification problem here. One of the most successful representations for image segmentation and object tracking is superpixels [118–120]. A discriminative appearance model based on superpixel to distinguish the target and the background, having mid-level cues, is proposed in [121]. A confidence map for target background is computed here to formulate the tracking task.

In scale-invariant feature transform (SIFT)-based technique [122–124], the image information is transformed into scale-invariant features which may be applied for matching different scenes or views related to the target object [125]. A set of image features are generated through SIFT by following four stages of computations namely—extrema detection, keypoint localization, orientation assignment and keypoint descriptor.

In extrema detection stage, all image locations are searched. A Gaussian difference function is used to detect the probable interesting points that remain unperturbed to the orientation and scale.

In keypoint localization stage, the location and scale are determined by fitting a detailed model at each candidate location.

In orientation assignment stage, local image gradient directions are used to assign one or more orientations to each of the key point locations. Image data are transformed based on the assigned orientation, scale and position of the each feature. All future operations are performed on the transformed image data and invariance is provided to these transformations.

In keypoint descriptor stage, the selected scale is used to measure the local image gradients in the region surrounding each keypoint. A significant amount of local distortion in shape and illumination changes is allowed in the transformed representation.

SIFT-based techniques have its wide use in literature because of its invariance to the scene background change during the tracking. A real-time, low-power system based on SIFT algorithm was proposed in [126]. A database of the features of the known objects is maintained and the individual features are matched with it. A modified version of the approximation of nearest neighbor search algorithm based on the K-d tree and BBF algorithm is used here. SIFT- and PowerPC-based infrared (IR) imaging system is used in [127] to automatically recognize the target object in unknown environments. First, the positional interest points and scale are localized for a moving object. Thereafter, the description of the interest points is built. SIFT and Kalman filter are used in [128] to handle occlusion. In an image sequence, the objects are identified using the SIFT algorithm with the help of the extracted invariant features. The presence of occlusion degrades the accuracy of SIFT. Kalman filter [129] is used here to minimize the effect of occlusion

because the estimation of the location of the object in the subsequent frame is done based on the location information about the object in the previous frame.

Saliency detection-based method is applicable to individual images if there is a presence of a well-centered single salient object [130]. Two stages of saliency detection are mentioned in the literature [131]. The first stage involves the detection of the most prominent object and the accurate region of the object is segmented in the second stage. These two stages are rarely separated in practice rather they are often overlapped [132, 133]. In [134], a novel method of real-time extraction of saliency features from the video frames is proposed. Conditional random fields (CRF) [135] are combined with the saliency features and thereafter, a particle filter is applied to track the detected object. In [136], the mean-shift tracker in combination with saliency detection is used for object tracking in dynamic scenes. To minimize the interference of the complex background, first a spatial-temporal saliency feature extraction method is proposed. Furthermore, the tracking performance is enhanced by fusing the top-down visual mechanism in the saliency evaluation method. A novel method of detecting the salient object in images is proposed in [137], where the variability is computed statistically by two scatter matrices to measure the variability between the central and the surrounding objects. The pixel centric most salient regions are defined as a salience support region. The saliency of pixel is estimated through its saliency support region to detect variable-sized multiple salient objects in a scene.

Statistical Modeling

The visual tracking methods are continuously subjected to inevitable appearance changes. In statistical modeling, the object detection is performed dynamically [138]. Variations in shape, texture and the correlations between them are represented by the statistical model [139]. A statistical model is categorized into three classes [140] namely—generative model, discriminative model and hybrid model.

In visual tracking, the appearance templates are adaptively generated and updated by the generative model [141, 142]. The appearance model of the target is adaptively updated by the online learning strategy embedded in the tracking framework [143].

A framework based on an online EM algorithm to model the change in appearance during tracking is proposed in [144]. In the presence of image outliers, this model provides robustness when used in a motion-based tracking algorithm. In [145, 146] Adaptive Appearance model is incorporated in a particle filter to realize robust visual tracking. An online learning algorithm is proposed in [147] to generate an image-based representation of the video sequences for visual tracking. A probabilistic appearance manifold [148] is

constructed here from a generic prior and a video sequence of the object. An adaptive subspace representation of the target object is proposed in [149], where low-dimensional subspace is incrementally learned and updated. A compact representation of the target is provided here instead of representing the same as a set of independent pixels. Appearance changes due to internal or external factors are reflected since the subspace model is continuously updated by the incremental method. In [35], an incremental tensor subspace learning-based algorithm is proposed for visual tracking. The appearance changes of the target are represented by the algorithm through online learning of a low-dimensional eigenspace representation. In [150], Retinex algorithm [151, 152] is combined with the original image and the resultant is defined as weighted tensor subspace (WTS). WTS is adapted to the target appearance changes by an incremental learning algorithm. In [153], a robust tracking algorithm is proposed to combine sparse appearance models and adaptive template update strategy, which is less sensitive to occlusion. A weighted structural local sparse appearance model is adopted in [154], which combines patch-based gray value and histogram-oriented gradient features for the patch dictionary.

Tracking is defined as a classification problem in discriminative methods [155]. The target is discriminated from the background and updated online. Appearance and environmental changes are handled by a binary classifier which is trained to filter out the target from the background [156, 157]. As this method applies a discriminatively trained detector for tracking purposes, this is also called tracking by detection mechanism [158–162]. Discriminative methods pertain machine learning approaches to distinguish between the object and non-object [163]. To achieve constructive prophetic performances, online variants are proposed to progressively learn discriminative classification features for distinguishing object and non-object. The main problem is a discriminative feature (for, e.g., color, texture, shape, etc.) may be identical along with the varying background [164]. In [165], a discriminative correlation filter-based (DCF) approach is proposed which is used to evaluate the object in the next frame. Hand-crafted appearance features such as HOG [166], color name feature [167] or a combination of both [168] are usually utilized by DCF-based trackers. To remove ambiguity, a deep motion feature is used which differentiates the target based on discriminative motion pattern and leads to successful tracking after occlusion, addressed in [169]. A discriminative scale space tracking approach (DSST), which learns separate discriminative correlation filters for explicit translation and scale evaluation, is proposed in [170]. A support vector machine (SVM) tracking framework and dictionary learning based on discriminative appearance model are reported in [171]. To track arbitrary

object in videos, a real-time, online tracking algorithm is proposed based on discriminative model [172].

The generative and discriminative models have complementary strengths and weaknesses, though they have different characteristics. A combination of generative and discriminative model to get the best practices of both domains is proposed in [172]. A new hybrid model is proposed here to classify weakly labeled training data. A multi-conditional learning framework [173] is proposed in [174] for simultaneously clustering, classifying and dimensionality reduction. Favorable properties of both the models are observed in the multi-conditional learning model. In the same study, it is demonstrated that a generalized superior performance is achieved using the hybrid model of the foreground or background pixel classification problem [175].

From the appearance model, stable properties of appearance are identified and motion estimation is done by weighing on them [144]. Next section elaborates briefly about the motion estimation methodologies mentioned in the literatures.

Motion estimation

In motion estimation, motion vectors [176–180] are determined to represent the transformation through adjacent 2D image frames in a video sequence [181]. Motion vectors are computed in two ways [182]—pixel-based methods or direct method, and feature-based methods or indirect method. In direct methods [183], motion parameters are estimated directly by measuring the contribution of each pixel that results in optimal usage of the available information and image alignment. In indirect methods, features like corner detection are used and the corresponding features between the frames are matched with a statistical function applied over a local or global area [184]. Image areas are identified where a good correspondence is achievable and computation is concentrated in these areas. The initial estimation of the camera geometry is, thus, obtained. The correspondence of the image regions having less information is guided by this geometry.

In visual tracking, motion can be modeled using a particle filter [140] which is considered as a dynamic state estimation problem. Let the parameters for describing the affine motion of an object is represented by m_t and the subsequent observation vectors denoted by o_t . The following two rules are recursively applied to estimate the posterior probability

$$p(m_t|o_{1:t-1}) = \int p(m_t|m_{t-1})p(m_{t-1}|o_{1:t-1})dm_{t-1}, \quad (15)$$

$$p(m_t|o_{1:t}) = \frac{p(o_t|m_t)p(m_t|o_{1:t-1})}{P(o_t|o_{1:t-1})}, \tag{16}$$

where $m_{1:t} = \{m_1, m_2, \dots, m_t\}$ represents state vectors at time t and $o_{1:t} = \{o_1, o_2, \dots, o_t\}$ represents the corresponding observatory states.

The motion model describes the transition of states between the subsequent frames and is denoted by $p(m_t|m_{t-1})$. The observation model is denoted by $p(o_t|m_t)$ which calculates the probability of an observed image frame to be in a particular object class.

Object Localization

The target location is estimated in subsequent frames by the motion estimation process. The target localization or positioning operation is performed by maximum posterior prediction or greedy search, based on motion estimation [185].

A brief description about visual tracking and the associated models is given in the above section. Visual tracking is one of the rapidly growing fields in computer vision. Numerous algorithms are proposed in literature every year. Several measures to evaluate the visual tracking algorithms are briefly described in the below section.

Visual Tracking Performance

The performance measures represent the difference or correspondence between the predicted and actual ground truth annotations. Several performance measures, widely used in visual tracking [186, 187] are—center location error, bounding box overlaps, tracking length, failure rate, area under the lost-track-ratio curve, etc. A brief description of each of these measures is given below.

Center Location Error

The center location error is one of the widely used measures for evaluating the performance of object tracking. The difference between the center of the manually marked ground truth position (r_t^G) and the tracked target's center (r_t^T) is computed by computing the Euclidean distance between them [188]. The same is formulated as follows.

In a sequence of length n , the state description of the object (φ) is given by:

$$\varphi = \left\{ (r_t, c_t) \right\}_{t=1}^n, \tag{17}$$

where the center of the object is denoted by $r_t \in \mathcal{R}^2$ and r_t represents the object region at time t .

The central error (E_c) is formulated as follows:

$$E_c(\varphi^G, \varphi^T) = \sqrt{\frac{1}{n} \sum_{t=1}^n |r_t^G - r_t^T|^2}. \tag{18}$$

Randomness of the output location is frequent when the track of a target object is lost by the tracking algorithm. In such a scenario, it is difficult to measure the accurate tracking performance [188]. The error due to randomness is minimized in [163] where a threshold distance is maintained from the ground truth object and the percentage of frames within this threshold is calculated to estimate the tracking accuracy.

Bounding Box Overlap

In central location error, the pixel difference is measured, but the scale and size of the target object are not reflected [163]. A popular evaluation metric that minimizes the limitation of the central location error is the overlapping score [189, 190]. The overlap of the ground truth region and the predicted target's region is considered as overlap score (S_r) and the same is formulated as below [191].

$$S_r = \frac{\text{Area}(r_t^G \cap r_t^T)}{\text{Area}(r_t^G \cup r_t^T)}, \tag{19}$$

where \cup and \cap represent the union and intersection of two boundary region boxes and the region area is represented by the function Area().

Both position and size of the bounding boxes of ground truth object and predicted target are considered here and as a result, the significant errors due to tracking failures are minimized.

Tracking Length

Tracking length is a measure which is used in literature [192, 193]; it denotes the number of frames successfully tracked from the initialization of the tracker until its first failure. The tracker's failure cases are explicitly addressed here but it is not effective in the presence of a difficult tracking condition at the initialization of the video sequence.

Failure Rate

The problem of tracking length is addressed in the failure rate measure [194, 195]. This is a supervised system where the tracker is reinitialized by a human operator once it suffers failure. The system records the number of manual interventions and the same is used as a comparative performance score. The entire video sequence is considered in the performance evaluation and hence, the dependency of the beginning part, unlike the tracking length measure, is diminished.

Area Under the Lost-Track-Ratio Curve

In [196], a hybrid measure is proposed where several measures are combined into a single measure. Based on the overlap measure (S_r) which is described in the earlier section, the lost-track ratio γ is computed. In a particular frame, the track is considered to be lost when overlap between the ground truth and the estimated target is smaller than a certain threshold value (β), i.e., $S_r \leq \beta$, where $\beta \in (0, 1)$

Lost-track ratio is represented by the following formula:

$$\gamma = \frac{F_t}{F}, \quad (20)$$

where F_t is the number of frames having a lost track and F is the total number frames belonging to the estimated target trajectory.

The area under the lost-track (AULT) is formulated as below:

$$\text{AULT} = \Delta\beta \sum_{\beta=0}^1 \gamma(\beta). \quad (21)$$

In this method a compact measure is presented where a tracker has to take into account two separate tracking aspects.

Visual tracking has its wide application in the literature. Some of the application areas of visual tracking are briefly described in the below section.

Applications of Visual Tracking

Different methods of visual tracking are used in a wide range of application domains. This section is mainly focused around seven application domains of visual tracking—Medical Science, Space Science, Augmented Reality Applications, Posture estimation, Robotics, Education, Sports, Cinematography, Business and Marketing, and Deep Learning Features.

Medical Science

To improve the robot-assisted laparoscopic surgery system, a human machine interface is presented for instrument localization and automated endoscope manipulation [197, 198]. An “Eye Mouse” based on a low-cost tracking system is implemented in [199], which is used to manipulate computer access for people with drastic disabilities. The study of discrimination between bipolar and schizophrenic disorders by using visual motion processing impairment is found in [200]. Three different applications for analyzing

the classification rate and accuracy of the tracking system, namely the control of the mobile robot in the maze, the text writing program “EyeWriter” and the computer game, were observed in [201]. A non-invasive, robust visual tracking method for pupils identification in video sequences captured by low-cost equipment is addressed in [202]. A detailed discussion of eye tracking application in medical science is described in [203].

Space Science

A visual tracking approach based on color is proposed in [204, 205] for astronauts, which presents a numeric analysis of accuracy on a spectrum of astronaut profiles. A sensitivity-based differential Earth mover’s distance (DEMD) algorithm of simplex approach is illustrated and empirically substantiated in the visual tracking context [206]. In [207], an object detection and tracking based on background subtraction, optical flow and CAMShift algorithm is presented to track unusual events successfully in video taken by UAV. A visual tracking algorithm based on deep learning and probabilistic model to form Personal Satellite for tracking the astronauts of the space stations in RGB-D videos, reported in [208].

Augmented Reality (AR) Applications

Augmented reality system on color-based and feature-based visual tracking is implemented on a series of applications such as Sixth Sense [209], Markerless vision-based tracking [210], Asiatic skin segmentation [211], Parallel Tracking and Mapping (PTAM) [212], construction site visualization [213], Face augmentation system [214, 215], etc., reported in [216]. A fully mobile hybrid AR system which combines a vision-based trackers with an inertial tracker to develop energy efficient applications for urban environments is proposed in [217]. An image-based localization of mobile devices using an offline data acquisition is reported in [218]. A robust visual tracking AR system for urban environments by utilizing appearance-based line detection and textured 3D models is addressed in [219].

Posture Estimation

This application domain deals with the images involving humans, which covers facial tracking, hand gesture identification, and the whole-body movement tracing. A model-based non-invasive visual hand tracking system, named ‘DigitEyes’ for high DOF articulated mechanisms, is described in [220]. The three main approaches

for analyzing human gesture and whole-body tracking, namely 2D perspective without explicit shape models, 2D perspective with explicit shape models and 3D outlook, were discussed in [221]. A kinematic real-time model for hand tracking and pose evaluation is proposed to lead a robotic arm in gripping gestures [222]. A 3D LKT algorithm based on model for evaluating 3D head postures from discrete 2D visual frames is proposed in [223].

Robotics

A real-time system for ego-motion estimation on autonomous ground vehicles with stereo cameras using feature detection algorithm is illustrated in [224]. A visual navigation system is proposed in [225] which can be applied to all kinds of robots. In this paper, the authors categorized and illustrated the visual navigation techniques majorly into map-based navigation [226] and mapless navigation [227]. The motionCUT framework is presented in [228] to detect motion in visual scenes generated by moving cameras and the said technique is applied on the humanoid robot iCub for experimental validation. A vision-based tracking methodology using a stereoscopic vision system for mobile robots is introduced in [229].

Education

Visual tracking technology is widely applicable in the field of educational research. To increase the robustness of the visual prompting for a remedial reading system that helps the end users with identification and pronunciation of terms, a reading assistant is presented in [230]. To implement the said system, a GWGazer system is proposed which combines two different methods, namely interaction technique evaluation [231–233] and observational research [234–236]. An ESA (empathic software agent) interface using real-time visual tracking to ease empathetic pertinent behavior is applicable in the virtual education environment within a learning community, reported in [237]. An effective approach towards students' visual attention tracking using an eye tracking methodology to solve multiple choice type problems is addressed in [238]. An information encapsulating process of teacher's consciousness towards the student's requirement using visual tracking is presented in [239], which is beneficial for classroom management system. To facilitate computer educational research using eye tracking methods, a gaze estimation methodology is proposed, which keeps record of a person's visual behavior, reported in [240]. A realistic solution of mathematics teaching based on visual tracking is addressed in [241]. A detailed study of visual tracking in computer programming is described in [242].

Sports

Visual tracking holds a strong application field towards Sports. There are several approaches under this domain using different models of visual tracking. The precise tracking of the golfer during a conventional golf swing using dynamic modeling is presented in [243]. A re-sampling and re-weighting particle filter method is proposed to track overlapping athletes in a beach volleyball or football sequence using a single camera, reported in [244]. Improvement in performance of the underwater hockey athletes has been addressed in [245] by inspecting their vision behavior during breath holding exercise and eye tracking. A detailed discussion in this domain is presented in [246, 247].

Apart from these, visual tracking can be broadly used in the field of cinematography [248–251], cranes systems [252, 253], business and marketing [254–260] and deep learning applications [260–265].

Discussion

The traditional visual tracking methods perform competently in well-controlled environments. The image representations used by the trackers may not be sufficient for accurate robust tracking in complex environments. Moreover, the visual tracking problem becomes more challenging due to the presence of occlusion, un-organized background, abrupt fast random motion, dramatic changes in illumination, and significant changes in pose and viewpoints.

Support vector machine (SVM) classifier was fused with optical flow-based tracker in [266] for visual tracking. The classifier helps to detect the location of the object in the next frame even though a certain part of the object is missing. In this method, the next frame is not only matched with the previous frame, but also against all possible patterns learned by the classifier. More precise bounding boxes are identified in [267] using a joint classification–regression random forest model. Here, authors demonstrated that the aspect ratio of the variable bounding boxes was accurately predicted by this model. In [268], a neural network-based tracking system was proposed to describe a collection of tracking structures that enhance the effectiveness and adaptability of a visual tracker. Multinetwork architectures are used here that increase the accuracy and stability of visual tracking.

An extensive bibliographic study has been carried out based on the previously published works listed in Scopus database for the period of last 5 years (2014–2018). Amongst 2453 listed works, 48.9% articles were published in journals and 44.5% in conferences. It is observed that major contributions in this area are from computer science engineers (42%). Medical science and related domains (6%) also have notable contribution in this arena. The leading contributors

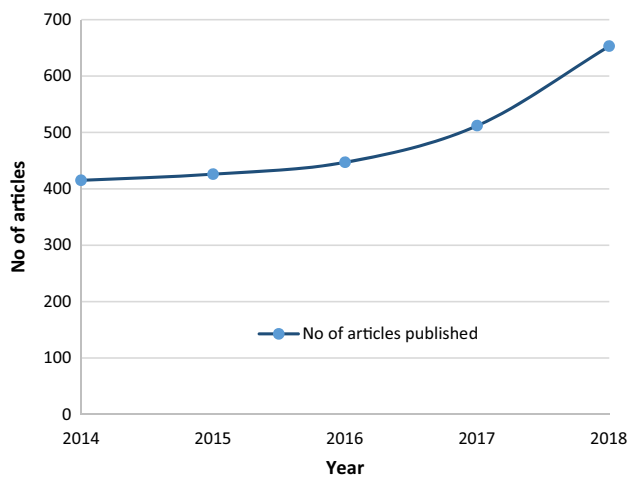


Fig. 2 Trends of visual tracking research

are from countries like China (57%), USA (12%), UK (5%), etc. Figure 2 clearly depicts the increasing interest in vision tracking in the last few years.

The above study clearly shows that, in recent years with the advent of deep learning, the challenging problem to track a moving object with a complex background has made significant progress [269]. Unlike previous trackers, more emphasis is put on unsupervised feature learning. A noteworthy performance improvement in visual tracking is observed with the introduction of deep neural networks (DNN) [269, 270] and convolutional neural networks (CNN) [271–275]. DNN, especially CNN, demonstrate a strong efficiency in learning feature representations from huge annotated visual data unlike handcrafted features. High-level rich semantic information is carried out by the object classes which assist in categorizing the objects. These features are also tolerant to data corruption. A significant improvement in accuracy is observed in object and saliency detection besides image classification due to the combination of CNNs with the traditional trackers.

Conclusion

An overall study on visual tracking and its performance measures is presented in this study. Object initialization is the first stage of visual tracking. Initialization could be manual or automatic. The object properties like appearance, velocity, location, etc. are represented by observation model or appearance model. Special features like color, gradient, texture, shape, super-pixel, depth, motion, optical flow, etc. are used for robust visual tracking, that describe the appearance model. Appearance modeling consists of visual representation and statistical modeling. In visual representation, various visual features are used to form robust object

descriptors; whereas in statistical modeling, a mathematical model for identifying the target object is developed. In the last few decades, a huge number of visual tracking algorithms are proposed in the literature. A comprehensive review of different measures to evaluate the tracking algorithms is presented in this study. Visual tracking is applied in a wide range of applications including medical science, space science, robotics, education, sports, etc. Some of the application areas of visual tracking and related studies in the literature are presented here.

Compliance with ethical standards

Conflict of interest On behalf of all authors, the corresponding author states that there is no conflict of interest.

References

1. Sun Y, Meng MQH. Multiple moving objects tracking for automated visual surveillance. In: 2015 IEEE international conference on information and automation. 2015; IEEE. pp. 1617–1621.
2. Wei W, Yunxiao A. Vision-based human motion recognition: a survey. In: 2009 Second international conference on intelligent networks and intelligent systems. IEEE; 2009. pp. 386–389.
3. Zha ZJ, Wang M, Zheng YT, Yang Y, Hong R, Chua TS. Interactive video indexing with statistical active learning. *IEEE Trans Multimed.* 2012;14(1):17–27.
4. Ying S, Yang Y. Study on vehicle navigation system with real-time traffic information. In: 2008 International conference on computer science and software engineering. vol. 4. IEEE; 2008. pp. 1079–1082.
5. Huang K, Petkovsek S, Poudel B, Ning T. A human-computer interface design using automatic gaze tracking. In: 2012 IEEE 11th international conference on signal processing. vol. 3. IEEE; 2012. pp. 1633–1636.
6. Alenljung B, Lindblom J, Andreasson R, Ziemke T. User experience in social human-robot interaction. In: *Rapid automation: concepts, methodologies, tools, and applications.* IGI Global; 2019. pp. 1468–1490.
7. Chincholkar AA, Bhojar MSA, Dagwar MSN. Moving object tracking and detection in videos using MATLAB: a review. *Int J Adv Res Comput Electron.* 2014;1(5):2348–5523.
8. Abdelkader MF, Chellappa R, Zheng Q, Chan AL. Integrated motion detection and tracking for visual surveillance. In: *Fourth IEEE International Conference on Computer Vision Systems (ICVS'06).* IEEE; 2006. p. 28.
9. Courtney JD. Automatic video indexing via object motion analysis. *Pattern Recogn.* 1997;30(4):607–25.
10. Chae KH, Moon YS, Ko NY. Visual tracking of objects for unmanned surface vehicle navigation. In: 2016 16th International Conference on Control, Automation and Systems (ICCAS). IEEE; 2016. pp. 335–337.
11. Phung MD, Tran QV, Hara K, Inagaki H, Abe M. Easy-setup eye movement recording system for human-computer interaction. In: 2008 IEEE international conference on research, innovation and vision for the future in computing and communication technologies. 2008; IEEE. pp. 292–297.

12. Kavya R. Feature extraction technique for robust and fast visual tracking: a typical review. *Int J Emerg Eng Res Technol*. 2015;3(1):98–104.
13. Kang B, Liang D, Yang Z. Robust visual tracking via global context regularized locality-constrained linear coding. *Optik*. 2019;183:232–40.
14. Yilmaz A, Javed O, Shah M. Object tracking: a survey. *Acm Comput Surv (CSUR)*. 2006;38(4):13.
15. Jalal, A. S., & Singh, V. (2012). The state-of-the-art in visual object tracking. *Informatica*, 36(3).
16. Li X, Hu W, Shen C, Zhang Z, Dick A, Hengel AVD. A survey of appearance models in visual object tracking. *ACM Trans Intell Syst Technol (TIST)*. 2013;4(4):58.
17. Anuradha K, Anand V, Raajan NR. Identification of human actor in various scenarios by applying background modeling. *Multimed Tools Appl*. 2019. <https://doi.org/10.1007/s11042-019-7443-5>.
18. Sghaier S, Farhat W, Souani C. Novel technique for 3D face recognition using anthropometric methodology. *Int J Ambient Comput Intell (IJACI)*. 2018;9(1):60–77.
19. Zhang Y, Xu X, Liu X. Robust and high performance face detector. *arXiv preprint arXiv:1901.02350*. 2019.
20. Surekha B, Nazare KJ, Raju SV, Dey N. Attendance recording system using partial face recognition algorithm. In: *Intelligent techniques in signal processing for multimedia security*. Springer, Cham; 2017. pp. 293–319.
21. Chaki J, Dey N, Shi F, Sherratt RS. Pattern mining approaches used in sensor-based biometric recognition: a review. *IEEE Sens J*. 2019;19(10):3569–80.
22. Dey N, Mukherjee A. *Embedded systems and robotics with open source tools*. USA: CRC Press; 2018.
23. Shell HSM, Arora V, Dutta A, Behera L. Face feature tracking with automatic initialization and failure recovery. In: 2010 IEEE conference on cybernetics and intelligent systems. IEEE; 2010. pp. 96–101.
24. Schmidt J. Automatic initialization for body tracking using appearance to learn a model for tracking human upper body motions. 2008.
25. Fan L, Wang Z, Cail B, Tao C, Zhang Z, Wang Y et al. A survey on multiple object tracking algorithm. In: 2016 IEEE international conference on information and automation (ICIA). IEEE; 2016. pp. 1855–1862.
26. Liu S, Feng Y. Real-time fast moving object tracking in severely degraded videos captured by unmanned aerial vehicle. *Int J Adv Rob Syst*. 2018;15(1):1729881418759108.
27. Lu J, Li H. The Importance of Feature Representation for Visual Tracking Systems with Discriminative Methods. In: 2015 7th International conference on intelligent human-machine systems and cybernetics. vol. 2. IEEE; 2015. pp. 190–193.
28. Saleemi I, Hartung L, Shah M. Scene understanding by statistical modeling of motion patterns. In: 2010 IEEE computer society conference on computer vision and pattern recognition. IEEE; 2010. pp. 2069–2076.
29. Zhang K, Liu Q, Yang J, Yang MH. Visual tracking via Boolean map representations. *Pattern Recogn*. 2018;81:147–60.
30. Ernst D, Marée R, Wehenkel L. Reinforcement learning with raw image pixels as input state. In: *Advances in machine vision, image processing, and pattern analysis*. Springer, Berlin; 2006. pp. 446–454.
31. Sahu DK, Jawahar CV. Unsupervised feature learning for optical character recognition. In: 2015 13th International conference on document analysis and recognition (ICDAR). IEEE; 2015. pp. 1041–1045.
32. Silveira G, Malis E. Real-time visual tracking under arbitrary illumination changes. In: 2007 IEEE conference on computer vision and pattern recognition. IEEE; 2007. pp. 1–6.
33. Lucas BD, Kanade T. An iterative image registration technique with an application to stereo vision. 1981.
34. Ho J, Lee KC, Yang MH, Kriegman D. Visual tracking using learned linear subspaces. In: *CVPR (1)*. 2004. pp. 782–789.
35. Li X, Hu W, Zhang Z, Zhang X, Luo G. Robust visual tracking based on incremental tensor subspace learning. In: 2007 IEEE 11th international conference on computer vision. IEEE; 2007. pp. 1–8.
36. Wen J, Li X, Gao X, Tao D. Incremental learning of weighted tensor subspace for visual tracking. In: 2009 IEEE international conference on systems, man and cybernetics. IEEE; 2009. pp. 3688–3693.
37. Hu W, Li X, Zhang X, Shi X, Maybank S, Zhang Z. Incremental tensor subspace learning and its applications to foreground segmentation and tracking. *Int J Comput Vis*. 2011;91(3):303–27.
38. Yang S, Xie Y, Li P, Wen H, Luo H, He Z. Visual object tracking robust to illumination variation based on hyperline clustering. *Information*. 2019;10(1):26.
39. Dey N. Uneven illumination correction of digital images: a survey of the state-of-the-art. *Optik*. 2019;183:483–95.
40. Wang T, Gu IY, Shi P. Object tracking using incremental 2D-PCA learning and ML estimation. In: 2007 IEEE international conference on acoustics, speech and signal processing-ICASSP'07. vol. 1. IEEE; 2007. pp. 1–933.
41. Li X, Hu W, Zhang Z, Zhang X, Zhu M, Cheng J. Visual tracking via incremental log-euclideanriemannian subspace learning. In: 2008 IEEE conference on computer vision and pattern recognition. IEEE; 2008. pp. 1–8.
42. Wang H, Suter D, Schindler K, Shen C. Adaptive object tracking based on an effective appearance filter. *IEEE Trans Pattern Anal Mach Intell*. 2007;29(9):1661–7.
43. Allili MS, Ziou D. Object of interest segmentation and tracking by using feature selection and active contours. In: 2007 IEEE conference on computer vision and pattern recognition. IEEE; 2007. pp. 1–8.
44. Akpinar S, Alpaslan FN. Video action recognition using an optical flow based representation. In: *Proceedings of the international conference on image processing, computer vision, and pattern recognition (IPCV) (p. 1)*. The Steering Committee of the World Congress in Computer Science, Computer Engineering and Applied Computing (WorldComp). 2014.
45. Horn BK, Schunck BG. Determining optical flow. *Artif Intell*. 1981;17(1–3):185–203.
46. Barron JL, Fleet DJ, Beauchemin SS. Performance of optical flow techniques. *Int J Comput Vis*. 1994;12(1):43–77.
47. Uras S, Girogi F, Verri A, Torre V. A computational approach to motion perception. *Biol Cybern*. 1988;60(2):79–87.
48. Camus T. Real-time quantized optical flow. *Real-Time Imaging*. 1997;3(2):71–86.
49. Proesmans M, Van Gool L, Pauwels E, Oosterlinck A. Determination of optical flow and its discontinuities using non-linear diffusion. In: *European Conference on Computer Vision*. Springer, Berlin; 1994. pp. 294–304.
50. Fuh CS, Maragos P. Region-based optical flow estimation. In: *Proceedings CVPR'89: IEEE computer society conference on computer vision and pattern recognition*. IEEE; 1989. pp. 130–135.
51. O'Donovan P. Optical flow: techniques and applications. *Int J Comput Vis*. 2005;1–26.
52. Anandan P. A computational framework and an algorithm for the measurement of visual motion. *Int J Comput Vis*. 1989;2(3):283–310.
53. Singh A. An estimation-theoretic framework for image-flow computation. In: *Proceedings third international conference on computer vision*. IEEE; 1990. pp. 168–177.

54. Li Y, Huttenlocher DP. Learning for optical flow using stochastic optimization. In: European conference on computer vision. Springer, Berlin; 2008. pp. 379–391.
55. Barniv Y. Velocity filtering applied to optical flow calculations. 1990.
56. Argyriou V. Asymmetric bilateral phase correlation for optical flow estimation in the frequency domain. arXiv preprint [arXiv:1811.00327](https://arxiv.org/abs/1811.00327). 2018.
57. Buxton BF, Buxton H. Computation of optic flow from the motion of edge features in image sequences. *Image Vis Comput*. 1984;2(2):59–75.
58. Fleet DJ, Jepson AD. Computation of component image velocity from local phase information. *Int J Comput Vis*. 1990;5(1):77–104.
59. Lee JY, Yu W. Visual tracking by partition-based histogram backprojection and maximum support criteria. In: 2011 IEEE International Conference on Robotics and Biomimetics. IEEE; 2011. pp. 2860–2865.
60. Zhi-Qiang H, Xiang L, Wang-Sheng Y, Wu L, An-Qi H. Mean-shift tracking algorithm with improved background-weighted histogram. In: 2014 Fifth international conference on intelligent systems design and engineering applications. IEEE; 2014. pp. 597–602.
61. Birchfield S. Elliptical head tracking using intensity gradients and color histograms. In: Proceedings. 1998 IEEE Computer Society conference on computer vision and pattern recognition (Cat. No. 98CB36231). IEEE; 1998. pp. 232–237.
62. Comaniciu D, Ramesh V, Meer P. Real-time tracking of non-rigid objects using mean shift. In: Proceedings IEEE conference on computer vision and pattern recognition. CVPR 2000 (Cat. No. PR00662). vol. 2. IEEE; 2000. pp. 142–149.
63. Viola P, Jones M. Rapid object detection using a boosted cascade of simple features. *CVPR*. 2001;1(1):511–8.
64. Porikli F. Integral histogram: a fast way to extract histograms in cartesian spaces. In: 2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05). Vol. 1. IEEE; 2005. pp. 829–836.
65. Parameswaran V, Ramesh V, Zoghlami I. Tunable kernels for tracking. In: 2006 IEEE computer society conference on computer vision and pattern recognition (CVPR'06). Vol. 2. IEEE; 2006. pp. 2179–2186.
66. Fan Z, Yang M, Wu Y, Hua G, Yu T. Efficient optimal kernel placement for reliable visual tracking. In: 2006 IEEE computer society conference on computer vision and pattern recognition (CVPR'06). Vol. 1. IEEE; 2006. pp. 658–665.
67. Nejhum SS, Ho J, Yang MH. Visual tracking with histograms and articulating blocks. In: 2008 IEEE conference on computer vision and pattern recognition. IEEE; 2008. pp. 1–8.
68. Birchfield ST, Rangarajan S. Spatiograms versus histograms for region-based tracking. In: 2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05). Vol. 2. IEEE; 2005. pp. 1158–1163.
69. Zhao A. Robust histogram-based object tracking in image sequences. In: 9th Biennial conference of the Australian pattern recognition society on digital image computing techniques and applications (DICTA 2007). IEEE; 2007. pp. 45–52.
70. Djouadi A, Snorrason O, Garber FD. The quality of training sample estimates of the bhattacharyya coefficient. *IEEE Trans Pattern Anal Mach Intell*. 1990;12(1):92–7.
71. Kailath T. The divergence and Bhattacharyya distance measures in signal selection. *IEEE Trans Commun Technol*. 1967;15(1):52–60.
72. Aherne FJ, Thacker NA, Rockett PI. The Bhattacharyya metric as an absolute similarity measure for frequency coded data. *Kybernetika*. 1998;34(4):363–8.
73. Wu Y, Wang J, Lu H. Real-time visual tracking via incremental covariance model update on Log-Euclidean Riemannian manifold. In: 2009 Chinese conference on pattern recognition. IEEE; pp. 1–5.
74. Tuzel O, Porikli F, Meer P. Region covariance: a fast descriptor for detection and classification. In: European conference on computer vision. Springer, Berlin; 2006. pp. 589–600.
75. Porikli F, Tuzel O, Meer P. Covariance tracking using model update based on lie algebra. In: 2006 IEEE computer society conference on computer vision and pattern recognition (CVPR'06). Vol. 1. IEEE; 2006. pp. 728–735.
76. Duflot LA, Reisenhofer R, Tamadazte B, Andreff N, Krupa A. Wavelet and shearlet-based image representations for visual servoing. *Int J Robot Res*. 2018; 0278364918769739.
77. Krueger V, Sommer G. Efficient head pose estimation with Gabor wavelet networks. In: BMVC. pp. 1–10.
78. Krüger V, Sommer G. Gabor wavelet networks for object representation. In: Multi-image analysis. Springer, Berlin; 2001. pp. 115–128.
79. Feris RS, Krueger V, Cesar RM Jr. A wavelet subspace method for real-time face tracking. *Real-Time Imaging*. 2004;10(6):339–50.
80. He C, Zheng YF, Ahalt SC. Object tracking using the Gabor wavelet transform and the golden section algorithm. *IEEE Trans Multimed*. 2002;4(4):528–38.
81. Paragios N, Deriche R. Geodesic active contours and level sets for the detection and tracking of moving objects. *IEEE Trans Pattern Anal Mach Intell*. 2000;22(3):266–80.
82. Cremers D. Dynamical statistical shape priors for level set-based tracking. *IEEE Trans Pattern Anal Mach Intell*. 2006;28(8):1262–73.
83. Allili MS, Ziou D. Object of interest segmentation and tracking by using feature selection and active contours. In: 2007 IEEE conference on computer vision and pattern recognition. IEEE; 2007. pp. 1–8.
84. Vaswani N, Rathi Y, Yezzi A, Tannenbaum A. Pf-mt with an interpolation effective basis for tracking local contour deformations. *IEEE Trans. Image Process*. 2008;19(4):841–57.
85. Sun X, Yao H, Zhang S. A novel supervised level set method for non-rigid object tracking. In: CVPR 2011. IEEE; 2011. pp. 3393–3400.
86. Musavi SHA, Chowdhry BS, Bhatti J. Object tracking based on active contour modeling. In: 2014 4th International conference on wireless communications, vehicular technology, information theory and aerospace and electronic systems (VITAE). IEEE; 2014. pp. 1–5.
87. Hu W, Zhou X, Li W, Luo W, Zhang X, Maybank S. Active contour-based visual tracking by integrating colors, shapes, and motions. *IEEE Trans Image Process*. 2013;22(5):1778–92.
88. Kass M, Witkin A, Terzopoulos D. Snakes: active contour models. *Int J Comput Vis*. 1988;1(4):321–31.
89. Caselles V, Kimmel R, Sapiro G. Geodesic active contours. *Int J Comput Vis*. 1997;22(1):61–79.
90. Hore S, Chakraborty S, Chatterjee S, Dey N, Ashour AS, Van Chung L, Le DN. An integrated interactive technique for image segmentation using stack based seeded region growing and thresholding. *Int J Electr Comput Eng*. 2016;6(6):2088–8708.
91. Ashour AS, Samanta S, Dey N, Kausar N, Abdessalemkaraa WB, Hassanien AE. Computed tomography image enhancement using cuckoo search: a log transform based approach. *J Signal Inf Process*. 2015;6(03):244.
92. Araki T, Ikeda N, Dey N, Acharjee S, Molinari F, Saba L, et al. Shape-based approach for coronary calcium lesion volume measurement on intravascular ultrasound imaging and its association with carotid intima-media thickness. *J Ultrasound Med*. 2015;34(3):469–82.

93. Tuan TM, Fujita H, Dey N, Ashour AS, Ngoc VTN, Chu DT. Dental diagnosis from X-ray images: an expert system based on fuzzy computing. *Biomed Signal Process Control*. 2018;39:64–73.
94. Samantaa S, Dey N, Das P, Acharjee S, Chaudhuri SS. Multilevel threshold based gray scale image segmentation using cuckoo search. *arXiv preprint arXiv:1307.0277*. 2013.
95. Rajinikanth V, Dey N, Satapathy SC, Ashour AS. An approach to examine magnetic resonance angiography based on Tsallis entropy and deformable snake model. *Futur Gener Comput Syst*. 2018;85:160–72.
96. Kumar R, Talukdar FA, Dey N, Ashour AS, Santhi V, Balas VE, Shi F. Histogram thresholding in image segmentation: a joint level set method and lattice boltzmann method based approach. In: *Information technology and intelligent transportation systems*. Springer, Cham; 2017. pp. 529–539.
97. Srikham M. Active contours segmentation with edge based and local region based. In: *Proceedings of the 21st international conference on pattern recognition (ICPR2012)*. IEEE; 2012. pp. 1989–1992.
98. Chan TF, Vese LA. Active contours without edges. *IEEE Trans Image Process*. 2001;10(2):266–77.
99. Feng H, Castanon DA, Karl WC. A curve evolution approach for image segmentation using adaptive flows. In: *Proceedings eighth IEEE international conference on computer vision. ICCV 2001. Vol. 2*. IEEE; 2001. pp. 494–499.
100. Tsai A, Yezzi A, Willisky AS. Curve evolution implementation of the Mumford-Shah functional for image segmentation, denoising, interpolation, and magnification. 2001.
101. Osher S, Sethian JA. Fronts propagating with curvature-dependent speed: algorithms based on Hamilton–Jacobi formulations. *J Comput Phys*. 1988;79(1):12–49.
102. Zhu SC, Yuille A. Region competition: unifying snakes, region growing, and Bayes/MDL for multiband image segmentation. *IEEE Trans Pattern Anal Mach Intell*. 1996;9:884–900.
103. Yilmaz A, Li X, Shah M. Object contour tracking using level sets. In: *Asian conference on computer vision*. 2004.
104. Wang F. Particle filters for visual tracking. In: *International conference on computer science and information engineering*. Springer, Berlin; 2011. pp. 107–112.
105. Varas D, Marques F. Region-based particle filter for video object segmentation. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. pp. 3470–3477.
106. Li H, Wang Y. Object of interest tracking based on visual saliency and feature points matching. 2015.
107. Chantara W, Mun JH, Shin DW, Ho YS. Object tracking using adaptive template matching. *IEIE Trans Smart Process Comput*. 2015;4(1):1–9.
108. Baker S, Matthews I. Lucas-kanade 20 years on: a unifying framework. *Int J Comput Vis*. 2004;56(3):221–55.
109. Benhimane S, Malis E. Homography-based 2d visual tracking and servoing. *Int J Robot Res*. 2007;26(7):661–76.
110. Kwon J, Lee HS, Park FC, Lee KM. A geometric particle filter for template-based visual tracking. *IEEE Trans Pattern Anal Mach Intell*. 2014;36(4):625–43.
111. Lin Z, Davis LS, Doermann D, DeMenthon D. Hierarchical part-template matching for human detection and segmentation. In: *2007 IEEE 11th international conference on computer vision*. IEEE; 2007. pp. 1–8.
112. Ren X, Malik J. Tracking as repeated figure/ground segmentation. In: *CVPR. Vol. 1*. 2007. p. 7.
113. Chuang YY, Agarwala A, Curless B, Salesin DH, Szeliski R. Video matting of complex scenes. In: *ACM transactions on graphics (ToG)*. Vol. 21, No. 3. ACM; 2002. pp. 243–248.
114. Wang J, Bhat P, Colburn RA, Agrawala M, Cohen MF. Interactive video cutout. In: *ACM transactions on graphics (ToG)*. Vol. 24, No. 3. ACM; pp. 585–594.
115. Li Y, Sun J, Tang CK, Shum HY. Lazy snapping. *ACM Trans Graph (ToG)*. 2004;23(3):303–8.
116. Rother C, Kolmogorov V, Blake A. Interactive foreground extraction using iterated graph cuts. *ACM Trans Graph*. 2004;23:3.
117. Lu L, Hager GD. A nonparametric treatment for location/segmentation based visual tracking. In: *2007 IEEE conference on computer vision and pattern recognition*. IEEE; pp. 1–8.
118. Levinshtein A, Stere A, Kutulakos KN, Fleet DJ, Dickinson SJ, Siddiqi K. Turbopixels: fast superpixels using geometric flows. *IEEE Trans Pattern Anal Mach Intell*. 2009;31(12):2290–7.
119. Achanta R, Shaji A, Smith K, Lucchi A, Fua P, Sisstrunk S. SLIC superpixels compared to state-of-the-art superpixel methods. *IEEE Trans Pattern Anal Mach Intell*. 2012;34(11):2274–82.
120. Hu J, Fan XP, Liu S, Huang L. Robust target tracking algorithm based on superpixel visual attention mechanism: robust target tracking algorithm. *Int J Ambient Comput Intell (IJACI)*. 2019;10(2):1–17.
121. Wang S, Lu H, Yang F, Yang MH. Superpixel tracking. In: *2011 International conference on computer vision*. IEEE; 2011. pp. 1323–1330.
122. Dey N, Ashour AS, Hassanien AE. Feature detectors and descriptors generations with numerous images and video applications: a recap. In: *Feature detectors and motion detection in video processing*. IGI Global; 2017. pp. 36–65.
123. Hore S, Bhattacharya T, Dey N, Hassanien AE, Banerjee A, Chaudhuri SB. A real time dactylogy based feature extraction for selective image encryption and artificial neural network. In: *Image feature detectors and descriptors*. Springer, Cham; 2016. pp. 203–226.
124. Tharwat A, Gaber T, Awad YM, Dey N, Hassanien AE. Plants identification using feature fusion technique and bagging classifier. In: *The 1st international conference on advanced intelligent system and informatics (AISII2015)*, November 28–30, 2015, Beni Suef, Egypt. Springer, Cham; 2016. pp. 461–471.
125. Lowe DG. Distinctive image features from scale-invariant keypoints. *Int J Comput Vis*. 2004;60(2):91–110.
126. Wang Z, Xiao H, He W, Wen F, Yuan K. Real-time SIFT-based object recognition system. In: *2013 IEEE international conference on mechatronics and automation*. IEEE; 2013; pp. 1361–1366.
127. Park C, Jung S. SIFT-based object recognition for tracking in infrared imaging system. In: *2009 34th International conference on infrared, millimeter, and terahertz waves*; IEEE; 2009. pp. 1–2.
128. Mirunalini P, Jaisakthi SM, Sujana R. Tracking of object in occluded and non-occluded environment using SIFT and Kalman filter. In: *TENCON 2017-2017 IEEE Region 10 Conference*. IEEE; 2017. pp. 1290–1295.
129. Li Q, Li R, Ji K, Dai W. Kalman filter and its application. In: *2015 8th International Conference on Intelligent Networks and Intelligent Systems (ICINIS)*. IEEE; 2015. pp. 74–77.
130. Cane T, Ferryman J. Saliency-based detection for maritime object tracking. In: *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*. 2016. pp. 18–25.
131. Borji A, Cheng MM, Hou Q, Jiang H, Li J. Salient object detection: a survey. *arXiv preprint arXiv:1411.5878*. 2014.
132. Itti L, Koch C, Niebur E. A model of saliency-based visual attention for rapid scene analysis. *IEEE Trans Pattern Anal Mach Intell*. 1998;11:1254–9.
133. Liu T, Yuan Z, Sun J, Wang J, Zheng N, Tang X, Shum HY. Learning to detect a salient object. *IEEE Trans Pattern Anal Mach Intell*. 2011;33(2):353–67.
134. Zhang G, Yuan Z, Zheng N, Sheng X, Liu T. Visual saliency based object tracking. In: *Asian conference on computer vision*. 2009; Springer, Berlin. pp. 193–203.

135. Taycher L, Shakhnarovich G, Demirdjian D, Darrell T. Conditional random people: tracking humans with crfs and grid filters (No. MIT-CSAIL-TR-2005-079). Massachusetts Inst of Tech Cambridge Computer Science and Artificial Intelligence Lab. 2005.
136. Jeong J, Yoon TS, Park JB. Mean shift tracker combined with online learning-based detector and Kalman filtering for real-time tracking. *Expert Syst Appl*. 2017;79:194–206.
137. Xu L, Zeng L, Duan H, Sowah NL. Saliency detection in complex scenes. *EURASIP J Image Video Process*. 2014;2014(1):31.
138. Liu Q, Zhao X, Hou Z. Survey of single-target visual tracking methods based on online learning. *IET Comput Vis*. 2014;8(5):419–28.
139. Bacivarov I, Ionita M, Corcoran P. Statistical models of appearance for eye tracking and eye-blink detection and measurement. *IEEE Trans Consum Electron*. 2008;54(3):1312–20.
140. Dou J, Qin Q, Tu Z. Robust visual tracking based on generative and discriminative model collaboration. *Multimed Tools Appl*. 2017;76(14):15839–66.
141. Kawamoto K, Yonekawa T, Okamoto K. Visual vehicle tracking based on an appearance generative model. In: The 6th international conference on soft computing and intelligent systems, and the 13th international symposium on advanced intelligence systems. IEEE; 2012. pp. 711–714.
142. Chakraborty B, Bhattacharyya S, Chakraborty S. Generative model based video shot boundary detection for automated surveillance. *Int J Ambient Comput Intell (IJACI)*. 2018;9(4):69–95.
143. Remya KV, Vipin Krishnan CV. Survey of generative and discriminative appearance models in visual object tracking. *Int J Adv Res Ideas Innov Technol*. 2018;4(1). www.IJARIIT.com.
144. Jepson AD, Fleet DJ, El-Maraghi TF. Robust online appearance models for visual tracking. *IEEE Trans Pattern Anal Mach Intell*. 2003;25(10):1296–311.
145. Zhou SK, Chellappa R, Moghaddam B. Visual tracking and recognition using appearance-adaptive models in particle filters. *IEEE Trans Image Process*. 2004;13(11):1491–506.
146. Gao M, Shen J, Jiang J. Visual tracking using improved flower pollination algorithm. *Optik*. 2018;156:522–9.
147. Yang H, Shao L, Zheng F, Wang L, Song Z. Recent advances and trends in visual tracking: a review. *Neurocomputing*. 2011;74(18):3823–31.
148. Lee KC, Ho J, Yang MH, Kriegman D. Video-based face recognition using probabilistic appearance manifolds. In: IEEE computer society conference on computer vision and pattern recognition. Vol. 1. IEEE Computer Society; 1999. pp. I–313.
149. Ross DA, Lim J, Lin RS, Yang MH. Incremental learning for robust visual tracking. *Int J Comput Vision*. 2008;77(1–3):125–41.
150. Wen J, Li X, Gao X, Tao D. Incremental learning of weighted tensor subspace for visual tracking. In: 2009 IEEE international conference on systems, man and cybernetics. IEEE; 2009. pp. 3688–3693.
151. Funt BV, Ciurea F, McCann JJ. Retinex in matlab tm. *J Electron Imaging*. 2004;13(1):48–58.
152. Ju MH, Kang HB. Illumination invariant face tracking and recognition. 2008.
153. Jia X, Lu H, Yang MH. Visual tracking via adaptive structural local sparse appearance model. In: 2012 IEEE Conference on computer vision and pattern recognition. IEEE. 2012. pp. 1822–1829.
154. Dou Jianfang, Qin Qin, Tu Zimei. Robust visual tracking based on generative and discriminative model collaboration. *Multimed Tools Appl*. 2016. <https://doi.org/10.1007/s11042-016-3872-6>.
155. Zhang K, Zhang L, Yang MH. Real-time compressive tracking. In: European conference on computer vision. Springer, Berlin; 2012. pp. 864–877.
156. Zhou T, Liu F, Bhaskar H, Yang J. Robust visual tracking via online discriminative and low-rank dictionary learning. *IEEE Trans Cybern*. 2018;48(9):2643–55.
157. Fan H, Xiang J, Li G, Ni F. Robust visual tracking via deep discriminative model. In: 2017 IEEE international conference on acoustics, speech and signal processing (ICASSP). IEEE; 2017. pp. 1927–1931.
158. Babenko B, Yang MH, Belongie S. Robust object tracking with online multiple instance learning. *IEEE Trans Pattern Anal Mach Intell*. 2011;33(8):1619–32.
159. Hare S, Saffari A, Struck PHT. Structured output tracking with kernels. In: IEEE international conference on computer vision. IEEE; 2012. pp. 263–270.
160. Avidan S. Support vector tracking. In: Proceedings of the 2001 IEEE computer society conference on computer vision and pattern recognition. CVPR 2001. Vol. 1. IEEE; 2001. pp. I–I.
161. Grabner H, Leistner C, Bischof H. Semi-supervised on-line boosting for robust tracking. In: European conference on computer vision. Springer, Berlin; 2008. pp. 234–247.
162. Saffari A, Leistner C, Santner J, Godec M, Bischof H. On-line random forests. In: 2009 IEEE 12th international conference on computer vision workshops, ICCV workshops. IEEE; 2009. pp. 1393–1400.
163. Henriques JF, Caseiro R, Martins P, Batista J. Exploiting the circulant structure of tracking-by-detection with kernels. In: European conference on computer vision. Springer, Berlin; 2012. pp. 702–715.
164. Li X, Liu Q, He Z, Wang H, Zhang C, Chen WS. A multi-view model for visual tracking via correlation filters. *Knowl-Based Syst*. 2016;113:88–99.
165. Bolme DS, Beveridge JR, Draper BA, Lui YM. Visual object tracking using adaptive correlation filters. In: 2010 IEEE computer society conference on computer vision and pattern recognition. IEEE; 2010. pp. 2544–2550.
166. Danelljan M, Häger G, Khan F, Felsberg M. Accurate scale estimation for robust visual tracking. In: British machine vision conference, Nottingham, September 1–5, 2014. BMVA Press.
167. Danelljan M, Shahbaz Khan F, Felsberg M, Van de Weijer J. Adaptive color attributes for real-time visual tracking. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2014. pp. 1090–1097.
168. Li Y, Zhu J. A scale adaptive kernel correlation filter tracker with feature integration. In: European conference on computer vision. Springer, Cham; 2014. pp. 254–265.
169. Danelljan M, Bhat G, Gladh S, Khan FS, Felsberg M. Deep motion and appearance cues for visual tracking. *Pattern Recogn Lett*. 2019;124:74–81.
170. Danelljan M, Häger G, Khan FS, Felsberg M. Discriminative scale space tracking. *IEEE Trans Pattern Anal Mach Intell*. 2017;39(8):1561–75.
171. Duffner S, Garcia C. Using discriminative motion context for online visual object tracking. *IEEE Trans Circuits Syst Video Technol*. 2016;26(12):2215–25.
172. Ulusoy I, Bishop CM. Generative versus discriminative methods for object recognition. In: 2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05). Vol. 2. IEEE; 2005. pp. 258–265.
173. McCallum A, Pal C, Druck G, Wang X. Multi-conditional learning: generative/discriminative training for clustering and classification. In: AAAI. 2006. pp. 433–439.
174. Kelm BM, Pal C, McCallum A. Combining generative and discriminative methods for pixel classification with

- multi-conditional learning. In: 18th International conference on pattern recognition (ICPR'06). Vol. 2. IEEE; 2006. pp. 828–832.
175. Blake A, Rother C, Brown M, Perez P, Torr P. Interactive image segmentation using an adaptive GMMRF model. In: European conference on computer vision. Springer, Berlin. 2004. pp. 428–441.
 176. Acharjee S, Dey N, Biswas D, Das P, Chaudhuri SS. A novel Block Matching Algorithmic Approach with smaller block size for motion vector estimation in video compression. In: 2012 12th International conference on intelligent systems design and applications (ISDA). IEEE; 2012. pp. 668–672.
 177. Acharjee S, Biswas D, Dey N, Maji P, Chaudhuri SS. An efficient motion estimation algorithm using division mechanism of low and high motion zone. In: 2013 International multi-conference on automation, computing, communication, control and compressed sensing (iMac4s). IEEE; 2013. pp. 169–172.
 178. Acharjee S, Ray R, Chakraborty S, Nath S, Dey N. Watermarking in motion vector for security enhancement of medical videos. In: 2014 International conference on control, instrumentation, communication and computational technologies (ICCICT). IEEE; 2014. pp. 532–537.
 179. Acharjee S, Chakraborty S, Karra WBA, Azar AT, Dey N. Performance evaluation of different cost functions in motion vector estimation. *Int J Service Sci Manag Eng Technol (IJSSMET)*. 2014;5(1):45–65.
 180. Acharjee S, Chakraborty S, Samanta S, Azar AT, Hassanien AE, Dey N. Highly secured multilayered motion vector watermarking. In: International conference on advanced machine learning technologies and applications. Springer, Cham; 2014. pp. 121–134.
 181. Acharjee S, Pal G, Redha T, Chakraborty S, Chaudhuri SS, Dey N. Motion vector estimation using parallel processing. In: International Conference on Circuits, Communication, Control and Computing. IEEE; 2014. pp. 231–236.
 182. Rawat P, Singhai J. Review of motion estimation and video stabilization techniques for hand held mobile video. *Sig Image Proc Int J (SIPIJ)*. 2011;2(2):159–68.
 183. Irani M, Anandan P. About direct methods. In: International workshop on vision algorithms. Springer, Berlin; 1999. pp. 267–277.
 184. Torr PH, Zisserman A. Feature based methods for structure and motion estimation. In: International workshop on vision algorithms. Springer, Berlin; 1999. pp. 278–294.
 185. Fiaz M, Mahmood A, Jung SK. Tracking noisy targets: a review of recent object tracking approaches. *arXiv preprint arXiv:1802.03098*. 2018.
 186. Kristan M, Matas J, Leonardis A, Felsberg M, Cehovin L, Fernandez G, et al. The visual object tracking vot2015 challenge results. In: Proceedings of the IEEE international conference on computer vision workshops. 2015. pp. 1–23.
 187. Čehovin L, Leonardis A, Kristan M. Visual object tracking performance measures revisited. *IEEE Trans Image Process*. 2016;25(3):1261–74.
 188. Wu Y, Lim J, Yang MH. Online object tracking: a benchmark. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2013. pp. 2411–2418.
 189. Everingham M, Van Gool L, Williams CK, Winn J, Zisserman A. The pascal visual object classes (voc) challenge. *Int J Comput Vis*. 2010;88(2):303–38.
 190. Hare S, Golodetz S, Saffari A, Vineet V, Cheng MM, Hicks SL, Torr PH. Struck: structured output tracking with kernels. *IEEE Trans Pattern Anal Mach Intell*. 2016;38(10):2096–109.
 191. Fang Y, Yuan Y, Li L, Wu J, Lin W, Li Z. Performance evaluation of visual tracking algorithms on video sequences with quality degradation. *IEEE Access*. 2017;5:2430–41.
 192. Kwon J, Lee KM. Tracking of a non-rigid object via patch-based dynamic appearance modeling and adaptive basin hopping Monte Carlo sampling. In: 2009 IEEE conference on computer vision and pattern recognition. IEEE; 2009. pp. 1208–1215.
 193. Yang F, Lu H, Yang MH. Robust superpixel tracking. *IEEE Trans Image Process*. 2014;23(4):1639–51.
 194. Kristan M, Kovacic S, Leonardis A, Pers J. A two-stage dynamic model for visual tracking. *IEEE Trans Syst Man Cybern Part B (Cybernetics)*. 2010;40(6):1505–20.
 195. Kristan M, Pers J, Perse M, Kovacic S, Bon M. Multiple interacting targets tracking with application to team sports. In: ISPA 2005. Proceedings of the 4th international symposium on image and signal processing and analysis. IEEE; 2005. pp. 322–327.
 196. Nawaz T, Cavallaro A. A protocol for evaluating video trackers under real-world conditions. *IEEE Trans Image Process*. 2013;22(4):1354–61.
 197. Zhang X, Payandeh S. Application of visual tracking for robot-assisted laparoscopic surgery. *J Robot Syst*. 2002;19(7):315–28.
 198. Dey N, Ashour AS, Shi F, Sherratt RS. Wireless capsule gastrointestinal endoscopy: direction-of-arrival estimation based localization survey. *IEEE Rev Biomed Eng*. 2017;10:2–11.
 199. Su MC, Wang KC, Chen GD. An eye tracking system and its application in aids for people with severe disabilities. *Biomed Eng Appl Basis Commun*. 2006;18(06):319–27.
 200. Chen Y, Levy DL, Sheremata S, Holzman PS. Bipolar and schizophrenic patients differ in patterns of visual motion discrimination. *Schizophr Res*. 2006;88(1–3):208–16.
 201. Raudonis V, Simutis R, Narvydas G. Discrete eye tracking for medical applications. In: 2009 2nd International Symposium on Applied Sciences in Biomedical and Communication Technologies. IEEE; 2009. pp. 1–6.
 202. De Santis A, Iacoviello D. A robust eye tracking procedure for medical and industrial applications. In: Advances in computational vision and medical image processing. Springer, Dordrecht; 2009. pp. 173–185.
 203. Harezlak K, Kasprowski P. Application of eye tracking in medicine: a survey, research issues and challenges. *Comput Med Imaging Graph*. 2018;65:176–90.
 204. Lennon J, Atkins E. Color-based vision tracking for an astronaut EVA assist vehicle (No. 2001-01-2135). SAE Technical Paper. 2001.
 205. Borra S, Thanki R, Dey N. Satellite image classification. In: Satellite image analysis: clustering and classification. Springer, Singapore. pp. 53–81.
 206. Zhao Q, Yang Z, Tao H. Differential earth mover's distance with its applications to visual tracking. *IEEE Trans Pattern Anal Mach Intell*. 2010;32(2):274–87.
 207. Kamate S, Yilmazer N. Application of object detection and tracking techniques for unmanned aerial vehicles. *Proc Comput Sci*. 2015;61:436–41.
 208. Zhang R, Wang Z, Zhang Y. Astronaut visual tracking of flying assistant robot in space station based on deep learning and probabilistic model. *Int J Aerosp Eng*. 2018.
 209. Mistry P, Maes P, Chang L. WUW-wear Ur world: a wearable gestural interface. In: CHI'09 extended abstracts on human factors in computing systems. ACM; 2009. pp. 4111–4116.
 210. Kerdvibulvech C. Markerless vision-based tracking for interactive augmented reality game. *Int J Interact Worlds (IJIW'10)*. 2010.
 211. Kerdvibulvech C. Asiatic skin color segmentation using an adaptive algorithm in changing luminance environment. 2011.
 212. Klein G, Murray D. Parallel tracking and mapping on a camera phone. In: 2009 8th IEEE international symposium on mixed and augmented reality. IEEE; 2009. pp. 83–86.

213. Woodward C, Hakkarainen M. Mobile mixed reality system for architectural and construction site visualization. In: *Augmented reality-some emerging application areas*. IntechOpen; 2011.
214. Dantone M, Bossard L, Quack T, Van Gool L. Augmented faces. In: *2011 IEEE international conference on computer vision workshops (ICCV Workshops)*. IEEE; 2011. pp. 24–31.
215. Kerdvibulvech C. Augmented reality applications using visual tracking. *วารสารเทคโนโลยีสารสนเทศลาดกระบัง*. 2016;2(1).
216. Casas S, Olanda R, Dey N. Motion cueing algorithms: a review: algorithms, evaluation and tuning. *Int J Virtual Augment Reality (IJVAR)*. 2017;1(1):90–106.
217. Ribo M, Lang P, Ganster H, Brandner M, Stock C, Pinz A. Hybrid tracking for outdoor augmented reality applications. *IEEE Comput Graph Appl*. 2002;22(6):54–63.
218. Klopschitz M, Schall G, Schmalstieg D, Reitmayr G. Visual tracking for augmented reality. In: *2010 International conference on indoor positioning and indoor navigation*. IEEE; 2010. pp. 1–4.
219. Reitmayr G, Drummond T. Going out: robust model-based tracking for outdoor augmented reality. In: *ISMAR*. Vol. 6. 2006. pp. 109–118.
220. Rehg JM, Kanade T. Visual tracking of high dof articulated structures: an application to human hand tracking. In: *European conference on computer vision*. Springer, Berlin; 1994. pp. 35–46.
221. Gavrilu DM. The visual analysis of human movement: a survey. *Comput Vis Image Underst*. 1999;73(1):82–98.
222. Lathuiliere F, Herve JY. Visual hand posture tracking in a gripper guiding application. In: *Proceedings 2000 ICRA. Millennium Conference. IEEE International Conference on Robotics and Automation. Symposia Proceedings (Cat. No. 00CH37065)*. Vol. 2. IEEE; 2000. pp. 1688–1694.
223. Chen ZW, Chiang CC, Hsieh ZT. Extending 3D Lucas-Kanade tracking with adaptive templates for head pose estimation. *Mach Vis Appl*. 2010;21(6):889–903.
224. Nistér D, Naroditsky O, Bergen J. Visual odometry for ground vehicle applications. *J Field Robot*. 2006;23(1):3–20.
225. Bonin-Font F, Ortiz A, Oliver G. Visual navigation for mobile robots: a survey. *J Intell Rob Syst*. 2008;53(3):263–96.
226. Borenstein J, Koren Y. Real-time obstacle avoidance for fast mobile robots. *IEEE Trans Syst Man Cybern*. 1989;19(5):1179–87.
227. Bernardino A, Santos-Victor J. Visual behaviours for binocular tracking. *Robot Auton Syst*. 1998;25(3–4):137–46.
228. Ciliberto C, Pattacini U, Natale L, Nori F, Metta G. Reexamining lucas-kanade method for real-time independent motion detection: application to the icub humanoid robot. In: *2011 IEEE/RSJ international conference on intelligent robots and systems*. IEEE; 2011. pp. 4154–4160.
229. Das PK, Mandhata SC, Panda CN, Patro SN. Vision based object tracking by mobile robot. *Int J Comput Appl*. 2012;45(8):40–2.
230. Sibert JL, Gokturk M, Lavine RA. The reading assistant: eye gaze triggered auditory prompting for reading remediation. In: *Proceedings of the 13th annual ACM symposium on user interface software and technology*. ACM; 2000. pp. 101–107.
231. Bolt RA. Eyes at the interface. In: *Proceedings of the 1982 conference on Human factors in computing systems*. ACM; 1982. pp. 360–362.
232. Jacob RJ. Eye movement-based human-computer interaction techniques: toward non-command interfaces. *Adv Hum Comput Interact*. 1993;4:151–90.
233. Sibert LE, Jacob RJ. Evaluation of eye gaze interaction. In: *Proceedings of the SIGCHI conference on human factors in computing systems*. ACM; 2000. pp. 281–288.
234. McConkie GW, Zola D. Eye movement techniques in studying differences among developing readers. *Center for the study of reading technical report*; no. 377. 1986.
235. O'Regan JK. Eye movements and reading. *Rev Oculomot Res*. 1990;4:395–453.
236. Rayner K. Eye movements in reading and information processing: 20 years of research. *Psychol Bull*. 1998;124(3):372.
237. Wang H, Chignell M, Ishizuka M. Empathic tutoring software agents using real-time eye tracking. In: *Proceedings of the 2006 symposium on eye tracking research and applications*. ACM; 2006. pp. 73–78.
238. Tsai MJ, Hou HT, Lai ML, Liu WY, Yang FY. Visual attention for solving multiple-choice science problem: an eye-tracking analysis. *Comput Educ*. 2012;58(1):375–85.
239. Dessus P, Cosnefroy O, Luengo V. “Keep Your Eyes on'em all!”: a mobile eye-tracking analysis of teachers' sensitivity to students. In: *European conference on technology enhanced learning*. Springer, Cham; 2016. pp. 72–84.
240. Busjahn T, Schulte C, Sharif B, Begel A, Hansen M, Bednarik R, et al. Eye tracking in computing education. In: *Proceedings of the tenth annual conference on International computing education research*. ACM; 2014. pp. 3–10.
241. Sun Y, Li Q, Zhang H, Zou J. The application of eye tracking in education. In: *International conference on intelligent information hiding and multimedia signal processing*. Springer, Cham; 2017. pp. 27–33.
242. Obaidallah U, Al Haek M, Cheng PCH. A survey on the usage of eye-tracking in computer programming. *ACM Comput Surv (CSUR)*. 2018;51(1):5.
243. Smith AW, Lovell BC. Visual tracking for sports applications. 2005.
244. Mauthner T, Bischof H. A robust multiple object tracking for sport applications. 2007.
245. Battal Ö, Balçioğlu T, Duru AD. Analysis of gaze characteristics with eye tracking system during repeated breath holding exercises in underwater hockey elite athletes. In: *2016 20th National Biomedical Engineering Meeting (BIYOMUT)*. IEEE; 2016. pp. 1–4.
246. Kredel R, Vater C, Klostermann A, Hossner EJ. Eye-tracking technology and the dynamics of natural gaze behavior in sports: a systematic review of 40 years of research. *Front Psychol*. 2017;8:1845.
247. Discombe RM, Cotterill ST. Eye tracking in sport: a guide for new and aspiring researchers. *Sport Exerc Psychol Rev*. 2015;11(2):49–58.
248. Mademlis I, Mygdalis V, Nikolaidis N, Pitas I. Challenges in autonomous UAV cinematography: an overview. In: *2018 IEEE international conference on multimedia and expo (ICME)*. IEEE; 2018. pp. 1–6.
249. Passalis N, Tefas A, Pitas I. Efficient camera control using 2D visual information for unmanned aerial vehicle-based cinematography. In: *2018 IEEE international symposium on circuits and systems (ISCAS)*. IEEE; 2018. pp. 1–5.
250. Hubbard AW, Seng CN. Visual movements of batters. *Res Q Am Assoc Health Phys Educ Recreat*. 1954;25(1):42–57.
251. Zachariadis O, Mygdalis V, Mademlis I, Nikolaidis N, Pitas I. 2D visual tracking for sports UAV cinematography applications. In: *2017 IEEE global conference on signal and information processing (GlobalSIP)*. IEEE; 2017. pp. 36–40.
252. Ramli L, Mohamed Z, Abdullahi AM, Jaafar HI, Lazim IM. Control strategies for crane systems: a comprehensive review. *Mech Syst Signal Process*. 2017;95:1–23.
253. Peng KCC, Singhoose W, Bhaumik P. Using machine vision and hand-motion control to improve crane operator performance. *IEEE Trans Syst Man Cybern Part A Syst Hum*. 2012;42(6):1496–503.
254. Wedel M, Pieters R. A review of eye-tracking research in marketing. In: *Review of marketing research*. Emerald Group Publishing Limited; 2008. pp. 123–147.

255. Koller M, Salzberger T, Brenner G, Walla P. Broadening the range of applications of eye-tracking in business research. *Análise Porto Alegre*. 2012;23(1):71–7.
256. Zamani H, Abas A, Amin MKM. Eye tracking application on emotion analysis for marketing strategy. *J Telecommun Electron Comput Eng (JTEC)*. 2016;8(11):87–91.
257. Wedel M, Pieters R. Eye tracking for visual marketing. *Found Trends Market*. 2008;1(4):231–320.
258. dos Santos RDOJ, de Oliveira JHC, Rocha JB, Giraldi JDME. Eye tracking in neuromarketing: a research agenda for marketing studies. *Int J Psychol Stud*. 2015;7(1):32.
259. Boraston Z, Blakemore SJ. The application of eye-tracking technology in the study of autism. *J Physiol*. 2007;581(3):893–8.
260. Babenko B, Yang MH, Belongie S. Visual tracking with online multiple instance learning. In: 2009 IEEE conference on computer vision and pattern recognition. IEEE;2009. pp. 983–990.
261. Hu D, Zhou X, Yu X, Hou Z. Study on deep learning and its application in visual tracking. In: 2015 10th International conference on broadband and wireless computing, communication and applications (BWCCA). IEEE; 2015. pp. 240–246.
262. Liu W, Wang Z, Liu X, Zeng N, Liu Y, Alsaadi FE. A survey of deep neural network architectures and their applications. *Neurocomputing*. 2017;234:11–26.
263. Lan K, Wang DT, Fong S, Liu LS, Wong KK, Dey N. A survey of data mining and deep learning in bioinformatics. *J Med Syst*. 2018;42(8):139.
264. Dey N, Ashour AS, Borra S. (Eds.). *Classification in bioapps: automation of decision making*. Vol. 26. Springer; 2017.
265. Avidan S. Support vector tracking. In: Proceedings of the 2001 IEEE computer society conference on computer vision and pattern recognition. CVPR 2001. Vol. 1. IEEE; 2001. pp. I–I.
266. Schulter S, Leistner C, Wohlhart P, Roth PM, Bischof H. Accurate object detection with joint classification-regression random forests. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2014. pp. 923–930.
267. Anguita D, Parodi G, Zunino R. Neural structures for visual motion tracking. *Mach Vis Appl*. 1995;8(5):275–88.
268. Zhang, J., Yang, L., & Wu, X. (2016, October). A survey on visual tracking via convolutional neural networks. In 2016 2nd IEEE International Conference on Computer and Communications (ICCC) (pp. 474–479). IEEE.
269. Sultana M, Mahmood A, Javed S, Jung SK. Unsupervised deep context prediction for background estimation and foreground segmentation. *Mach Vis. Appl*. 2019;30(3):375–95.
270. Hu L, Hong C, Zeng Z, Wang X. Two-stream person re-identification with multi-task deep neural networks. *Mach Vis Appl*. 2018;29(6):947–54.
271. Li Z, Dey N, Ashour AS, Cao L, Wang Y, Wang D, et al. Convolutional neural network based clustering and manifold learning method for diabetic plantar pressure imaging dataset. *J Med Imaging Health Inf*. 2017;7(3):639–52.
272. Wang Y, Chen Y, Yang N, Zheng L, Dey N, Ashour AS, et al. Classification of mice hepatic granuloma microscopic images based on a deep convolutional neural network. *Appl Soft Comput*. 2019;74:40–50.
273. Wang D, Li Z, Dey N, Ashour AS, Moraru L, Biswas A, Shi F. Optical pressure sensors based plantar image segmenting using an improved fully convolutional network. *Optik*. 2019;179:99–114.
274. Hu S, Liu M, Fong S, Song W, Dey N, Wong R. Forecasting China future MNP by deep learning. In: *Behavior engineering and applications*. Springer, Cham. 2018. pp. 169–210.
275. Zhuo L, Jiang L, Zhu Z, Li J, Zhang J, Long H. Vehicle classification for large-scale traffic surveillance videos using convolutional neural networks. *Mach Vis Appl*. 2017;28(7):793–802.
276. Dey N, Fong S, Song W, Cho K. Forecasting energy consumption from smart home sensor network by deep learning. In: *International Conference on Smart Trends for Information Technology and Computer Communications*. Springer, Singapore. 2017. pp. 255–265.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.