Visual object tracking—classical and contemporary approaches

Article *in* Frontiers of Computer Science (print) · November 2015

DOI: 10.1007/s11704-015-4246-3

Ahmad Ali
Pakistan Institute of Engineering and Applied Sciences

Abdul Jalil
Pakistan Institute of Engineering and Applied Sciences

Jianwei Niu
Beihang University (BUAA)

Xiaoke Zhao
Beihang University (BUAA)

Front. Comput. Sci.
DOI 10.1007/s11704-015-4246-3

RESEARCH ARTICLE

Visual object tracking — classical and contemporary approaches

Ahmad ALI[1], Abdul JALIL[1], Jianwei NIU (✉)[2], Xiaoke ZHAO[2], Saima RATHORE[1], Javed AHMED[3], Muhammad AKSAM IFTIKHAR[4]

1 Department of Computer and Information Sciences, Pakistan Institute of Engineering & Applied Sciences, Inlamabad 44000, Pakistan
2 State Key Laboratory of Virtual Reality Technology and Systems, School of Computer Science and Engineering, Beihang University, Beijing 100191, China
3 Department of Computer Science, NUST Military College of Signals, Islamabad 44000, Pakistan
4 COMSATS Institute of Information Technology, Lahore 54000, Pakistan

Abstract Visual object tracking (VOT) is an important subfield of computer vision. It has widespread application domains, and has been considered as an important part of surveillance and security system. VOA facilitates finding the position of target in image coordinates of video frames. While doing this, VOA also faces many challenges such as noise, clutter, occlusion, rapid change in object appearances, highly maneuvered (complex) object motion, illumination changes. In recent years, VOT has made significant progress due to availability of low-cost high-quality video cameras as well as fast computational resources, and many modern techniques have been proposed to handle the challenges faced by VOT. This article introduces the readers to 1) VOT and its applications in other domains, 2) different issues which arise in it, 3) various classical as well as contemporary approaches for object tracking, 4) evaluation methodologies for VOT, and 5) online resources, i.e., annotated datasets and source code available for various tracking techniques.

Keywords visual object tracking, computer vision, image processing, point tracking, kernel tracking, silhouette tracking

Received May 23, 2014; accepted March 24, 2015

E-mail: niujianwei@buaa.edu.cn

1 Introduction

Visual object tracking (VOT) is a well known sub-field of computer vision and hence image processing. Its main objective is to find the locus of points that the target of interest follows in image coordinates by assigning labels in temporarily spaced video frames. This information may be of pertinent importance for further analysis such as to calculate the centre of mass, area, perimeter and motion vector of the target. Thus, target tracking may play a vital role in high level image analysis tasks, such as object recognition [1,2], activity analysis [3,4], and intelligent scene understanding [5]. Usability spectrum of VOT is very wide, and it has found its applications in real world systems. A few of its applications are shown in Fig. 1, which includes:

- **Human machine interaction (HMI):** VOT may play an important role in this emerging field to make better community life by making easy-to-use interaction with machines, e.g., sixth-sense (a wearable gesture interface) [6], perceptual user interfaces [7], eye gaze tracking for disabled people [8].
- **Visual surveillance and security systems:** These systems are now ubiquitous, and VOT is a vital part of intelligent visual surveillance, e.g., third generation surveillance systems (3GSS) [9], Siemens Sistore CX

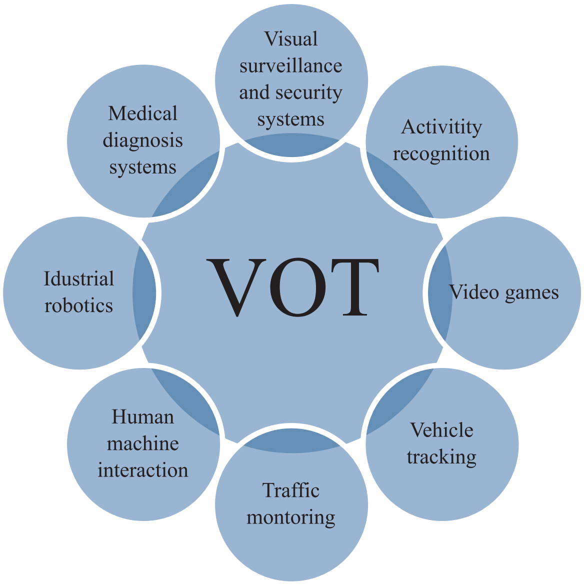


Fig. 1 Usage of VOT in different applications

EDS [10], surveillance of places and buildings related to public and defense interests for intruder detection [11], monitoring human activities [12–17].

- **Traffic monitoring:** VOT may be used for motoring and managing traffic on roads, e.g., monitoring of traffic flow [18], detection of traffic accidents [19], counting of pedestrian [20].
- **Industrial robotics:** VOT may be used in control system of industrial and humanoid robotics, e.g., using vision sensor with tracking algorithm in feed-back loop [21], ASIMO humanoid robot [22], visual control for unmanned aerial vehicle (UAV) [23].
- **Vehicle tracking:** VOT may be used for automobile tracking, e.g., tracking a vehicle by UAV [24], tracking vehicles on the road to assist driver [25,26], and autopilot of a UGV [27].
- **Video games:** VOT may be used for video games to provide better user control, e.g., tracking user movements [28], face tracking for playing game [29].
- **Medical diagnosis systems:** VOT has pushed its importance in medical field for diagnosis of different diseases, e.g., tracking of ventricular wall [30], reconstruction of vocal tract shape [31,32].
- **Activity recognition:** VOT may be used for activity recognition for indoor and out-door monitoring, e.g., learning activity patterns [33] and human activity recognition [34].

1.1 Issues in VOT

Great efforts have been made by researchers in the field of VOT for the last four decades [35,36]; however, it is still an open avenue for computer vision research community due to various issues as shown in Fig. 2. The issues are also described as follows:

Fig. 2 Different issues that may arise during tracking. (a) Occlusion; (b) Changing appearance; (c) Cluttered background; (d) Changing size in image; (e) Illumination variations; (f) Noise in image; (g) Similar objects; (h) Complex object motion

- **Occlusion:** It is the state when target is hidden by another object. Target may be partially or fully occluded. Occlusion detection and occlusion handling is a common issue occurring during tracking, but there is no universal technique to tackle it. Therefore, strategies are opted according to the nature of target and environment of tracking.
- **Changing appearance:** Most of the targets, especially non-rigid objects, change their appearance during motion. Therefore, it becomes mandatory for target model to adapt these changes for a long term tracking session. Small inaccuracies may include in target model during updating, which accumulates as time passes and ultimately result in unstable tracking. If the model is made fixed and is not updated, it cannot incorporate changing target appearance and lose the target. Thus, a trade-off is required and it is called stability vs. plasticity dilemma.
- **Cluttered background:** When the background of the target in scene contains many other objects, it is called cluttered environment for the target. If the background of the target is already known (e.g., indoor tracking),

it is easy to handle cluttered environment; but for unknown background or outdoor tracking, the severity of the problem is increased.

- **Changing target size in image:** When the target moves towards or away from camera, its size in image increases or decreases respectively. Therefore, it should be handled by the tracking strategies.
- **Illumination variations:** Illumination variations also important issues, because many features of the target which are prominent in high luminance become obscure in low luminance.
- **Noise in image:** Image of the target scene may be noisy (e.g., electronic circuit noise), so some preprocessing is required to remove the noise from the image.
- **Similar objects:** When there are similar objects near the target, it becomes difficult to track the target.
- **Complex object motion:** When target motion is complex such as out-of-plane movement, or it abruptly changes its speed and direction, e.g., motion of a fighter plane or motion of people during skating, tracking becomes a difficult task due to inexact approximation of motion model.

1.2 Our contribution to the existing surveys

Several surveillance and tracking related surveys can be found in literature as shown in Table 1. Most of these surveys are old (i.e., of last decade), e.g., surveys in Refs. [37,40,42–49]. Some cover only a specific field or technique for tracking (e.g., pedestrian tracking [41], Bayesian method [49], tracking under sea water [43], wavelet for tracking [50]), a few discuss tracking within different principle category (e.g., crowd analysis [48], human motion analysis [44], intelligent visual surveillance [46], appearance model [51]), and the recent surveys discuss only modern trends in VOT (e.g., [39]), or recent algorithms using classical techniques, (e.g., [3,50]). Our survey presents 1) classical as well as contemporary approaches for VOT (shown in Fig. 3), 2) qualitative and quantitative comparison of different tracking algorithms, and 3) online available resources of different algorithms such as source code, annotated data set.

1.3 Paper organization

The rest of the paper is organized as follows. Section 2 describes the classical VOT approaches. Section 3 explains the recent techniques of object tracking. Section 4 discusses different methods for evaluation of tracking algorithms and benchmark resources available online. Finally, Section 5 concludes the paper and provides future directions for VOT.

Table 1 Several related surveys

Related Surveys	Year	Topic
Chau et al. [3]	2013	Tracking any object
Yilmaz et al. [37]	2006	
Joshi et al. [38]	2012	
Yang et al. [39]	2011	
Cannons [40]	2008	
Geronimo et al. [41]	2010	Pedestrian tracking
Ogale [42]	2006	
Trucco et al. [43]	2006	Surveillance and motion analysis
Moeslund et al. [44]	2006	
Aggarwal et al. [45]	1997	
Kang et al. [46]	2007	
Forsyth et al. [47]	2006	
Kim et al. [9]	2010	
Zhan et al. [48]	2008	
Arulampalam et al. [49]	2002	Bayesian tracking
Jalal et al. [50]	2012	Wavelet for object tracking
Li et al. [51]	2013	Appearance Models

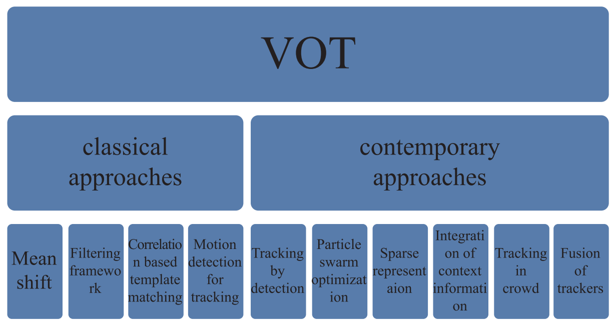


Fig. 3 Different classical as well as contemporary approaches for visual object tracking

2 Classical tracking approaches

In this article, we will describe the following widely known classical approaches for visual object tracking: 1) mean-shift, 2) filtering framework, 3) correlation based template matching, and 4) motion detection based tracking algorithms. Main aim of this section is to highlight different tracking algorithms using aforementioned approaches.

2.1 Mean shift for VOT

Mean shift is non-parametric statistically iterative method. It is originally developed by Fukunaga and Hostetler [52]. It is used to find the mode of a distribution provided its discrete points are given, hence useful in data clustering. Cheng unleashes it to image processing community [53]. It is a very simple and straightforward algorithm. It randomly picks large number of image pixels as representatives of cluster centers. A hypothesized multidimensional ellipsoid is centered on cluster center, and cluster center is moved to the mean of the data lying inside ellipsoid. Similar process is repeated for all the clusters. Mean is iteratively calculated, and cluster centers are moved accordingly until there is no change in mean value. Adjacent and similar regions (similarity depends upon application type and user defined criteria) are merged during iterations and number of final clusters may be much less than the initial number of clusters. Eq. (1) describes the calculation of mean shift vector as given by [54]:

$$\mathbf{m}(\mathbf{x}) = \frac{\sum_{i=1}^{n} \mathbf{X}_i g(\frac{\|\mathbf{X} - \mathbf{X}_i\|^2}{h})}{\sum_{i=1}^{n} g(\frac{\|\mathbf{X} - \mathbf{X}_i\|^2}{h})} - \mathbf{X}, \tag{1}$$

where $g(.)$ is kernel function, x is the centre point and x_i are the data points. Mean shift based schemes suffer from a few drawbacks. They require manual adjustment of system parameters such as the smallest and the largest possible window size, spatial kernel bandwidths. Starting with its application in image segmentation [53,55], mean shift gains popularity in the field of VOT following the research work of Comaniciu et al. [56]. Their paper is awarded as best paper in CVPR 2000 and the technique is filed for patent. In the paper, mean shift is used for real-time tracking of non-rigid objects when viewed using a moving camera. Probability density (histogram) is used to model the target and color is used as feature for tracking. Mean shift algorithm finds the most probable position of target in each upcoming frame. Comparison of probable target candidates with original target model is done using metric based on Bhattacharya coefficient. This work is extended as kernel-based object tracking in Ref. [57]. The proposed scheme is proved to be computationally fast, and robust against clutter, occlusions, camera orientation, and scale changes, in several scenarios, but it is not successful against illumination changes and unpredicted object motion. Moreover, spatial information of target is lost due to use of color histogram as target representative, and Bhattacharya coefficient is also not considered strong discriminative measure [58]. Yang et al. [59] introduce a new similarity measure using RBF kernel that is expectation of the spatially smoothed density estimates over the model image, which improves the robustness and frame rate of tracking. Beleznai et al. [60–64] exploit mode seeking capability of mean shift and apply on difference image for detecting and tracking humans in a video. They use fast version of mean shift for change detection in video. Model based validation scheme is used to approve detected change as humans. Fast mean shift finds clusters in the difference image and updated cluster parameters (e.g., cluster centers) are used for tracking purposes. The idea of mean shift for VOT is extended by Zivkovic et al. [65], and it is used to find out not only the local mode of density function but also for estimation of local mode shape. The algorithm shows robustness for scale changes and adaption of shape, but it is fragile in the case of clutter in background, presence of multiple targets, rapidly changing appearance of target and its motion. Zhou et al. [66] introduce new cost function for tracking of non-rigid objects and improve the performance of mean shift proposed by Zivkovic et al. [65] in complex scenes. Their algorithm optimally adapts ellipse for marking the TOI. The new cost function contains Lagrange base regularization factor which decreases the difference between estimated probability distribution and the expected one. The algorithm shows better results, but it requires more prior information and is computationally slower than that in Ref. [65]. Ning et al. [67] use mean shift approach with joint color texture feature to track the target in complex environment. Shan et al. [68] introduce mean shift with particle filtering for its sampling efficiency. Their work generates good results for rapid motion with less number of particles than that of particle filter alone, but it does not perform well for occlusion and cluttered background. Wang et al. [69] apply mean shift on infrared imagery to track humans. They use motion guided gray and edge cues to improve mean shift results. Their algorithm works only for fixed camera. Following issues arise when using histogram as target representative and mean shift approach for VOT:

- mean shift approach converges locally due to local basin of attraction;
- spatial information is lost due to use of histogram;
- due to global nature of template model, it cannot handle occlusion (even if it is partial) with good accuracy.

The first two issues were handled using different variants of mean shift algorithm such as in Refs. [59,70] and the third one is tackled using fragment based approach. Adam et al. [71] highlight the fragment based VOT to handle the last two of the aforementioned issues with mean shift approach. In spite of using model based patches (e.g., head, limb, torso), fragments are selected randomly. These spatially non-overlapping patches help in preserving the spatial information. Multiple histograms are used to represent each subregion or patch of template and the template position in upcoming image frame is calculated using vote map formed by each patch individual vote. Integral histogram technique is used to make the algorithm efficient. Their algorithm shows robustness to partial occlusion, but lacking the method of selection for different patches. Jeyakar et al. [72] combine fragment based approach with mean shift. User is taken into the loop by making selection of fragments manually. The patches may be overlapping or non-overlapping. Bhattacharya coefficient based metric is used for similarity measure. The algorithm shows impressive results in case of partial occlusion; it also handles illumination, appearance and scale changes as well as clutter in the background. Khan et al. [73] use normalized correlation as similarity measure for occluded and clutter imagery. Template is partitioned into nine non-overlapping fragments. Table 2 summarizes the comparison of different algorithms using mean shift approach. It is clear from this table that Wang et al. [69] handle most of the issues.

2.2 Filtering framework

In this Section, we will discuss different filters used for target tracking. Seminal Kalman filter, is a statistical parametric recursive algorithm specially designed for discrete time system. This filter is based on motion model of linear dynamic system; therefore, it requires its state space representation as shown in Eq. (2) and Eq. (3) [74]:

$$\mathbf{X}_{n+1} = \mathbf{\Phi}\mathbf{X}_n + \mathbf{U}_n, \tag{2}$$

$$\mathbf{Y}_n = \mathbf{M}\mathbf{X}_n + \mathbf{V}_n, \tag{3}$$

where $\mathbf{X}_n$ symbolizes the state vector, $\mathbf{\Phi}$ represents state transition matrix, $\mathbf{U}_n$ denotes the system noise vector, $\mathbf{V}_n$ is the observation noise vector, $\mathbf{Y}_n$ is the measurement vector, and $\mathbf{M}$ shows the observation matrix. KF estimates the states of the dynamic system in the presence of 1) noisy measurement (Gaussian noise), and 2) uncertainty in the model of dynamic system. It works in prediction-correction cycle format. KF, based on observed (measured) states, corrects its predicted states as well as update its gain matrix for better future predictions as described by Eqs. (4–9) [75–77].

$$\mathbf{X}^*_{n|n} = \mathbf{X}^*_{n|n-1} + \mathbf{K}_n(\mathbf{Y}_n - \mathbf{M}\mathbf{X}^*_{n|n-1}), \tag{4}$$

where $\mathbf{X}^*_{n|n}$ represents the posteriori measurement, $\mathbf{X}^*_{n|n-1}$ the prior measurement, and $\mathbf{K}_n$ the *Kalman gain matrix* defined as:

$$\mathbf{K}_n = \mathbf{S}^*_{n|n-1}\mathbf{M}^T[\mathbf{R}_n + \mathbf{M}\mathbf{S}^*_{n|n-1}\mathbf{M}^T]^{-1}, \tag{5}$$

where $\mathbf{R}_n$ is the *observation noise covariance* calculated by Eq. (6), $\mathbf{S}^*_{n|n-1}$ represents the *predictor error covariance* defined by Eq. (7).

$$\mathbf{R}_n = COV(\mathbf{V}_n) = E[\mathbf{V}_n\mathbf{V}_n^T], \tag{6}$$

where $E[.]$ is the expected value.

$$\mathbf{S}^*_{n|n-1} = COV(\mathbf{X}^*_{n|n-1}) = \mathbf{\Phi}\mathbf{S}^*_{n-1|n-1}\mathbf{\Phi}^T + \mathbf{Q}_n, \tag{7}$$

$$\mathbf{S}^*_{n-1|n-1} = COV(\mathbf{X}^*_{n-1|n-1}) = [\mathbf{I} - \mathbf{K}_{n-1}\mathbf{M}]\mathbf{S}^*_{n-1|n-2}, \tag{8}$$

where $\mathbf{Q}_n$ is the *noise covariance matrix* and is calculated by Eq. (9):

$$\mathbf{Q}_n = COV(\mathbf{U}_n) = E[\mathbf{U}_n\mathbf{U}_n^T]. \tag{9}$$

Interested reader is referred to [78] for derivation of KF equations. In VOT, KF is widely used in conjunction with other algorithms [57,78–89]. During the tracking session, KF normally acts in two modes: 1) *normal tracking mode*, in which KF predicts the next target coordinates in image plane on the basis of measurement to define the search window with

Table 2 Comparison of different VOT algorithms using mean shift

Representative work	Target representation	Similarity measure	Issues				
			S/M	O	IV	SV	SC
Comaniciu et al. [56,57]	Color histogram	Bhattacharya Coefficient	S	√	×	×	√
Yang et al. [59]	Joint Spatial-feature space	Expectation of density estimates	S	√	×	√	√
Beleznai et al. [60–64]	Difference image	No similarity measure	M	√	√	√	×
Zivkovic et al. [65]	Color histogram	Expectation Maximize like Algorithm	S	√	×	×	√
Zhou et al. [66]	Color histogram	Expectation Maximize like algorithm with ellipse outlining the target	S	√	×	×	√
Ning et al. [67]	Joint color-texture histogram	Bhattacharya Coefficient	S	√	×	×	√
Shan et al. [68]	Motion color	Distance function	S	√	√	√	×
Wang et al. [69]	Motion and gray edge cues	Bhattacharya Coefficient	S	√	√	√	×
Adam et al. [71]	Fragment base histogram representation	Earth Mover's Distance	S	√	×	×	√
Jeyakar et al. [72]	Fragment base representation	Bhattacharya Coefficient	S	√	√	×	√
Khan et al. [73]	Fragment base edge representation	Normalized correlation	S	√	√	×	√

Note: S/M — single target or multiple target, O — occlusion, IV — high illumination variations, SV — sudden and large change in target velocity, SC — scale change. Symbols √ and × respectively shows that algorithm does or does not handle the issue.

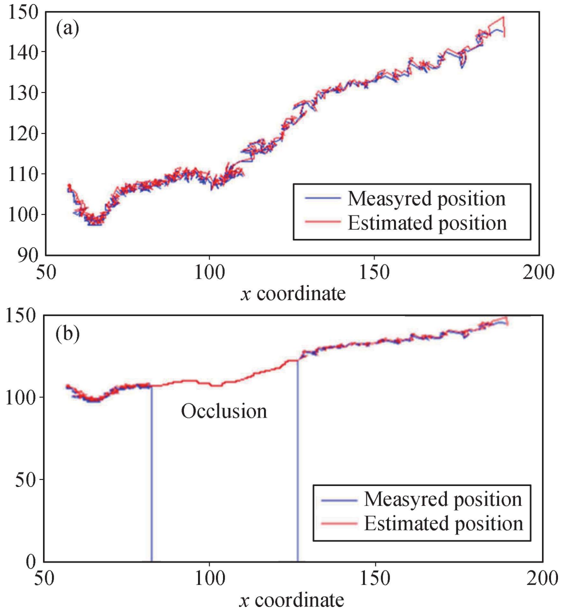


Fig. 4 (a) Normal tracking, and (b) critical issue of occlusion and Kalman Filter successfully handling the issue

optimum [83], and 2) *occlusion mode*, in which KF ignores the measured value and uses its predicted value for next state prediction. Thus, it is used to handle short-term occlusion. Figure 4(a) [90] shows the normal tracking mode and Fig. 4(b) illustrates the occlusion mode during tracking. It may be noted that KF successfully handles the issue of occlusion. Ahmed et. al. [83] combine KF with normalized correlation to handle short-term occlusion, as well as to find the optimum position and size of the search window for the next frame. Jang et al. [91] use KF for target motion prediction in order to reduce search space for matching the target. Comaniciu et al. [79] combines KF with mean shift tracker but it cannot handle the large movement of target. Ali et al. [82] use correlation and fast mean shift algorithms with KF to handle complex maneuvered motion of an air borne object. Li et al. [80] use KF with mean shift and fast motion estimation algorithm to handle the large and sudden movement of target. Li et al. [81] use Bhattacharya Coefficient for adjusting the KF estimation parameters adaptively. Their results show robustness against partial or full occlusion, fast target motion, and sudden changes in the target velocity. Ridder et al. [92] use KF for discriminative tracking approach. They modeled each pixel with KF in order to handle the variation in illumination. In this way, KF is used for adaptive background estimation and foreground detection. Peterfreund [93] use KF with active snakes for robust tracking of position and velocity of non-rigid as well as rigid objects. He uses image gradient along the contour, and its optical flow is used as system measurement. KF assumption linear dynamic model of target motion and Gaussian noise in measurement are not always true in real world. Therefore, its different variants such as extended KF (EKF), unscented KF (UKF) are introduced [94]. EKF apply first order Taylor series to approximate a non-linear system as the linear one. UKF does not apply such approximation, but uses unscented transformation, and generates a set of sigma points which are transferred to dynamic model of state and observation for final result. UKF results are better than that of EKF, but it assumes Gaussian distribution for posterior. In this way, UKF cannot work in case of multi-model distributions. Therefore, particle filter (PF) [49] is used to cater these issues. PF is a non-parametric Monte Carlo simulations based method [95]; it is used in tracking application first by Isard and Blake [96] with the name of Condensation. PF represents state of target by a set of weighted particles. The weight to each particle is assigned according to its contribution in finding the target's location. The position of each particle is updated according to the motion model and the measurement data. PF suffers from the problem of sample improvishment in which samples contribute no useful information in estimating the target position. Interest readers may refer [49,97] for further detail. Table 3 briefly describes different VOT approaches exploiting KF. It is clear from the table that the performance of method proposed by Li et al. [81] is better than other algorithms in case of gray scale imagery. For color images, Wang et al. [69] performs better.

Table 3 Comparison of different VOT approaches exploiting KF

VOT approaches exploiting KF	Representative work	Issues			
		OS	O	LM	SV
Mean Shift and KF	Comaniciu et al. [79]	√	√	×	×
	Jang et al. [91]	√	×	×	×
	Li et al. [80]	×	×	√	√
	Li et al. [81]	×	√	√	√
Correlation and KF	Ali et al. [82]	×	√	√	√
	Ahmed et al. [83]	√	√	×	×
Background / Foreground Detection and KF	Ridder et al. [92]	×	×	√	√

Note: OS — optimum search, O — occlusion, LM — large target movement, SV — sudden change in velocity. Symbol √ stands for tracking algorithm handles the issue and symbol × means it does not tackle the issue.

2.3 Correlation based template matching

Template matching or correlation tracking is the classical method in the field of VOT. Its history dates back to the beginning of computer vision field [35,36,98]. The process of tracking is started by selecting the target in the first frame manually or by some automatic target detection system. The representation of the target is called template which is used to locate the target by correlating it with the video frame in each iteration. The location with the highest correlation score is considered as the new target position. Different correlation metrics, e.g., standard correlation (SC) [99] (Eq. (10)), phase correlation (PC) [100] (Eq. (11)), normalized correlation (NC) [99] (Eq. (12)), normalized cross correlation (NCC) [101,102] (Eq. (13)) are usually used as similarity measure in tracking applications. Details of these metrics can be found in Ref. [83].

$$c(m,n) = \sum_{i=0}^{K-1}\sum_{j=0}^{L-1} f(m+i, n+j)t(i,j), \tag{10}$$

$$c = real[idft(\frac{F}{|F|}.\frac{T^*}{|T|})], \tag{11}$$

$$c(m,n) = \frac{\sum_{i=0}^{K-1}\sum_{j=0}^{L-1} f(m+i, n+j)t(i,j)}{\sqrt{\sum_{i=0}^{K-1}\sum_{j=0}^{L-1} f^2(m+i, n+j)}\sqrt{\sum_{i=0}^{K-1}\sum_{j=0}^{L-1} t^2(i,j)}}, \tag{12}$$

$$c(m,n)=\frac{\sum_{i=0}^{K-1}\sum_{j=0}^{L-1}[f(m+i,n+j)-\mu_s][t(i,j)-\mu_t]}{\sqrt{\sum_{i=0}^{K-1}\sum_{j=0}^{L-1}[f(m+i,n+j)-\mu_f]^2}\sqrt{\sum_{i=0}^{K-1}\sum_{j=0}^{L-1}[t(i,j)-\mu_t]^2}}, \tag{13}$$

where f is the image, t is the template, F and T are their Fourier transforms, T^* is conjugate of T, $idft(.)$ is inverse discrete Fourier transform operator, $real(.)$ extracts real part of its operand, μ_f and μ_t shows mean of image and template respectively.

Table 4 Comparison of different correlation metrics

No.	Correlation Type	Discriminatory power	Robustness to noise
1	Standard correlation	Poor	Poor
2	Phase correlation	Strong	Poor
3	Normalized correlation	Strong	Strong
4	Normalized cross correlation	Strong	Strong

SC does not have any bounding value, so, no threshold value can be set to validate the match score and update the template. Moreover, it is sensitive to illuminate and produce peak value at the brightest spot in the image. PC computes correlation in the Fourier domain. It is insensitive to variations in image intensity because it ignores the Fourier magnitude and calculates the phase component only. PC has strong discriminatory power and produces sharp peak, but it is not robust to noise as compared to SC [103]. Moreover, it assigns equal weight to all of its components, which looks inappropriate as significant components should be assigned more weights than other components [101]. Due to these discrepancies, PC may yield false positive [83,104,105]. Different variants of PC are also proposed [106–108], yet they are not as robust to variation in appearance, illumination, and contrast as NC and NCC. These two metrics have their values in the ranges [0, 1] and [−1, 1] respectively, so, it is easy to set a threshold for template updating and occlusion handling. Updating the template is mandatory for tracking an object when changing its appearance. Ali et al. [82] update the template completely in every next frame, if the peak correlation value is higher than a threshold. The updating scheme suffers from fast template drift problem, if newly found template position is not the exact position. To handle this problem, Ahmed et al. [83] update the template smoothly using first order IIR filter. Most of the times, NCC is used as similarity measure in image registration [99,101,109,110], but NC performs better than NCC when edge-enhancement is performed as preprocessing step of target tracking [83]. Ali et al. [111] use NC for edge enhanced template with a new updating method which considers the rate of appearance of target while updating the template. Their algorithm uses adaptive threshold, as compared to fixed threshold used by Ahmed et al. [83] and the algorithm works better than other discussed correlation methods. Asgrizadeh et al. [78] integrate region mutual information (RMI) with edge correlation tracking for more robust tracking of aerial objects. RMI provides information about the clutter and clear backgrounds as well as high luminance changes. Table 4 shows the comparison of above-mentioned correlation metrics.

2.4 Motion detection for tracking

There are various methods for motion detection including background subtraction, temporal difference, background modeling, and optical flow.

2.4.1 Background subtraction

Target or area of interest in a scene is referred to as foreground, and anything else in the image is termed as background. Background subtraction or foreground detection may be used for two purposes. Firstly, it may be used to initialize the tracking, and secondly, it is used to detect the target of interest from frame to frame. The simplest method for foreground detection is to subtract each frame from a fixed background in case of stationary background. The pixels corresponding to background will yield a very low value, and the pixels related to foreground will create high values in the subtracted image. Thus, a threshold can be set to distinguish the foreground pixels from the background pixels. Connected component algorithm is used to group the foreground pixels. After detecting foreground, we search the target only at the foreground regions, so as to get rid off exhaustive search, which improves efficiency of the algorithm. This straightforward method of background subtraction normally works in structured environment, and but fails in case of un-structured or outdoor environment where illumination and background does not remain stationary.

Since fixed background does not work in case of outdoor, adaptive background model is adopted. Wren et al. [112] uses a uni-model Gaussian representation for each pixel using its mean and variance in YUV color space. Their algorithm works well in order to handle small illumination changes but it does not show efficacy in case of sudden illumination changes (e.g., flashing light, swaying trees or bushes, moving fountains, and rotating fins of a fan). These issues are handled by the work of Stauffer et al. [33,113,114]. They adopt multi-model Gaussian representation for each pixel, and update the models online to learn the changing background. Usually, three to five models are used to model each pixel distribution. If match of a current pixel is found with any Gaussian distribution, it is considered as background, otherwise, it is classified as foreground pixel, and background model is updated according. This multi-model Gaussian approach does not tackle the problems of drastic illumination change, and moving shadows. KaewTraKulPong et al. [115] improve the learning rate Guassian mixture model and introduce shadow detection method. Their algorithm compares the foreground pixel with the background model. If the difference between chromatic and brightness components is within certain threshold, it is considered to be a part of shadow. Similar technique was presented by Horprasert et al. [116,117]. Haritaoglu et al. [12] develop a real-time surveillance system that trains background model by three features, i.e., minimum pixel value (m), maximum pixel value (n), and maximum intensity difference between consecutive frames (d). A pixel is classified as foreground, if its difference from m or n is greater than d, otherwise, it is taken as background pixel. Oliver et al. [119] use principle component analysis and eigen decomposition to build eigen background. The project image of the current frame is subtracted from the eigen background to detect foreground objects.

2.4.2 Temporal differencing

Temporal differencing means subtraction of previous frame from current frame to detect any change or moving objects in the scene. Lipton et al. [119] use temporal differencing between two consecutive frames for foreground detection. They adopt multiple hypothesis for classification of foreground regions as targets of interest. The classification metric employs perimeter and area to identify the targets in the difference image. In order to improve the detection of foreground regions, three inter-frames temporal difference schemes are also be used [120,121]. Temporal difference methods are sensitive to the threshold as well as illumination changes. Moreover, when the target stops moving, it cannot be detected as foreground object.

2.4.3 Optical flow

Optical flow is the apparent motion pattern in an image of a scene due to relative motion between objects of the scene and the camera. The calculation of optical flow assumes brightness consistency between corresponding pixels in the scene. There are various methods for calculating dense optical flow in an image such as Lucas and Kanade [35], Horn and Schunck [122], Black and Anandan [123], and Szeliski and Couglan [124]. Optical flow is used as a feature in segmentation, tracking applications based on motion of objects. Shi and Tomasi [125] exploit optical flow to find out the motion of a region in an image and developed their well known KLT tracker. The tracker is sensitive to illumination changes and large frame motion. Rangarajan and Shah [126] use optical flow to find initial inter-frame correspondence between the first two frames to their proposed greedy search algorithm. Papageorgiou et al. [127] use optical flow to reduce search space of their SVM based pedestrian and face detection algorithm. Cremers et al. [128] use optical flow as a feature in contour based tracking algorithm. Li et al. [129] use optical flow for silhouette tracking. Bertalmio et al. [130] and Mansouri [131] use optical flow for the minimization of contour energy.

3 Contemporary tracking approaches

In this Section, we will investigate the recent approaches for VOT which includes 1) tracking by detection, 2) sparse representation, 3) particle swarm optimization, 4) integration of context information, and 5) tracking in crowd.

3.1 Tracking-by-detection

Tracking-by-detection includes the class of algorithms which considers tracking phenomenon as detection process of target in consecutive image frames by training a binary classifier to discriminate the target from its background. The class of algorithms is termed as tracking-by-detection or tracking-by-repeated -recognition algorithms [132]. These methods have gained popularity in recent years due to their efficacy in performance and simplicity of classification task [26,133–138]. Detail discussion on different classifiers is not in scope of this paper, interested readers referred to Refs. [139,140] for further study. Normally, a classifier requires training data for its performance, no prior knowledge is available about target position in case of tracking application. Therefore, training data is generated online during tracking, and classifier is updated accordingly. It is called Adaptive tracking-by-detection. Collin et al. [134] present an approach for online selection of features to discriminate the target from its background. The estimated position of target in each frame is considered as positive example and its nearby locations are treated as negative examples for updating the classifier. This step is called Generation and Labeling of Samples [141]. During tracking, classifier finds the target position by maximizing the classification score in a local region, normally around the target position found in the previous frame, using sliding window method. Figure 5 explains this tracking and updating process. Avidan [26] introduces support vector tracking (SVT) which integrates support vector machine (SVM) classifier with optical flow for vehicle tracking. Grabner et al. [2,133] present online version of AdaBoost approach for real-time tracking. The tracking approaches [26,133,134] update their classifiers by considering a single only positive example consisting of current position of target, and many negative examples, i.e., samples around the current target position as shown in Fig. 6. Small inexactness in the target position results in poorly labeled training samples, which is called label jittering that abates the performance of classifier and ultimately causes the drift problem. Therefore, most of the recent tracking-by-detection approaches try to improve tracking performance by making the classifier more robust to incorrectly labeled examples [132,143–147]. Babenko et al. [132] present online multiple instance learning boosting (Online MILBoost) algorithm for robust tracking. Instead of assigning label to each individual example, their algorithm combines instances into bags and label is assigned to each bag. A positive bag should contain at least one positive example; otherwise, negative label is assigned to it as shown in Fig. 7. Their algorithm shows prominent results against drift issue, but it fails to recapture the target if it gets out of scene and returns. This is the problem with all adaptive appearance model based algorithms that they start updating themselves with the false object, if the target is fully occluded or it gets out of scene for a while. Grabner et al. [143] solve this issue by using semi-supervised appearance model updating method. Their method combines the labeled data (prior knowledge), i.e., the target selected by the user in the first frame, with current unlabeled data. In this way, the method becomes robust to drift issue, but it shows less adaptability to appearance changes. Zeisl et al. [144] combine the strength of semi-supervised and multiple instance learning into a single framework. Zhang et al. [148] improve the work of Babenko et al. [132] by assigning different weights to different instances. The closer the instance to the target, the higher the weight assigned to it. William et al. [149] point out that the highest classification score does not necessarily belong to the target position, as there is no explicit relationship between classification confidence and the target spatial position.

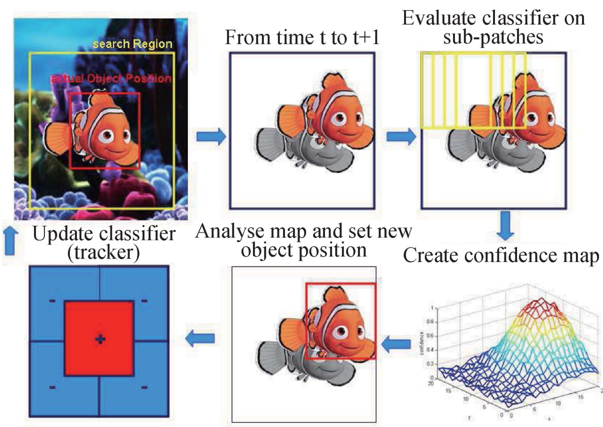


Fig. 5 (Source [142]): adaptive tracking-by-detection process, i.e., tracking the target and updating the classifier

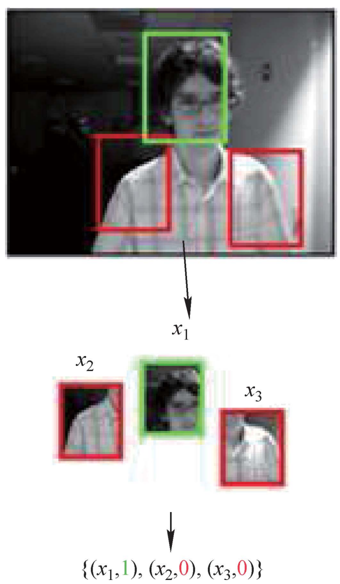


Fig. 6 Positive and negative samples for online AdaBoost [134]

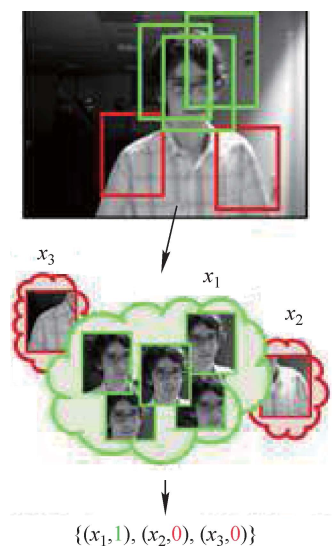


Fig. 7 Positive and negative bags for MIL classifier [133]

Hare et al. [142] present a framework based on structured output prediction which explicitly incorporates the tracker need of labeled training examples into the output space. Instead of learning classifier, the frame work focuses on estimating the target transformations by using structured output SVM. Thus, it gets rid of intermediating step of Generation and Labeling the Samples. Table 5 summarizes the representative work of tracking-by-detection technique, and Table 6 gives quantitative measures (Pascal score) of different tracking algorithms on publicly available video sequences. It is clear from this table that the tracking method proposed by Zhang et al. [148] works better than the rest of the algorithms. The last row of the table shows the frame process per second of each method in MATLAB at system with CPU 2.1 GHz and 2 GB RAM.

Table 5 Representative work of tracking-by-detection technique

No.	Representative work	Discriminatory technique
1	Collins et al. [134]	Selection of features by ranking
2	Hare et al. [141]	Structured SVM learning
3	Avidan [26]	SVM learning
4	Grabner et al. [133]	Boosting by selection of features using feature-ranking
5	Babenko et al. [132]	Boosting by multiple instance learning
6	Grabner et al. [143]	Semi supervised boosting
7	Zhang et al. [148]	Boosting by weighted multiple instance learning
8	Zeisl et al. [144]	Semi-supervised multiple instance learning

Table 6 Performance comparison of different algorithms

Videos	Zeisl et al. [144]	Babenko et al. [132]	Zhang et al. [148]	Grabner et al. [143]	Adam et al. [71]	Hare et al. [142]
Sylvester	0.63	0.61	0.73	0.67	0.44	0.68
David	0.71	0.54	1	0.45	0.08	0.80
Faceocc	0.68	0.71	0.97	0.96	1	0.86
Faceocc2	0.78	0.65	0.97	0.44	0.52	0.86
Tiger1	0.60	0.51	0.66	0.29	0.19	0.70
Tiger2	0.46	0.50	0.61	0.19	0.13	0.56
FPS	1.5	1.5	27	6	3	12

3.2 Particle swarm optimization

Particle swarm optimization (PSO), inspired by the birds searching for food, was first introduced by Kennedy et al. in 1995 [150,151]. Since then, its use in different applications is increasing day by day, and it has drawn attention of researchers in different fields [152–154]. PSO is a stochastic process exploiting the phenomenon of swarm intelligence and works on collective wisdom which prevails by each of its particles. Each particle in PSO updates its position by considering its own best position as well as its neighborhood best position until all the particles finds a common converging position or the maximum number of iterations are completed . Neighborhood may be the entire swarm around the current particle. An objective function is required to calculate the fitness map of each particle. The position of a particle with the highest value on fitness map is considered as its best position, and the best of these best positions is the global or swarm best position. PSO has very simple formulation consisting of velocity and position update equations described by Eq. (14) and Eq. (15).

$$v_d^i(n+1) = wv_d^i(n) + c_1r_1(p_d^i - x_d^i(n)) + c_2r_2(g_d - x_d^i(n)), \quad (14)$$

$$x_d^i(n+1) = x_d^i(n) + v_d^i(n), \quad (15)$$

where $v_d^i(n)$ and $x_d^i(n)$ represent the velocity and position of i^{th} particle in d-dimension at iteration n, repectively; p_d^i is the particle personal best position; g_d is the global best position of the swarm; w is the inertia weight; c_1, c_2 are constants; r_1 and r_2 are random values have values in the range of [0, 1]. Different variants of PSO and their applications can be found in Refs. [152,155,156]. In visual tracking applications, PSO is used to search for the best candidate position of the target in the current frame. Zhang et al. [157] introduce the sequence of temporal information of the target into PSO and name it as Sequential PSO. They present particle filter based tracking algorithm with hierarchical importance sampling process guided by the sequential PSO. Thus, their approach helps to cope with the classic sample impoverishment problem of particle filter. Zhang et al. [158] propose species based PSO for tracking of multiple objects. Each species is used to track the individual object. Thus, different trackers run under single framework. The inter-object occlusion is handled by species competition and repulsion. The number of objects and hence species are initialized by user at the very beginning of the tracking process. Akbari et al. [159] combine PSO and KF to track multiple objects in cluttered environment. They use non-overlapping fragments based representation for objects. Each fragment is represented by a particle. The particles of PSO are guided by KF in a hybrid framework using region as well as object information. Kwolek et al.

[160] propose a multi-object tracking algorithm which uses PSO to improve the targets, positions found by discriminative appearance models. The objective function is based on fragments based representation of targets and their covariance matrix. Anton-Canalis et al. [161] use PSO for tracking target in predator-prey style. Each particle directly interacts with a pixel, and tracking is performed by interaction of particle with its environment. Zheng et al. [162] represent the target into multi-dimensional feature space and employ PSO algorithm to expedite the search process. Bhattacharya coefficient is used as fitness function for PSO in their algorithm to track for faces and vehicles. Tawab et al. [163] propose PSO-based fast gray level object-tracking. They employ combination of SIMilarity (SIM) and Bhattacharya coefficients as fitness functions to evaluate the score of particles of PSO. Borra et al. [164] propose PSO-Fuzzy C Means (PSO-FCM) based tracking algorithm. PSO-FCM is used to segment the objects in the scene, and a pattern matching approach is used to track the target. Table 7 summarizes the representative work of PSO in VOT.

Table 7 Representative work of using variants of PSO in VOT

No.	Representative work	PSO variants
1	Zhang et al. [156]	Sequential PSO
2	Zhang et al. [157]	Species PSO
3	Akbari et al. [158]	Standard PSO
4	Kwolek et al. [159]	Standard PSO
5	Anton-Canalis et al. [160]	Predator-Prey PSO
6	Zheng et al. [161]	Standard PSO
7	Tawab et al. [162]	Standard PSO
8	Borra et al. [163]	PSO-FCM

3.3 Sparse representation

Compressive or sparse representation [164, 165] of a signal exhibits the signal as linear combination of small number of basis vectors. The representation is becoming popular in various pattern recognitions and image processing applications [166–169]. Mei et al. [170–172] use sparse representation for object tracking. The algorithm is capable to cope with occlusion problem using trivial templates and λ_1 minimization approach for sparse representation. The algorithm is expensive with respect to computations due to λ_1 minimization algorithm. Moreover, the trivial templates are used to model targets as well as background which make reconstruction error small for both of the regions. The candidate region with minimum reconstruction errors is considered the target. Thus, a confusion is generated while tracking. Liu et al. [173] improve tracking efficiency and robustness by exploiting sparseness and using a set of discriminative features. The algorithm uses fixed number of features, so it is not effective in complex or dynamic environment. Liu et al. [174] use mean shift and histograms based local sparse representation for appearance model. However, histogram representation is unable to distinguish between the targets and the background due to its inherent problem of lost spatial information. Zhong et al. [175] make the object tracking robust to occlusion by collaboration of discriminative sparsity-based discriminative classifier (SDC) and sparsity-based generative model (SGM) as well as generative tracking capabilities. SDC assigns higher confidence to foreground objects than background objects. SGM proposes a new method to calculate histograms which accounts for spatial position of each patch as well. Jia et al. [176] propose a structured sparse representation for appearance model for target tracking. The algorithm proposes an alignment-pooling method for partial as well as spatial information in order to tackle the occlusion problem. Moreover, the algorithm proposes a novel template updating method based on incremental subspace and sparse representations. Present tracking algorithms update appearance model using current image frame. Therefore, these methods are data-dependent. Zhang et al. [177] exploit multi-space features for appearance modeling based on data-independent basis. The algorithm uses random projections to protect the feature space of objects in image. Sparse representation is employed to extract the features. Detail review and experimental comparison of sparse coding based visual tracking may be studied from the paper of Zhang et al. [178].

3.4 Integration of context information

Integration of context with the target of interest for robust object tracking has gained significant importance in recent years. Various psychophysics studies have emphasized the role of context in image understanding for human perception system [179]. Interested readers may study [180] for the role of context in object detection. Yang et al. [181] propose a number of nearby objects to the target as spatial context to enhance the appearance model. These objects are automatically extracted from the video during run-time and are named as auxiliary objects. Auxiliary objects are chosen at least for short time interval according to following criterion: 1) straightforward to track, 2) persistent co-occurrence with the target, and 3) consistent motion correlation with the target. Li et al. [182] model contextual relationship by dynamic Markov random field to simultaneously recognize, localize, and track multiple objects of different categories in meeting room videos. Spatio-temporal relationship is used to get information of object category and its state. Nguyen et al. [183] use spatio-temporal context for multi-target tracking. Spatio context includes nearby objects and temporal context contains all previous target models based on probabilistic principle components analysis (PPCA). Wen et al. [184] also propose spatio-temporal context relationship for robust object tracking. Grabner et al. [185] use Hough transform to integrate temporal context (supporter) and distinguish between strong and weak coupled motions. Their algorithm works well in case of full occlusion, and when target changing its appearance heavily and rapidly. Table 8 summarizes the representative work of exploiting context information for VOT.

Table 8 Representative work of exploiting context information for VOT

No.	Representative work	Contextual information
1	Yang et al. [181]	Spatial position
2	Li et al. [182]	Spatio-temporal relationship
3	Nguyen et al. [183]	Spatio-temporal relationship
4	Wen et al. [184]	Spatio-temporal relationship
5	Grabner et al. [185]	Temporal context

3.5 Tracking in crowd

Tracking has gained importance in the last few years for visual analysis of crowd [186, 187] due to ubiquitously available security cameras, dire need of activity monitoring in sensitive public places (e.g., airport, railway station), and crowd density count (e.g., the number of protestors in D Chowk Islamabad during sit-in of Imran Khan, Chairman Pakistan Tehreek-e-Insaf). Interested readers may refer to Ref. [48] for further study of crowd analysis. Tracking strategies for un-crowded environment may not work for crowd because of the increased tracking complexity, e.g., the number of pixels on target decreases with the crowd density increasing, many inter objects occlusions, and cluttered background with similar objects [188]. The algorithm proposed by Zhao et al. [189] is one of the first methods for tracking in crowded environments. Their method is object-centric which tracks multiple pedestrians by modeling human shapes with articulated ellipsoid, and the background with a Gaussian distribution. Color histogram was used to model the appearance of pedestrians. It is impossible to fit ellipsoidal models of the human body in high density crowd, as complete human body is not visible. Therefore, the method is not suitable for dense crowd. Other object-centric tracking techniques in crowd, such as the work by Betke et al. [190], Li et al. [191], and Wu et al. [192], use object-detection to create short trajectories which are merged into the final tracking results based on a global model of all of the objects within the scene. The performance of the underlying object detection method in such methods is deteriorated in high-density crowd due to frequent inter objects occlusion. Brostow et al. [193] present a probabilistic method for tracking of individuals in crowds. They detect pedestrian by clustering low-level image features. They assume that pairs of points which move together are part of the same individual. If the individual tracking is camouflaged, features would not be grouped resulting in astray tracking. Ali et al. [188] propose force model based on scene structure to track objects in high density structured crowds. The force model considers that the objects in the scene bear global and local forces which are function of layout of scene and locomotive of neighborhood objects on individuals being tracked. Their model consists of three floor fields, which are static floor field (SFF), dynamic floor field (DFF), and boundary floor field (BFF). SFF specifies exit locations; DFF is about the influence of neighborhood objects on individual being tracked; BFF describes the hurdles to the crowd, e.g., wall, no-go area. As these fields are prior learned for tracking, the method fails in the situations where crowd flow is dynamic or moves to a new region. Pellegrini et al. [194] use social force model and propose linear trajectory avoidance to predict motion patterns of individuals in crowd. Their method requires strict offline parameter tuning of each test video. Rodriguez et al. [195] introduce a tracking method for un-structured crowded environment by employing a correlated topical model to represent multimodal motion at each spatial location. Their method assumes a fixed number of motion directions at each spatial location as a quantized optical flow vector, but neglects the temporal relationship between sequentially occurring local motions. This method also requires prior learning of crowd behavior similar to that proposed by Ali et al. [188]. Krate et al. [196] method also assumes prior learning of crowd behavior by considering 3D Guassian and hidden Markov models (HMM) at each spatio-temporal location. The problem is handled by recent work of Rodriguez et al. [197]. Their algorithm proposes to learn crowd behavior by a data-driven approach which considers that distinguishable crowd motion patterns are finite and small in number. Their method works by learning the crowd behavior using a large database of 500 crowded videos, therefore, it does not require the test video of crowded scene at prior. The patches of test videos are matched with the database to find the similar motion patterns. Idrees et al. [198] propose to use visual and contextual information of a crowded scene for online tracking, i.e., without any prior learning. Their method uses neighborhood motion [199] structure to preserve multi-object tracking by modeling their spatial constraint, and requires manual selection of patches with fixed structure during tracking process. Zhu et al. [200] propose to track distinct and stable mid-level patches jointly with dynamic evolution of group structure. Their method generates patches with stable internal motion, automatically, by low-level keypoint tracking which also hierarchically organizes patches using collective motion. Table IX represents the Pascal score for Zhu et al. [200], Zhang et al. [199], and Idrees et al. [198] for video sequences (Traffic, Crowds, Marathon, Split, Merge and Cross) which can be downloaded from http://home.ustc.edu.cn/ zhufengx/crowdTracking/. It is obvious from these results that the algorithm proposed by Zhu et al. [200] outperforms the rest of the algorithms.

Table 9 Pascal score for different methods in the field of tracking in crowd

	Zhu et al. [200]	Zhang et al. [199]	Idrees et al. [198]
Traffic	0.99	0.81	0.74
Crowds	0.90	0.64	0.81
Marathon	0.88	0.02	0.70
Split	0.86	0.25	0.70
Merge	0.82	0.57	0.77
Cross	0.89	0.27	0.69

3.6 Fusion of trackers

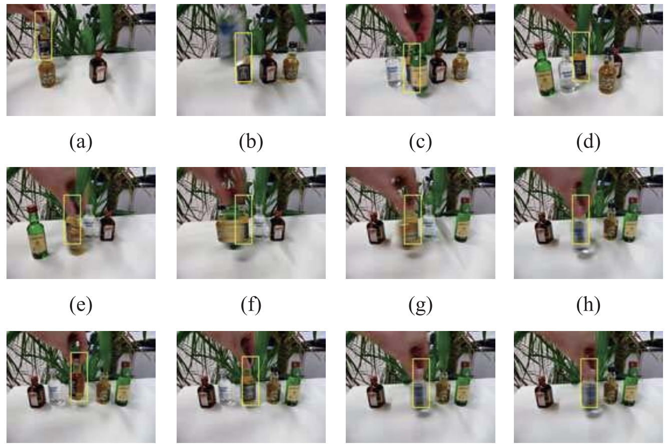
The long-term persistence tracking is still a challenging task due to absence of any prior information about the target and its surroundings. There is not a single tracker which can work well in all circumstances, therefore, the idea of using multiple trackers is propagating rapidly. The results of different trackers are combined to enhance the overall performance of target tracking. Ali et al. [111] combine, heuristically, the strength of correlation, KF, and mean shift trackers and minimize individual tracker's weakness. Correlation tracker has inherent problem of template drift, and it does not work in case of occlusion. KF tracker counters this issue and tracks the target during occlusion by predicting the next spatial location of the target. The problem with KF filter is that it is a "measurement follower", if due to some reason, the correlation tracker measurement gets wrong, KF prediction may not be reliable. Mean shift addresses this issue and finds the target location. The outputs of all three trackers are combined, and decision of target location is made on majority voting criteria. Gao et al. [201] propose a symbiotic tracker ensemble framework in which different trackers are run in parallel as black boxes, i.e., only input and output are taken into account without bothering about individual trackers' detail. Their algorithm focuses on learning an optimal combination of these tracking results. The relationship among individual trackers is considered with respect to two aspects: 1) consistency between two successive frames is calculated for each tracker, and 2) pair wise correlation among different trackers is calculated in upcoming frame by graph propagation process. Zhong et al. [202] tackle the template drift problem by considering visual tracking in a weakly supervised learning scenario where labels are provided by multiple imperfect trackers. Target position and accuracy of each tracker are simultaneously inferred by their proposed probabilistic framework. Yao et al. [203] use the association of kernel-based object tracking and particle filter. Particle filter is used for multi-modal visual tracking problem with non-Gaussian distribution, but it requires a large number of particles which make it inefficient especially for real time application. Kernel-based object tracking is efficient and simple to implement technique with good tracking results, but it is unable to handle multi-model problem that arises due to similar objects, and clutter in background. It is clear from this discussion that Kernel tracking and PF can complement each other. Therefore, their algorithm proposes a constrained gradient-based mean shift optimization which makes it possible to efficiently refine the particles' position states.

4 Evaluation methods for VOT algorithms and benchmark resources

VOT algorithms are evaluated qualitatively as well as quantitatively. For qualitative comparison, sample image frames are shown and visually examined. The visually better results are considered as those which have tracked rectangle closer to the target of interest as shown in Fig. 8. Qualitative analysis does not provide fair comparison between different algorithms. Therefore, quantitative solution is calculated to have better understanding of robustness of the algorithms. For this, two measures are employed. One is the mean distance from centre location, which provides the error between center location of tracked rectangle and its ground truth value. The overall performance of the algorithm is summarized by computing the mean of the center location errors for all the frames in a video. The problem of this method is that if a particular method successfully tracks a target in most of the video frames or loses the target in a few frames with a large distance, its performance will be poor in comparison with a method which mostly does not track the target but clings to the background nearby the target. Therefore, this method is not a true representative for performance of a particular method. One modification to this method is made by Babenko et al. [131] and Henriques et al. [204]. They calculate the percentage of frames in which the distance between the tracked location and the ground truth location is less than a fixed thresholde (e.g., 20 pixels). The other quantitative measure is called Pascal score, which can find out the overlapping area between tracked target and its ground truth value as described by Eq. (16):

$$P = \frac{\mathrm{area}(R_t \cap R_g)}{\mathrm{area}(R_t \cup R_g)}, \tag{16}$$

where R_t and R_g are the tracked target region and its ground truth region, respectively. $\cap$ and $\cup$ show the intersection and union symbols respectively. Pascal score may have a value from 0 to 1 in a closed interval. If there is no overlapping region, its value is 0, and it gains value of 1 in case of full overlap. The target is considered to be successfully tracked in a frame, if its Pascal score is greater than 0.5, (i.e., at least fifty percent overlap). In order to have a fair comparison between different tracking algorithms, two things are required. One is test videos with annotations, and the other is implementations of the algorithms. Wu et al. [205] have organized a dataset comprising of 50 videos with their ground truth values and a code library having implementation of 29 tracking algorithms. They provide performance evaluation and comparison of these algorithms over different parameters, e.g., scale changes, illumination changes, occlusion handling, and overall tracking performance. In order to make this survey self-contained, we are providing a list of a few publically available VOT resources as shown in Table 10.



(a) (b) (c) (d) (e) (f) (g) (h) (i) (j) (k) (l)

Fig. 8 A few tracked frames of Liquor video sequence. The yellow rectangle shows the tracked window. The more closer to the target, the better the result. (a) Frame 360; (b) Frame 607; (c) Frame 776; (d) Frame 1115; (e) Frame 1183; (f) Frame 1236; (g) Frame 1319; (h) Frame 1355; (i) Frame 1438; (j) Frame 1462; (k) Frame 1504; (l) Frame 1517

Table 10 List of a few online publically available tracking resources

No.	Name	Dataset	Ground truth	Source Code	URL
1	Fragtrack [71]	√	√	√	cs.technion.ac.il[1)]
2	Incremental visual tracker [206]	√	√	√	cs.utoronto.ca[2)]
3	ℓ_1 tracker [170]	×	×	√	ist.temple.edu[3)]
4	Kernel based tracker [57]	×	×	√	google.com[4)]
5	Boosting tracker	√	×	√	vision.ee.ethz.ch[5)]
6	MIL tracker [131]	√	√	√	vision.ucsd.edu[6)]
7	Visual tracking decomposition [207]	√	√	√	cv.snu.ac.kr[7)]
8	Structural SVM tracker [140]	×	×	√	samhare.net[8)]
9	PROST tracker [134]	√	√	√	icg.tugraz.at[9)]
10	KLT tracker [35]	×	×	√	ces.clemson.edu[10)]
11	Condensation tracker [95]	√	×	√	robots.ox.ac.uk[11)]
12	Caviar sequences	√	√	×	inf.ed.ac.uk[12)]
13	PETS sequences	√	√	×	hitech-projects.com[13)]
14	Compressive tracking [177]	√	√	√	comp.polyu.edu.hk[14)]
15	Structural local sparse tracker [176]	√	√	√	ice.dlut.edu.cn[15)]
16	Sparsity-based collaborative tracker [175]	√	√	√	ice.dlut.edu.cn[16)]

1) http://www.cs.technion.ac.il/~amita/fragtrack/fragtrack.htm 2) http://www.cs.utoronto.ca/~dross/ivt/
3) http://www.ist.temple.edu/~hbling/code_data.htm 4) http://code.google.com/p/detect/
5) http://www.vision.ee.ethz.ch/boostingTrackers/ 6) http://vision.ucsd.edu/~bbabenko/project_miltrack.shtml
7) http://cv.snu.ac.kr/research/~vtd/ 8) http://www.samhare.net/research/struck
9) http://gpu4vision.icg.tugraz.at/index.php?content=subsites/prost/prost.php 10) http://www.ces.clemson.edu/~stb/klt/
11) http://www.robots.ox.ac.uk/~misard/condensation.html 12) http://homepages.inf.ed.ac.uk/rbf/CAVIARDATA1/
13) http://www.hitech-projects.com/euprojects/cantata/datasets_cantata/dataset.html 14) http://www4.comp.polyu.edu.hk/~cslzhang/CT/CT.htm
15) http://ice.dlut.edu.cn/lu/Project/cvpr12_jia_project/cvpr12_jia_project.htm 16) http://ice.dlut.edu.cn/lu/Project/cvpr12_scm/cvpr12_scm.htm

5 Conclusion and future direction

In this survey, we have investigated classical as well as contemporary approaches for object tracking algorithms. The mean shift, Kalman filter, motion detection, and template matching based algorithms are presented as classical approaches for visual tracking, whereas tracking-by-detection, swarm intelligence, sparse representation, integration of context, and tracking in crowded environment are discussed as contemporary tracking algorithms. Issues involved in robust tracking are not prior information about target, online learning, template drift, occlusion, clutter, etc. Recently, Ali et al. [110] propose a novel template updating method to alleviate the drift problem, which updates the template using second order IIR filter. The first parameter of the filter is the rate of appearance of the target, and the second parameter considers the appearance difference from its actual appearance. They propose a heuristic tracking method which combines correlation, KF, and fast mean shift algorithms for long-term robust tracking. The task of human mind to process video data is of great inspiration for the researchers, and these methods are getting their place in target tracking. For example, Wang et al. [208] present memory-based multi-agent co-evolutionary modeling for VOT. Their memory model can remember, forget, and retrieve the target appearance by its own experience, and the co-evolutionary process consists of competition, recombination, and migration behavior. Multiple memory agents are put in the search region which finds the current position of the target by co-evolutionary process. In a recent paper, Wang et al. [209] present memory-based cognitive modeling for robust tracking which consists of long-term, short-term, and ultra-short-term memory. The cognitive model can remember, recall, forget, learn, classify, and recognize visual information. Visual focus of attention (VFOA) helps to track targets in human, therefore, the algorithms are being introduced, e.g., [210], which try to mimic this capability for real time tracking. Context aware tracking approaches [181, 184, 196] have generally shown better results in recent years. Therefore, integration and automatic extraction of contextual objects (supporters or auxiliary objects) is brought into focus by the computer vision community.

Acknowledgements This work was supported by the National Natural Science Foundation of China (Grant Nos. 61170296 and 61190125), and the R&D Program (2013BAH35F01), PIEAS-administered HEC Endowment Fund for Higher education and R&D for IT and Telecom Sector Fund.

References

1. Ta D N, Chen W C, Gelfand N, Pulli K. Surftrac: efficient tracking and continuous object recognition using local feature descriptors. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. 2009, 2937–2944
2. Skrypnyk I, Lowe D G. Scene modelling, recognition and tracking with invariant image features. In: Proceedings of IEEE and ACM International Symposium on Mixed and Augmented Reality. 2004, 110–119
3. Chau D P, Bremond F, Thonnat M. Object tracking in videos: approaches and issues. 2013, arXiv preprint arXiv: 1304.5212
4. Ko T. A survey on behavior analysis in video surveillance for homeland security applications. In: Proceedings of the 37th IEEE Applied Imagery Pattern Recognition Workshop. 2008, 1–8
5. Ess A, Schindler K, Leibe B, Van Gool L. Object detection and tracking for autonomous navigation in dynamic environments. The International Journal of Robotics Research, 2010, 29: 1707–1725
6. Mistry P, Maes P. SixthSense: a wearable gestural interface. In: Proceedings of ACM SIGGRAPH ASIA 2009 Sketches. 2009, 11
7. Bradski G R. Real time face and object tracking as a component of a perceptual user interface. In: Proceedings of the 4th IEEE Workshop on Applications of Computer Vision. 1998, 214–219
8. Zhu Z, Ji Q. Eye gaze tracking under natural head movements. In: Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition. 2005, 918–923
9. Kim I, Choi H S, Yi K M, Choi J Y, Kong S G. Intelligent visual surveillance — a survey. International Journal of Control, Automation and Systems, 2010, 8(5): 926–939
10. Siemens S. Sistore CX EDS-intelligent video detection system. Technical report. 2008
11. Collins R, Lipton A, Kanade T, Tolliver E. A system for video surveillance and monitoring. Technical Report CMU-RI-TR-00-12. 2000
12. Haritaoglu I, Harwood D, Davis L S. W4: real-time surveillance of people and their activities. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2000, 22(8): 809–830
13. Kettnaker V, Zabih R. Bayesian multi-camera surveillance. In: Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition. 1999, 242–259
14. Hu W, Tan T, Wang L, Maybank S. A survey on visual surveillance of object motion and behaviors. IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews, 2004, 34(3): 334–352
15. Collins R T, Lipton A J, Fujiyoshi H, Kanade T. Algorithms for cooperative multisensor surveillance. Proceedings of the IEEE, 2001, 89(10): 1456–1477
16. Greiffenhagen M, Comaniciu D, Niemann H, Ramesh V. Design, analysis, and engineering of video monitoring systems: an approach and a case study. Proceedings of the IEEE, 2001, 89(10): 1498–1517
17. Kumar R, Sawhney H, Samarasekera S, Hsu S, Tao H, Guo Y, Hanna K, Pope A, Wildes R, Hirvonen D, Hansen M, Burt P. Aerial video surveillance and exploitation. Proceedings of the IEEE, 2001, 89(10): 1518–1539
18. Coifman B, Beymer D, McLauchlan P, Malik J. A real-time computer vision system for vehicle tracking and traffic surveillance. Transporta-

tion Research Part C: Emerging Technologies, 1998, 6(4): 271–288

19. Tai J C, Tseng S T, Lin C P, Song K T. Real-time image tracking for automatic traffic monitoring and enforcement applications. Image and Vision Computing, 2004, 22(6): 485–501
20. Masoud O, Papanikolopoulos N P. A novel method for tracking and counting pedestrians in real-time using a single camera. IEEE Transactions on Vehicular Technology, 2001, 50(5): 1267–1278
21. Papanikolopoulos N P, Khosla P K. Adaptive robotic visual tracking: theory and experiments. IEEE Transactions on Automatic Control, 1993, 38(3): 429–445
22. Sakagami Y, Watanabe R, Aoyama C, Matsunaga S, Higaki N, Fujimura K. The intelligent asimo: system overview and integration. In: Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems. 2002, 2478–2483
23. Mondragon I F, Campoy P, Correa J F, Mejias L. Visual model feature tracking for UAV control. In: Proceedings of IEEE International Symposium on Intelligent Signal Processing. 2007, 1–6
24. Lee J, Huang R, Vaughn A, Xiao X, Hedrick J K, Zennaro M, Sengupta R. Strategies of path-planning for a UAV to track a ground vehicle. In: Proceedings of Annual Autonomous Intelligent Networks and Systems Conference. 2003
25. Handmann U, Kalinkea T, Tzomakas C, Werner M, von Seelen W. Computer vision for driver assistance systems. In: Proceedings of Aerospace/Defense Sensing and Controls. 1998, 136–147
26. Avidan S. Support vector tracking. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2004, 26(8): 1064–1072
27. Ahmed J, Shah M, Miller A, Harper D, Jafri M N. A vision-based system for a UGV to handle a road intersection. In: Proceedings of National Conference on Artificial Intelligence. 2007, 1077
28. Rand D, Kizony R, Weiss P T. The Sony playstation II eyetoy: low-cost virtual reality for use in rehabilitation. Journal of Neurologic Physical Therapy, 2008, 32(4): 153–163
29. Wang S, Xiong X, Xu Y, Wang C, Zhang W, Dai X, Zhang D. Face-tracking as an augmented input in video games: enhancing presence, role-playing and control. In: Proceedings of SIGCHI Conference on Human Factors in Computing Systems. 2006, 1097–1106
30. Amini A A, Owen R L, Anandan P, Duncan J. Non-rigid motion models for tracking the left ventricular wall. In: Proceedings of the 12th International Conference on Information Processing in Medical Imaging. 1991, 343–357
31. Vasconcelos M J M, Ventura S M R, Freitas D R S, Tavares J M R S. Using statistical deformable models to reconstruct vocal tract shape from magnetic resonance images. Institution of Mechanical Engineers, Part H: Journal of Engineering in Medicine, 2010, 224(10): 1153–1163
32. Vasconcelos M J M, Ventura S M R, Freitas D R S, Tavares J M R S. Towards the automatic study of the vocal tract from magnetic resonance images. Journal of Voice: Official Journal of the Voice Foundation, 2011, 25: 732–742
33. Stauffer C, Grimson W E L. Learning patterns of activity using real-time tracking. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2000, 22(8): 747–757
34. Bodor R, Jackson M, Papanikolopoulos N. Vision-based human tracking and activity recognition. In: Proceedings of the 11th Mediterranean Conference on Control and Automation. 2003, 18–20
35. Lucas B D, Kanade T. An iterative image registration technique with an application to stereo vision. In: Proceedings of International Joint Conference on Artificial Intelligence. 1981, 674–679
36. Fitts J M. Precision correlation tracking via optimal weighting functions. In: Proceedings of the 18th IEEE Conference on Decision and Control including the Symposium on Adaptive Processes. 1979, 280–283
37. Yilmaz A, Javed O, Shah M. Object tracking: a survey. ACM Computing Surveys, 2006, 38(4): 13
38. Joshi K A, Thakore D G. A survey on moving object detection and tracking in video surveillance system. International Journal of Soft Computing and Engineering, 2012: 2231–2307
39. Yang H, Shao L, Zheng F, Wang L, Song Z. Recent advances and trends in visual tracking: a review. Neurocomputing, 2011, 74(18): 3823–3831
40. Cannons K. A review of visual tracking. Technical Report CSE-2008-07. 2008
41. Geronimo D, Lopez A M, Sappa A D, Graf T. Survey of pedestrian detection for advanced driver assistance systems. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2010, 32(7): 1239–1258
42. Ogale N A. A survey of techniques for human detection. Master's Thesis. University of Maryland, 2006
43. Trucco E, Plakas K. Video tracking: a concise survey. IEEE Journal of Oceanic Engineering, 2006, 31(2): 520–529
44. Moeslund T B, Hilton A, Krüger V. A survey of advances in vision-based human motion capture and analysis. Computer Vision and Image Understanding, 2006, 104(2): 90–126
45. Aggarwal J K, Cai Q. Human motion analysis: a review. In: Proceedings of IEEE Nonrigid and Articulated Motion Workshop. 1997, 90–102
46. Kang W, Deng F. Research on intelligent visual surveillance for public security. In: Proceedings of IEEE/ACIS International Conference on Computer and Information Science. 2007, 824–829
47. Forsyth D A, Arikan O, Ikemoto L. Computational Studies of Human Motion: Tracking and Motion Synthesis. Boston: Now Publishers Inc., 2006
48. Zhan B, Monekosso D N, Remagnino P, Velastin S A, Xu L Q. Crowd analysis: a survey. Machine Vision and Applications, 2008, 19(5-6): 345–357
49. Arulampalam M S, Maskell S, Gordon N, Clapp T. A tutorial on particle filters for online nonlinear/non-gaussian bayesian tracking. IEEE Transactions on Signal Processing, 2002, 50(2): 174–188
50. Jalal A S, Singh J. The state-of-the-art in visual object tracking. Informatica Slovenia, 2012, 36(3): 227–248
51. Li X, Hu W, Shen C, Zhang Z, Dick A, Hengel A V D. A survey of appearance models in visual object tracking. ACM Transactions on Intelligent Systems and Technology, 2013, 4(4): 58
52. Fukunaga K, Hostetler L. The estimation of the gradient of a density function, with applications in pattern recognition. IEEE Transactions on Information Theory, 1975, 21(1): 32–40
53. Cheng Y. Mean shift, mode seeking, and clustering. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1995, 17(8): 790–799
54. Comaniciu D, Meer P. Mean shift: a robust approach toward feature space analysis. IEEE Transactions on Pattern Analysis and Machine

Intelligence, 2002, 24(5): 603–619

55. Comaniciu D, Meer P. Robust analysis of feature spaces: color image segmentation. In: Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition. 1997, 750–755
56. Comaniciu D, Ramesh V, Meer P. Real-time tracking of non-rigid objects using mean shift. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. 2000, 142–149
57. Comaniciu D, Ramesh V, Meer P. Kernel-based object tracking. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2003, 25(5): 564–575
58. Hero A O, Ma B, Michel O J J, Gorman J. Applications of entropic spanning graphs. IEEE Signal Processing Magazine, 2002, 19(5): 85–95
59. Yang C, Duraiswami R, Davis L. Efficient mean-shift tracking via a new similarity measure. In: Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition. 2005, 176–183
60. Beleznai C, Fruhstuck B, Bischof H. Human tracking by fast mean shift mode seeking. Journal of Multimedia, 2006, 1(1): 1–8
61. Beleznai C, Fruhstuck B, Bischof H. Human tracking by mode seeking. In: Proceedings of the 4th International Symposium on Image and Signal Processing and Analysis. 2005, 1–6
62. Beleznai C, Fruhstuck B, Bischof H. Tracking multiple humans by fast mean shift mode seeking. In: Proceedings of IEEE International Workshop on Performance Evaluation of Tracking and Surveillance. 2005, 25–32
63. Beleznai C, Fruhstuck B, Bischof H. Detecting humans in groups using a fast mean shift procedure. In: Proceedings of Workshop of the Austrian Association for Pattern Recogniton. 2004, 71–78
64. Beleznai C, Fruhstuck B, Bischof H. Human detection in groups using a fast mean shift procedure. In: Proceedings of International Conference on Image Processing. 2004, 349–352
65. Zivkovic Z, Krose B. An EM-like algorithm for color-histogram-based object tracking. In: Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition. 2004, 798–803
66. Zhou H, Yuan Y, Zhang Y, Shi C. Non-rigid object tracking in complex scenes. Pattern Recognition Letters, 2009, 30(2): 98–102
67. Ning J, Zhang L, Zhang D, Wu C. Robust object tracking using joint color-texture histogram. International Journal of Pattern Recognition and Artificial Intelligence, 2009, 23: 1245–1263
68. Shan C, Tan T, Wei Y. Real-time hand tracking using a mean shift embedded particle filter. Pattern Recognition, 2007, 40(7): 1958–1970
69. Wang X, Liu L, Tang Z. Infrared human tracking with improved mean shift algorithm based on multicue fusion. Journal of Applied Otics, 2009, 48(21): 4201–4212
70. Shen C, Brooks M J, Van Den Hengel A. Fast global kernel density mode seeking: applications to localization and tracking. IEEE Transactions on Image Processing, 2007, 16(5): 1457–1469
71. Adam A, Rivlin E, Shimshoni I. Robust fragments-based tracking using the integral histogram. In: Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition. 2006, 798–805
72. Jeyakar J, Babu R V, Ramakrishnan K R. Robust object tracking with background-weighted local kernels. Computer Vision and Image Understanding, 2008, 112(3): 296–309
73. Khan M I, Ahmed J, Ali A, Masood A. Robust edge-enhanced fragment based normalized correlation tracking in cluttered and occluded imagery. In: Proceedings of Signal Processing, Image Processing and Pattern Recognition. 2009, 169–176
74. Kalman R E, Bucy R S. New results in linear filtering and prediction theory. Journal of Basic Engineering, 1961, 83: 95–108
75. Brookner E. Tracking and Kalman Filtering Made Easy. New York: Wiley, 1998
bibitemWGrewal2008 Grewal M S, Andrews A P. Kalman filtering: theory and practice using MATLAB. New York, Chichester, Brisbane: John Wiley & Sons, 2008
76. Welch G, Bishop G. An introduction of the kalman filter. Technical Report, 2005
77. Asgarizadeh M, Pourghassem H. A robust object tracking synthetic structure using regional mutual information and edge correlation-based tracking algorithm in aerial surveillance application. Signal, Image and Video Processing, 2015, 9(1): 175–189
78. Comaniciu D, Ramesh V. Mean shift and optimal prediction for efficient object tracking. In: Proceedings of International Conference on Image Processing. 2000, 70–73
79. Li Z, Xu C, Li Y. Robust object tracking using mean shift and fast motion estimation. In: Proceedings of IEEE International Symposium on Intelligent Signal Processing and Communication Systems. 2007, 734–737
80. Li X, Zhang T, Shen X, Sun J. Object tracking using an adaptive kalman filter combined with mean shift. Optical Engineering, 2010, 49(2): 020503
81. Ali A, Mirza S M. Object tracking using correlation, kalman filter and fast means shift algorithms. In: Proceedings of International Conference on Emerging Technologies. 2006, 174–178
82. Ahmed J, Jafri M N, Shah M, Akbar M. Real-time edge-enhanced dynamic correlation and predictive open-loop car-following control for robust tracking. Machine Vision and Applications, 2008, 19(1): 1–25
83. Boykov Y, Huttenlocher D P. Adaptive bayesian recognition in tracking rigid objects. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. 2000, 697–704
84. Beymer D, McLauchlan P, Coifman B, Malik J. A real-time computer vision system for measuring traffic parameters. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. 1997, 495–501
85. Broida T J, Chellappa R. Estimation of object motion parameters from noisy images. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1986, 8(1): 90–99
86. Gennery D B. Visual tracking of known three-dimensional objects. International Journal of Computer Vision, 1992, 7(3): 243–270
87. Terzopoulos D, Szeliski R. Tracking with kalman snakes. In: Active Vision. Cambridge, MA, USA: MIT Press, 1993, 3–20
88. Blake A, Isard M. Active Contours: The Application of Techniques from Graphics, Vision, Control Theory and Statistics to Visual Tracking of Shapes in Motion. 1st ed. Secaucus, NJ, USA: Springer-Verlag New York, Inc., 1998
89. Cuevas E V, Zaldivar D, Rojas R. Kalman filter for vision tracking. Technical Report. 2005
90. Jang D S, Choi H I. Active models for tracking moving objects. Pattern

Recognition, 2000, 33(7): 1135–1146

91. Ridder C, Munkelt O, Kirchner H. Adaptive background estimation and foreground detection using kalman-filtering. In: Proceedings of International Conference on recent Advances in Mechatronics. 1995, 193–199
92. Peterfreund N. Robust tracking of position and velocity with kalman snakes. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1999, 21(6): 564–569
93. Anderson B D O, Moore J B. Optimal Filtering. Mincda: Courier Dover Publications, 2012
94. Doucet A, Godsill S, Andrieu C. On sequential monte carlo sampling methods for bayesian filtering. Statistics and Computing, 2000, 10(3): 197–208
95. Isard M, Blake A. Condensation–conditional density propagation for visual tracking. International Journal of Computer Vision, 1998, 29(1): 5–28
96. Rao G M, Satyanarayana C. Visual object target tracking using particle filter: a survey. International Journal of Image, Graphics and Signal Processing, 2013, 5(6): 57–71
97. Duda R O, Hart P E. Pattern Classification and Scene Analysis. New York: Wiley, 1973
98. Gonzalez R C, Woods R E. Digital Image Processing. Upper Saddle River, N.J.: Pearson/Prentice Hall, 2008
99. Kuglin C D, Hines D C. The phase correlation image alignment method. IEEE Conference on Cybernetics and Society, 1975, 163–165
100. Lewis J P. Fast normalized cross-correlation. Vision Interface, 1995, 10(1): 120–123
101. Chien S I, Sung S H. Adaptive window method with sizing vectors for reliable correlation-based target tracking. Pattern Recognition, 2000, 33(2): 237–249
102. Manduchi R, Mian G A. Accuracy analysis for correlation-based image registration algorithms. In: Proceedings of IEEE International Symposium on Circuits and Systems. 1993, 834–837
103. Stone H S, Tao B, McGuire M. Analysis of image registration noise due to rotationally dependent aliasing. Journal of Visual Communication and Image Representation, 2003, 14(2): 114–135
104. Stone H S. Fourier-based image registration techniques. Technical Report. 2002
105. Foroosh H, Zerubia J B, Berthod M. Extension of phase correlation to subpixel registration. IEEE Transactions on Image Processing, 2002, 11(3): 188–200
106. Keller Y, Averbuch A, Miller O. Robust phase correlation. In: Proceedings of the 17th International Conference on Pattern Recognition. 2004, 740–743
107. Ahmed J, Jafri M N. Improved phase correlation matching. In: Proceedings of International Conference on Image and Signal Processing. 2008, 128–135
108. Blackman S S, Popoli R F. Design and Analysis of Modern Tracking Systems. Boston, M A: Artech House, 1999
109. Nixon M S, Aguado A S. Feature Extraction & Image Processing. London: Academic Press, 2008
110. Ali A, Jalil A, Ahmed J, Iftikhar M A, Hussain M. Correlation, kalman filter and adaptive fast mean shift based heuristic approach for robust visual tracking. Signal, Image and Video Processing, 2014: 1–19
111. Wren C R, Azarbayejani A, Darrell T, Pentland A P. Pfinder: real-time tracking of the human body. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1997, 19(7): 780–785
112. Grimson W E L, Stauffer C, Romano R, Lee L. Using adaptive tracking to classify and monitor activities in a site. In: Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition. 1998, 22–29
113. Stauffer C, Grimson W E L. Adaptive background mixture models for real-time tracking. In: Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition. 1999
114. KaewTraKulPong P, Bowden R. An improved adaptive background mixture model for real-time tracking with shadow detection. Video-Based Surveillance Systems. 2002, 135–144
115. Horprasert T, Harwood D, Davis L S. A robust background subtraction and shadow detection. In: Proceedings of Asian Conference on Computer Vision. 1999, 983–988
116. Horprasert T, Harwood D, Davis L S. A statistical approach for real-time robust background subtraction and shadow detection. In: Proceedings of International Conference on Computer Vision. 1999, 1–19
117. Oliver N M, Rosario B, Pentland A P. A bayesian computer vision system for modeling human interactions. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2000, 22(8): 831–843
118. Lipton A J, Fujiyoshi H, Patil R S. Moving target classification and tracking from real-time video. In: Proceedings of the 4th IEEE Workshop on Applications of Computer Vision. 1998, 8–14
119. Dailey D J, Cathey F W, Pumrin S. An algorithm to estimate mean traffic speed using uncalibrated cameras. IEEE Transactions on Intelligent Transportation Systems, 2000, 1(2): 98–107
120. Dailey D J, Li L. An algorithm to estimate vehicle speed using uncalibrated cameras. In: Proceedings of IEEE/IEEJ/JSAI International Conference on Intelligent Transportation Systems. 1999, 441–446
121. Horn B K P, Schunck B G. Determining optical flow. Technical Report. 1980
122. Black M J, Anandan P. The robust estimation of multiple motions: Parametric and piecewise-smooth flow fields. Computer vision and image understanding, 1996, 63(1): 75–104
123. Szeliski R, Coughlan J. Spline-based image registration. International Journal of Computer Vision, 1997, 22(3): 199–218
124. Shi J, Tomasi C. Good features to track. In: Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition. 1994, 593–600
125. Rangarajan K, Shah M. Establishing motion correspondence. In: Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition. 1991, 103–108
126. Papageorgiou C P, Oren M, Poggio T. A general framework for object detection. In: Proceedings of the 6th IEEE International Conference on Computer Vision. 1998, 555–562
127. Cremers D, Schnorr C. Statistical shape knowledge in variational motion segmentation. Image and Vision Computing, 2003, 21(1): 77–86
128. Li B, Chellappa R, Zheng Q, Der S Z. Model-based temporal object verification using video. IEEE Transactions on Image Processing, 2001, 10(6): 897–908
129. Bertalmio M, Sapiro G, Randall G. Morphing active contours. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2000,

22(7): 733–737

130. Mansouri A R. Region tracking via level set pdes without motion computation. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2002, 24(7): 947–961
131. Babenko B, Yang M H, Belongie S. Robust object tracking with online multiple instance learning. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2011, 33(8): 1619–1632
132. Grabner H, Grabner M, Bischof H. Real-time tracking via on-line boosting. In: Proceedings of British Machine Vision Conference. 2006, 1(5): 6
133. Collins R T, Liu Y, Leordeanu M. Online selection of discriminative tracking features. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2005, 27(10): 1631–1643
134. Santner J, Leistner C, Saffari A, Pock T, Bischof H. Prost: parallel robust online simple tracking. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. 2010, 723–730
135. Liu X, Yu T. Gradient feature selection for online boosting. In: Proceedings of the 11th IEEE International Conference on Computer Vision. 2007, 1–8
136. Avidan S. Ensemble tracking. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2007, 29(2): 261–271
137. Wang J, Chen X, Gao W. Online selecting discriminative tracking features using particle filter. In: Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition. 2005, 1037–1042
138. Kuncheva L I. Combining pattern classifiers: Methods and algorithms. IEEE Transactions on Neural Networks, 2007, 18(3): 964–964
139. Bishop C M. Pattern Recognition and Machine Learning. Springer, 2006
140. Hare S, Saffari A, Torr P H S. Struck: structured output tracking with kernels. In: Proceedings of IEEE International Conference on Computer Vision. Nov 2011, 263–270
141. Stalder S, Grabner H. On-line Boosting Trackers. ETH-Zurich, 2009
142. Grabner H, Leistner C, Bischof H. Semi-supervised on-line boosting for robust tracking. In: Proceedings of European Conference on Computer Vision. 2008, 234–247
143. Zeisl B, Leistner C, Saffari A, Bischof H. On-line semi-supervised multiple-instance boosting. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. 2010, 1879–1879
144. Saffari A, Leistner C, Godec M, Bischof H. Robust multi-view boosting with priors. In: Proceedings of European Conference on Computer Vision, 2010, 776–789
145. Leistner C, Saffari A, Roth P M, Bischof H. On robustness of on-line boosting — a competitive study. In: Proceedings of IEEE International Conference on Computer Vision Workshops. 2009, 1362–1369
146. Masnadi-Shirazi H, Mahadevan V, Vasconcelos N. On the design of robust classifiers for computer vision. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. 2010, 779–786
147. Zhang K, Song H. Real-time visual tracking via online weighted multiple instance learning. Pattern Recognition, 2013, 46(1): 397–411
148. Williams O, Blake A, Cipolla R. A sparse probabilistic learning algorithm for real-time tracking. In: Proceedings of IEEE International Conference on Computer Vision. 2003, 353–360
149. Kennedy J, Eberhart R. Particle swarm optimization. In: Proceedings of IEEE International Conference on Neural Networks. 1995, 1942–1948
150. Eberhart R, Kennedy J. A new optimizer using particle swarm theory. In: Proceedings of the 6th International Symposium on Micro Machine and Human Science. 1995, 39–43
151. Poli R. Analysis of the publications on the applications of particle swarm optimisation. Journal of Artificial Evolution and Applications, 2008, 2008: 3
152. Clerc M, Kennedy J. The particle swarm — explosion, stability, and convergence in a multidimensional complex space. IEEE Transactions on Evolutionary Computation, 2002, 6(1): 58–73
153. Wachowiak M P, Smolikova R, Zheng Y, Zurada J M, Elmaghraby A S. An approach to multimodal biomedical image registration utilizing particle swarm optimization. IEEE Transactions on Evolutionary Computation, 2004, 8(3): 289–301
154. Engelbrecht A P. Computational Intelligence: an Introduction. 2nd ed. New York: Johu Wiley & Sons, 2007
155. Sedighizadeh D, Masehian E. Particle swarm optimization methods, taxonomy and applications. International Journal of Computer Theory and Engineering, 2009, 1(5): 486–502
156. Zhang X, Hu W, Maybank S, Zhu M. Sequential particle swarm optimization for visual tracking. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. 2008, 1–8
157. Zhang X, Hu W, Qu W, Maybank S. Multiple object tracking via species-based particle swarm optimization. IEEE Transactions on Circuits and Systems for Video Technology, 2010, 20(11): 1590–1602
158. Akbari R, Jazi M D, Palhang M. A hybrid method for robust multiple objects tracking in cluttered background. In: Proceedings of the 2nd International Conference on Information amp; amp; Communication Technologies. 2006, 1562–1567
159. Kwolek B. Multi-object tracking using particle swarm optimization on target interactions. In: Proceedings of Advances in Heuristic Signal Processing and Applications. 2013, 63–78
160. Anton-Canalis L, Hernandez-Tejera M, Sanchez-Nielsen E. Particle swarms as video sequence inhabitants for object tracking in computer vision. In: Proceedings of the 6th International Conference on Intelligent Systems Design and Applications. 2006, 604–609
161. Zheng Y, Meng Y. Adaptive object tracking using particle swarm optimization. In: Proceedings of International Symposium on Computational Intelligence in Robotics and Automation. 2007, 43–48
162. Tawab A M A, Abdelhalim M B, Habib S E D. Efficient multi-feature {PSO} for fast gray level object-tracking. Applied Soft Computing, 2014, 14: 317–337
163. Borra S K, Chaparala S K. Tracking of an object in video stream using a hybrid PSO-FCM and pattern matching. International Journal of Engineering Research and Technology, 2013, 2
164. Donoho D L. Compressed sensing. IEEE Transactions on Information Theory, 2006, 52(4): 1289–1306
165. Candes E J, Romberg J K, Tao T. Stable signal recovery from incomplete and inaccurate measurements. Communications on Pure and Applied Mathematics, 2006, 59(8): 1207–1223
166. Wright J, Ma Y, Mairal J, Sapiro G, Huang T S, Yan S. Sparse representation for computer vision and pattern recognition. Proceedings of the IEEE, 2010, 98(6): 1031–1044

167. Sapiro G, Mairal J, Wright J, Ma Y, Huang T, Yan S. Sparse representation for computer vision and pattern recognition. Technical Report. 2009

168. Yang J, Wright J, Huang T S, Ma Y. Image super-resolution via sparse representation. IEEE Transactions on Image Processing, 2010, 19(11): 2861–2873

169. Wright J, Yang A Y, Ganesh A, Sastry S S, Ma Y. Robust face recognition via sparse representation. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2009, 31(2): 210–227

170. Mei X, Ling H. Robust visual tracking using l1 minimization. In: Proceedings of IEEE International Conference on Computer Vision. 2009, 1436–1443

171. Mei X. Visual tracking and illumination recovery via sparse representation. Dissertation for the Doctoral Degree. University of Maryland, 2009

172. Mei X, Ling H. Robust visual tracking and vehicle classification via sparse representation. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2011, 33(11): 2259–2272

173. Liu B, Yang L, Huang J, Meer P, Gong L, Kulikowski C. Robust and fast collaborative tracking with two stage sparse optimization. In: Proceedings of European Conference on Computer Vision. 2010, 624–637

174. Liu J, Huang J, Yang L, Kulikowski C. Robust tracking using local sparse appearance model and k-selection. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. 2011, 1313–1320

175. Zhong W, Lu H, Yang H M. Robust object tracking via sparsity-based collaborative model. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. 2012, 1838–1845

176. Jia X, Lu X, Yang M H. Visual tracking via adaptive structural local sparse appearance model. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. 2012, 1822–1829

177. Zhang K, Zhang L, Yang M H. Real-time compressive tracking. In: Proceedings of European Conference on Computer Vision. 2012, 864–877

178. Zhang S, Yao H, Sun X, Lu X. Sparse coding based visual tracking: Review and experimental comparison. Pattern Recognition, 2013, 46(7): 1772–1788

179. Oliva A, Torralba A. The role of context in object recognition. Trends in Cognitive Sciences, 2007, 11(12): 520–527

180. Divvala S K, Hoiem D, Hays J H, Efros A A, Hebert M. An empirical study of context in object detection. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. 2009, 1271–1278

181. Yang M, Wu Y, Hua G. Context-aware visual tracking. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2009, 31(7): 1195–1209

182. Li Y, Nevatia R. Key object driven multi-category object recognition, localization and tracking using spatio-temporal context. In: Proceedings of Europian Conference on Computer Vision. 2008, 409–422

183. Nguyen H T, Ji Q, Smeulders A W M. Spatio-temporal context for robust multitarget tracking. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2007, 29(1): 52–64

184. Wen L, Cai Z, Lei Z, Yi D, Li S. Robust online learned spatio-temporal context model for visual tracking. IEEE Transactions on Image Processing, 2014, 23(2): 785–796

185. Grabner H, Matas J, Van Gool L, Cattin P. Tracking the invisible: Learning where the object might be. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. 2010, 1285–1292

186. Wu Z, Hristov N I, Hedrick T L, Kunz T H, Betke M. Tracking a large number of objects from multiple views. In: Proceedings of IEEE International Conference on Computer Vision. 2009, 1546–1553

187. Sugimura D, Kitani K M, Okabe T, Sato Y, Sugimoto A. Using individuality to track individuals: clustering individual trajectories in crowds using local appearance and frequency trait. In: Proceedings of IEEE International Conference on Computer Vision. 2009, 1467–1474

188. Ali S, Shah M. Floor fields for tracking in high density crowd scenes. Lecture Notes in Computer Science. 2008, 5303: 1–14

189. Zhao T, Nevatia R. Tracking multiple humans in crowded environment. In: Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition. 2004, II–406

190. Betke M, Hirsh D E, Bagchi A, Hristov N I, Makris N C, Kunz T H. Tracking large variable numbers of objects in clutter. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. 2007, 1–8

191. Li Y, Huang C, Nevatia R. Learning to associate: Hybridboosted multi-target tracker for crowded scene. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. 2009, 2953–2960

192. Wu B, Nevatia R. Tracking of multiple, partially occluded humans based on static body part detection. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. 2006, 951–958

193. Brostow G J, Cipolla R. Unsupervised bayesian detection of independent motion in crowds. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. 2006, 594–601

194. Pellegrini S, Ess A, Schindler K, Van Gool L. You'll never walk alone: Modeling social behavior for multi-target tracking. In: Proceedings of IEEE International Conference on Computer Vision. 2009, 261–268

195. Rodriguez M, Ali S, Kanade T. Tracking in unstructured crowded scenes. In: Proceedings of IEEE International Conference on Computer Vision. 2009, 1389–1396

196. Kratz L, Nishino K. Tracking with local spatio-temporal motion patterns in extremely crowded scenes. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. 2010, 693–700

197. Rodriguez M, Sivic J, Laptev I, Audibert J Y. Data-driven crowd analysis in videos. In: Proceedings of IEEE International Conference on Computer Vision. 2011, 1235–1242

198. Idrees H, Warner N, Shah M. Tracking in dense crowds using prominence and neighborhood motion concurrence. Image and Vision Computing, 2014, 32(1): 14–26

199. Zhang L, Maaten v. d L. Structure preserving object tracking. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. 2013, 1838–1845

200. Zhu F, Wang X, Yu N. Crowd tracking with dynamic evolution of group structures. In: Proceedings of the 13th European Conference on Computer Vision–ECCV 2014, 139–154

201. Gao Y, Ji R, Zhang L, Hauptmann A. Symbiotic tracker ensemble towards a unified tracking framework. IEEE Transactions on Circuits and Systems for Video Technology, 2014, 24(7): 1122–1131

202. Zhong B, Yao H, Chen S, Ji R, Chin T J, Wang H. Visual tracking via weakly supervised learning from multiple imperfect oracles. Pattern Recognition, 2014, 47(3): 1395–1410

203. Yao A, Lin X, Wang G, Yu S. A compact association of particle filtering and kernel based object tracking. Pattern Recognition, 2012, 45(7): 2584–2597

204. Henriques J F, Caseiro R, Martins P, Batista J. Exploiting the circulant structure of tracking-by-detection with kernels. In: Proceedings of the 12th European Conference on Computer Vision — ECCV 2012. 2012, 702–715

205. Wu Y, Lim J, Yang M H. Online object tracking: a benchmark. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. 2013, 2411–2418

206. Ross D A, Lim J, Lin R S, Yang M H. Incremental learning for robust visual tracking. International Journal of Computer Vision, 2008, 77(1-3): 125–141

207. Kwon J, Lee K M. Visual tracking decomposition. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. 2010, 1269–1276

208. Wang Y, Qi Y, Li Y. Memory-based multiagent coevolution modeling for robust moving object tracking. The Scientific World Journal, 2013, 2013

209. Wang Y, Qi Y. Memory-based cognitive modeling for robust object extraction and tracking. Applied Intelligence, 2013, 39(3): 614–629

210. Smith K, Ba S O, Odobez J M, Gatica-Perez D. Tracking the visual focus of attention for a varying number of wandering people. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2008, 30(7): 1212–1229

Ahmad Ali received his BS in computer sciences (Hons.) from University of Engineering & Technology (UET), Lahore, Pakistan. He completed his MS in system engineering from Pakistan Institute of Engineering and Applied Sciences (PIEAS), Islamabad, Pakistan. Currently, he is pursuing his PhD from PIEAS. His areas of interest are image processing, computer vision, object tracking, artificial intelligence, and speech processing.

Abdul Jalil received his MS and M.Phil in electronics from Qauid-i-Azam University, Islamabad, Pakistan in 1986 and 2000, respectively. He completed his PhD in image and signal processing from Mohammad Ali Jinnah University, Pakistan in 2006. Then, he remained affiliated with University of Sussex for Post Doctorate of 9 months during 2008–09 period. His research interests are related to machine vision and signal, and image processing.

Jianwei Niu received his PhD in Computer Science from Beihang University (BUAA), China in 2002. He is a professor in the School of Computer Science and Engineering, BUAA, and an IEEE senior member. He has published more than 100 referred papers, and filed more than 30 patents in mobile and pervasive computing. His current research interests include mobile and pervasive computing, mobile video analysis.

Xiaoke Zhao received his MS in computer science in 2014 from Beihang University, China. He received his BS in 2011 from Guiyang University. He has worked as a research intern at the Nokia Research Center in Beijing. His current research interests include computer vision and pattern recognition.

Saima Rathore received the BS degree in software engineering from Fatima Jinnah Women University, Rawalpindi, Pakistan, in 2006 and the MS degree in computer engineering from the University of Engineering and Technology, Taxila, Pakistan in 2008. She completed her PhD in computer science from Pakistan Institute of Engineering and Applied Sciences, Islamabad, Pakistan in 2015. Currently, she is working as a post-doctoral researcher in Perelman School of Medicine, University of Pennsylvania, USA. Her research interests include medical image analysis, segmentation, classification and evolutionary algorithms.

Javed Ahmed received his BE in electronics engineering from NED University of Engineering & Technology, Karachi, Pakistan in 1994. Then, he received his MS in systems engineering with 2nd position under the fellowship program from Pakistan Institute of Engineering & Applied Sciences (PIEAS), Islamabad, Pakistan in

1997. He joined National Engineering and Scientific Commission in 1997 and worked in the areas of electronics and signal processing. He received his PhD in electrical (telecom) engineering from National University of Sciences & Technology (NUST), Islamabad, Pakistan in 2008. He obtained the Certificate of Achievement from Computer Vision Lab at University of Central Florida, USA in February 2007 for conducting 8-month outstanding joint research with them. His current research areas are image processing, computer vision, signal processing, and soft computing.

Muhammad Aksam Iftikhar completed his BS in computer engineering from University of Engineering and Technology, Lahore, Pakistan in 2007. He received MS in computer science from the same university in 2010. He completed his PhD from Pakistan Institute of Engineering and Applied Sciences, Islamabad, Pakistan. His primary research areas include image processing (especially medical image processing and analysis), computer vision, machine learning and pattern recognition.