CrossMark

Dynamic and robust method for detection and locating vehicles in the video images sequences with use of image processing algorithm

Gholamreza Farahani

Abstract

There are various methods in the field of moving-object tracking in the video images that each of them implies on the specific features of object. Among tracking methods based on features, algorithms based on color are able to provide a precise description of the object and track the object with high speed. One of the efficient methods in the field of object tracking based on color information is mean-shift algorithm. If the color of moving object approaches toward a background model or image background has a low contrast and brightness, then the color information is not enough for target tracking. In this paper, the new tracking method is proposed which with combination of moved object information with color information, the new proposed method will be capable to track object under condition that color information is not enough for tracking. With use of background subtraction method based on Gaussian combination, the binary image which includes moving information will use in the mean-shift algorithm. Usage of object movement information will compensate the lack of spatial information and will increase robustness of algorithm especially in the complicated conditions. Also in order to achieve the robust algorithm against changes in shapes, size, and rotation of object, extended mean-shift algorithm is used. Results show the robustness of proposed algorithm in object tracking especially under conditions which object color is same as background color and have better results in the low contrast condition in comparison to mean-shift and extended mean-shift algorithms.

Keywords: Machine vision, Image clustering, Vehicle navigation, Environment modeling, Traffic control, Mean-shift algorithm

1 Introduction

Object tracking in the video images is one of the fields of applied science which for its difficulties has preoccupied many researches to itself. Although history of tracking systems comes back to design of radar system and GPS location, tracking systems based on two-dimensional signal processing do not have a long history. Human perennial favorite to robots for the industrial or home works helps to fast growth of digital image-tracking systems. Many studies are carried out in the detection and tracking field and different methods is proposed for them, these studies still ongoing, because so far, a perfect method that works well everywhere and be fast and robust has not been achieved. The reasons which make the detection and tracking so hard are changes in the brightness of image, shape changes of desired object, changes of target number, non-Gaussian noises, and occlusion problem.

These algorithms according to their weakness and strengths are used in the different applications such as medical, military and industrial. The most important factors in the procedure toward processing algorithms and machine vision techniques is insertion of intelligent systems in everyday life, so that to this day, robots and other intelligent machines are inserted into the human life.

Correspondence: farahani.gh@irost.org
Electrical and Information Technology Institute, Iranian Research Organization for Science and Technology (IROST), Sh. Ehsani Rad St., Enqelab St., Parsa Sq., Ahmadabad Mostoufi Rd., Azadegan Highway, Tehran 3353136846, Iran

Object tracking based on image has the difficulties and complexities. Movement of object from camera makes the image size of object larger, smaller or rotation in the object image. Environment light variations such as weather clouding or shadow creation during the day, causes changes in the object image. Despite many problems in object tracking, it has many applications.

Nowadays, control of traffic is one the most important and vital issues which it requires employing skilled and experienced people. Traffic engineering is based on control and management of traffic information collecting and analysis such as number of vehicles, speed and traffic flow. In this regard, recently, different strategies are proposed to control traffic mechanism automatically [1–3].

Different sensors such as microwave vehicle detector, inductive loop detector (ILD), video camera, and optical and electronic sensors are created to use for quantitative and qualitative analysis of traffic images [4].

But despite the application diversity, these sensors have some disadvantages. As an example, installation and maintenance of ILDs is expensive. Also ILDs cannot identify vehicles that are stopped or slow moving and only count the vehicles in the one point. Furthermore for measuring traffic flow or vehicles speed, several ILD sensors are needed [5].

Video cameras with covering a wide space of traffic scenes, will receive most information of these scenes compared to other sensors. Also, video camera installation has a lower cost in comparison with other sensors and services and maintenance of them is not requiring an interrupt to the passages traffic flow. Furthermore, specialist with use of analytical and processing methods that are applied to the received images from cameras can manage and control the traffic with lower cost and more efficient.

Limitations that will decrease the precision of vehicle tacking algorithms and recent researches are working on them are:

- Mobile or fixed camera
- Dynamic calibration of camera
- Image quality such as noise and video bit rate
- Night and day limitation
- Weather limitation
- Light reflection
- Vehicle speed
- Distance from camera
- Busy route and targets occlusion
- Processor limitations for complex algorithms

Sochor [6] uses a background subtraction method to detect vehicle and uses Kalman filter for tracking, then calculate movement direction and vehicle speed. The precision of this method will decease during vehicle occlusion and target detection in the night and rainy weather. Yang et al. [7] use background dynamic update to detect vehicle and spatial-temporal profile to calculate amount of traffic and type of vehicle. This method has a precision drop in the vehicle occlusion and dark shadows of vehicles but it is proper for mobile camera that background changes very fast. Jazayeri et al. [8] use Hidden Markov Model (HMM) to separate background from vehicle and vehicle tracking in the day light and night dark. The images used by Jazayeri et al. [8] are in-car moving type. This method has a precision drop in the very fast or slow speed of vehicle and far vehicles in the images and it is suitable for images of mobile camera that background changes rapidly.

Chiu et al. [9] utilize background statistical calculations to detect vehicle and uses visual features to track the vehicle. The precision of this method drops in the occlusion vehicles but for light variation conditions and different weather has an acceptable performance.

Zhang et al. [10] make use of the method to detect the vehicle headlights, and with pairing them, tracks the vehicles in the night. This method for different speed of vehicle conditions and rainy weather and high traffic has an acceptable performance. O'Malley et al. [11] used a HSV (hue, saturation, value) space model for vehicle tracking in the night to find vehicle rear-lamps and paired them. In this method to improve the performance, Kalman filter is used. Salvi et al. [12] have used the adaptive threshold method to find vehicle headlights and pairing them. Also for vehicle tracking and separation car and motorcycle will check the spatial and temporal analysis of emitted light pattern. This method has an error when the vehicle headlight is broken.

Other subject in the motion estimation is speed of estimation. Yan et al. [13] proposed a parallel framework for high-efficiency video coding (HEVC) motion estimation on a 64-core system, which compared with serial execution, achieves more than 30 and 40 times speedup for 1920×1080 and 2560×1600 video sequences, respectively. Also, Yan et al., [14] suggested a highly parallel framework for HEVC coding unit partitioning tree decision on the Tile64 platform that achieves averagely more than 11 and 16 times speedup for 1920×1080 and 2560×1600 video sequences, respectively, without any coding efficiency degradation.

2 Object detection

Object detection in different environment is one of the important subjects and challenges in the machine vision field. Many parameters such as light condition, size, and situation (speed and acceleration) are effective on the detection results. First step in the automatic machine vision systems is detection. Created problems in this step directly affects on the later steps.

Usually detection systems include three main parts: object searching and its motion detection, feature extraction, and classification. In the detection part, modeling algorithms and background elimination will use. After obtaining the moving object, for confirmation that it is the same desired object, its features should be extracted and compared with reference object features.

Algorithms that are used for feature extraction have high variability and each of them has different abilities. One of the existing methods in the image processing and machine vision is the detection based on color histogram. This method has a high speed but in the conditions such as weak contrast, homochromatic background and occlusion is not stable. Color histograms in addition to low cost computing and robustness against size variations and rotation, because in the HSV, space is calculated, therefore they are robust to the light variations and shadow too [15]. In this section, some detection methods that include parts such as background modeling and motion detection will be reviewed.

2.1 Scene analysis system

In the scene analysis and vision care systems, camera and other sensors will be used to receive information, monitore and analysis them and automatic understanding of the events. To analyze events from images, scene component should be recognized. Figure 1 shows a simple structure of scene analysis system.

In scene segmentation block, scene pixels will classify into the background and objects on the scene, then necessary features for processing will extract from objects. Segmentation can be based on spatial, temporal, or spatial-temporal methods. The spatial segmentation methods use the edges, region, texture, color, corners, and contours.

Temporal segmentation methods that used in video use subtraction of two or multiple frames or subtraction from background. Spatial-temporal segmentation methods like as optical flow [16] and subtraction between frames, use spatial, and temporal information simultaneously.

Figure 2 shows block diagram of spatial-temporal segmentation method with more details.

As shown in Fig. 2, spatial-temporal segmentation method includes image acquisition system, background updating, change detection, object positioning, and object tracking. In some methods, background is not used, but in most common methods, static, or dynamic background will use [17].
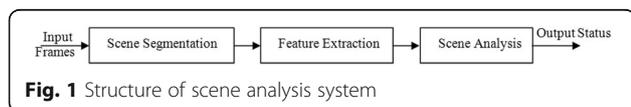
One of the simplest methods of object detection is usage of distance between pixels to the reference image pixels [18]. In this method, the slightest change in appearance such as size and speed has adversely affected on the detection results. Therefore, some geometric features are added such as edges and corners which need a large amount of calculation especially in the complicated scenes.

Other simple method is thresholding the gray images. In this method, thresholding is carried out on the different levels, then according to the changes in the image, object will identify. Carrillo and his colleagues proposed the sperms morphology identifier based on threshold, and used it separate the sperms [19].

Change detection will be used for object finding. To find a position of an object in the image, first, the position of the object in the image should be determined. In these applications, for object detection with use of some features of object, object should be searched in the whole image. Different algorithms are proposed in this field that block matching is one of them. In these algorithms, searching space can be reduced with use of wavelet subspaces or background modeling.
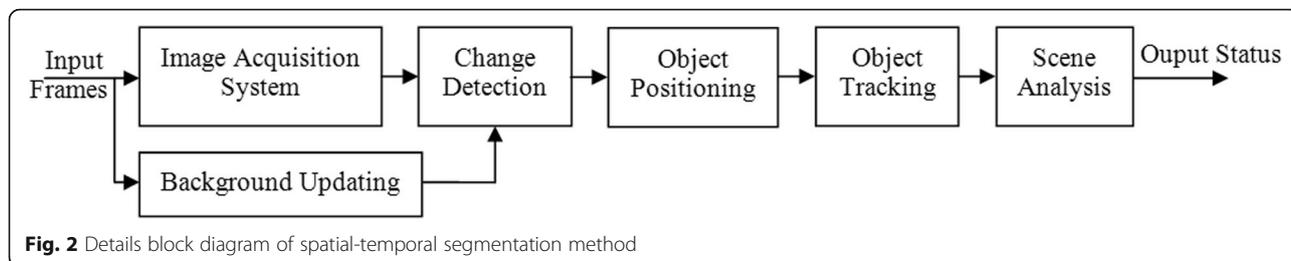
2.2 Background modeling

A group of researches use a frames subtraction as a subtraction between input image and reference image. In this situation, reference image will change according to the variations of input images that named it background learning [20]. One of the problems in background updating is the objects that remain unchanged in the scene for a long time. One solution for this problem is that for the points which recognize as a background model pixels, updating carry out with low forgetting coefficient and in the point which moving object recognize, updating carry out with high forgetting coefficient [21]. Other problems are lighting conditions, scene lighting variation with time, and object shadow.

There are different algorithms in the background modeling, such as adaptive filtering, neural networks, and Gaussian models [22]. In some papers, background model is obtained with averaging several consecutive frames and Gaussian distribution for background points is considered. Therefore, if point stay on the distribution, it will consider as an object inside of scene otherwise it will consider as a background. For calculation of Gaussian model parameters, pixel background in the previous frame is required. Also, multi-Gaussian models are used for background modeling [23].

3 Tracking algorithms

One of the developing problems in the image processing and machine vision is object tracking problem that its

Fig. 1 Structure of scene analysis system

Fig. 2 Details block diagram of spatial-temporal segmentation method

aim is presentation of position changes of an object in the video image sequences. In many object tracking applications, it is most important that tracking methods to be a real time. Therefore, methods should use the algorithms with low time cost.

These methods could be classified into the three clusters. Tracking based on region, tracking based on active contour and tracking based on feature and combined methods. At continuation of this section, each of these methods will explain shortly.

3.1 Tracking based on region
In this model, video images processing system, will find a connected region in the image which is named bubble. For example to each vehicle a bubble is assigned and then this bubble will track with use of cross correlation measurement over time. In this algorithm, the number of bubbles increases with increasing the diversity of object color. For each of the bubbles, a Gaussian function is considered and during searching with use of Gaussian functions and distances between a bubble with other bubbles, recognition and identification of the vehicles will carry out. This algorithm mostly use for detection of object with large size [24].

Although this algorithm works well on the highway which has a few vehicles but it cannot manage the occlusion of vehicles reliably. Also it cannot obtain the 3D of vehicles therefore this algorithm cannot be used to acquire supervision requirements in the busy backgrounds with multiple moving objects.

3.2 Tracking based on active contour
A Duggan for tracking based on region is tracking based on active contour or snakes. The main idea is a presentation of environmental contour of object and dynamic updating of it. Usage of presentation based on contour instead of presentation based on region will reduce the computational complexity.

Tracking algorithms based on active contour will display the scheme of moving objects as a contour that will update in the consecutive frames dynamically. These algorithms present effective descriptions of the objects

than algorithms based on region and have been applied for tracking more successfully.

The problem of these algorithms is that management of occlusion is hard because tracking precision due to precision weak at contour position is limited. Methods based on active contour according to the algorithm which is used to find object border will classify into the snakes and geodesic [25].

3.2.1 Snakes
Snakes are the models based on edge which are introduced for the first time by Kass et al. [26] and then it found many applications including object tracking. After that time, many researches present different versions of it to use in applications such as edge detection, shape modeling, and object tracking under various restrictions. However, due to the noise especially in the busy natural scenery, there are many problems for these models, therefore proper initial value for object and parameter setting is necessary. Apart from problems that arise due to the definition of this snake models, there are other items which are not manages with snakes. These items are:

- Efficiency of snakes at sequences with complicated background is low, for example, backgrounds containing textures that have strong boundaries close to the border of moving object.
- When the intended moving objects are covered with fix or mobile obstacles, these methods will have a problem.

These are main reasons that researchers present the other models to use active contours with use of information based on movement, color, texture, randomization procedures and more appropriate restrictions [27].

3.2.2 Geodesic
This method is a geometric method, which is as an alternative to snake method. This method is based on minimizing energy. This method has been used by Caselles et al. [28]. This method has several advantages than

snake method. For example, this method is not parametric but the main disadvantage of the method is nonlinearity. In this method, like as snakes, there is closed curve which should be within the bounds of the target object. This curve will show as a $C(p)$ that $0 \leq p \leq 1$.

The following function should be minimized to stay this curve on the boundaries of target object.

$$E[C(p)] = \int_0^1 \underbrace{g(|\nabla I(C(p))|)}_{Boundary\ attraction} \underbrace{|\dot{C}(p)|}_{Regularity} \quad (1)$$

where $\dot{C}(p)$ is a partial derivative of curve with respect to p parameter. This component will smooth the curve; g is a function that decreases monotonically [28]. This component drags the curve into the borders of a target object. Geodesic methods do not depend on the type of movement and shape of the object, therefore they are used frequently for object tracking. The main problem of geodesic methods is that they are not sustainable during the occlusion of objects.

3.3 Tracking based on feature
In these methods, identification and object tracking based on features of intended object for tracking will be carried out. In each frame of video images sequence, these features will extract and with feature matching in consecutive frames, tracking process will carry out. This method is used in different systems [29]. These algorithms can be adapted rapidly and used in real-time processing and multi-object tracking successfully. Feature-based tracking algorithms could be break down into the three categories: algorithms based on global features, algorithms based on local features, and algorithms based on dependency graph.

4 Tracking methods based on color feature
Tracking methods based on color feature are robust against camera view angle changes, object size changes, rotation, and partial occlusion. Firstly, in this section, usage of color information in object description will introduce and target localization based on color feature will describe. Then, primary mean-shift algorithm and extended mean-shift algorithm as a robust method against shape and size changes will present and finally deficiencies of mean-shift algorithm will explain.

4.1 Usage of color information in object description
One of the object tracking methods is tracking based on color feature. In the image processing and machine vision, color is as an important feature in the object description, because from one side usage of it does not have complexities of other methods and from other side it has a good robustness. Therefore, according to this method, while maintaining the robustness; amount of processing data can be reduced [30].

One of the criteria that show the object color information is color histogram. This information could be used for present or absence of object in the image. In fact, everywhere in the image that has a histogram like as pervious frame; it can be considered as a new position of the object [31].

The base of tracking in this section is according to color histogram. Histogram describes the share of different colors in the whole image. To obtain share of different colors, after transformation image color space to discrete space, repetition number of special color in image will count and this numeric value will show a histogram [32].

Description of an object based on color histogram is robust to size, angle, and rotation variations and under these conditions will change gradually [33].

Figure 3 shows match value variation of object related to angle variation and distance of camera based on histogram. As it is clear from this figure, variation speed is very slow.

In object description with color, it could be possible to use different color space such as RGB (Red, Green, Blue) or HIS (Hue, Intensity, Saturation). Each of these color space has a three components. In the color space RGB, each image pixel will present with combination of three
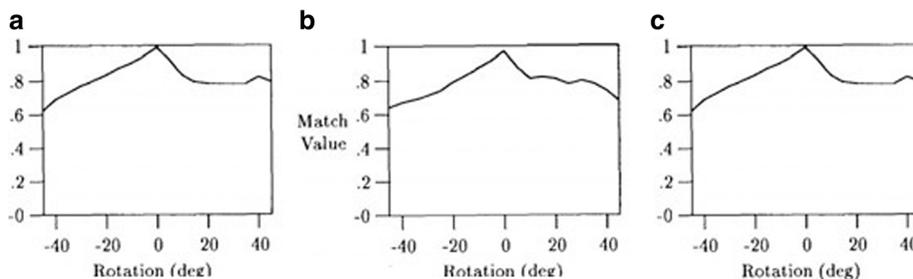

Fig. 3 Histogram match value variation graph related to camera variations, (**a**) distance variation effect, (**b**) rotation to left and right effect, (**c**) rotation to up and down effect [33]

main color components Red (R), Green (G), and Blue (B). In HIS color space, each pixel will split into Hue (H), Intensity (I), and Saturation (S) components. Description of object with color histogram clearly specifies the location of object in each frame of image sequences.

Figure 4 shows a histogram of an area of the object in two frames of image.

As shown in Fig. 4, histogram of the image areas in different location of object has no significant difference. This histogram similarity can be a worthy criterion to find object in consecutive frames of the image sequence.

4.2 Target localization based on color feature

Due to not change the color information during consecutive frames of a video sequence, color feature is a good criterion for target object localization. In next subsection object localization based on color histogram will describe.

4.2.1 Target reference model

With normalization of histogram, it can be considered as a probability density function (PDF) which this PDF is a good description of object in the feature space. With consideration $\overrightarrow{q}$ as a normalized histogram vector in a space with m color component, then

$$\overrightarrow{q} = \{q_u\}_{u=1\ldots m} \tag{2}$$

where.

$$\sum_{u=1}^{m} q_u = 1 \text{ and } q_{u\geq 0} \tag{3}$$

In eq. (2), $\overrightarrow{q}$ is a target reference model and q_u is u-th component of vector $\overrightarrow{q}$.

4.2.2 Target candidate model

Object model that searches in the consecutive frames will introduce as a target candidate which PDF presents with $\overrightarrow{p}(\overrightarrow{y})$ and $\overrightarrow{y}$ indicates the position vector of window center of target candidate. In order to reduce the amount of calculations, it is possible to reduce each component color from 255 to m. Therefore,

$$\overrightarrow{p}(\overrightarrow{y}) = \{p_u(\overrightarrow{y})\}_{u=1\ldots m} \tag{4}$$

where

$$\sum p_u(\overrightarrow{y}) = 1 \tag{5}$$

4.2.3 Similarity function

Target locating with selection of a similarity criterion for target candidate and reference model will fulfill. Equation (6) defines the similarity function between candidate and target reference model.

$$\rho\left(\overrightarrow{p}(\overrightarrow{y}), \overrightarrow{q}\right) = \sum_{u=1}^{m} \sqrt{p_u(\overrightarrow{y})} \cdot \sqrt{q_u} \tag{6}$$

where $\rho\left(\overrightarrow{p}(\overrightarrow{y}), \overrightarrow{q}\right)$ is similarity function, p_u is u-th component of vector $\overrightarrow{p}$, q_u is u-th component of vector $\overrightarrow{q}$ and $\overrightarrow{y}$ indicates the position vector of window center
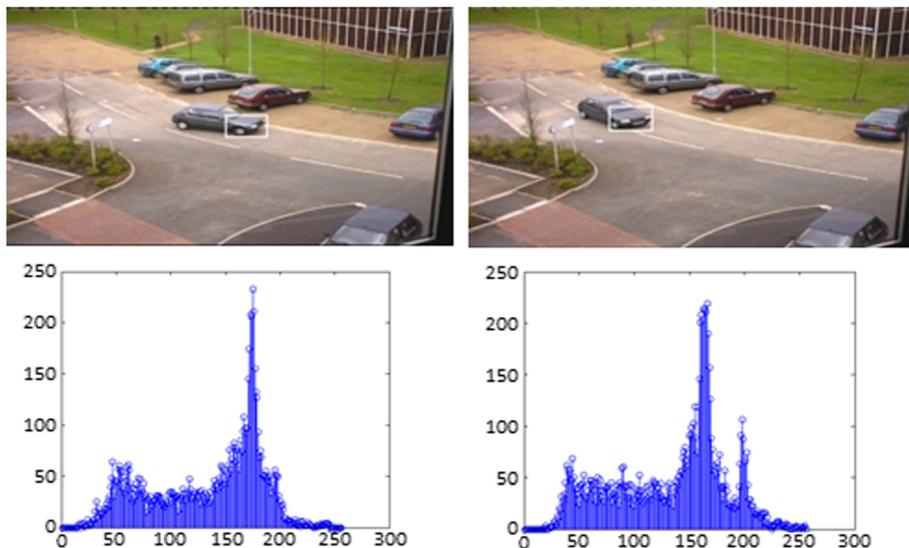

Fig. 4 Histogram of selected area of an object in two different frames

of target candidate. Maximum value of this function is description of object presence with maximum similarity to reference model in the first frame.

If only spectrum information of an object for finding target location in subsequent frames is considered, similarity function will include vast variations and will loss the spatial information of the target. To find maximums of similarity function, there are different ideas. Generally, gradient-based optimization methods have more complexity in implementation. Another general method is a block searching over the whole frame. In this method a searching window will slide over the whole frame and in each step, similarity function will calculate then $\overrightarrow{y}$ which has a most similarity function will choose as a center of target window. Because each pixel in the window does not have an equal importance, therefor in calculation of histogram, weighted histogram will be used [34].

4.2.4 Bhattacharyya distance measure

Similarity function is a distance between target candidate and target reference model. Whatever target candidate is closer to the target model, similarity function will close to the one. For the condition which target candidate exactly meets the target, similarity function has its maximum value one.

In the statistics, Bhattacharyya distance will measure the similarity between two probability distributions. Bhattacharyya coefficients will determine the amount of overlap between the two statistics samples. These coefficients are a_i and b_i which are raised for the first time in 1930 in Indian statistical institute [33] as eq. (7).

$$\text{Bhattacharyya distance} = \sum_{i=1}^{n} \sqrt{\sum a_i . \sum b_i} \quad (7).$$

With consideration of eq. (6), similarity function can be determined with scalar product of two vectors $\left[\sqrt{p_0(\overrightarrow{y})} \ \sqrt{p_1(\overrightarrow{y})} \cdots \sqrt{p_m(\overrightarrow{y})} \right]$ and $\left[\sqrt{q_0} \ \sqrt{q_1} \cdots \sqrt{q_m} \right]$. where $p_0(\overrightarrow{y})$, $p_1(\overrightarrow{y})$ ... $p_m(\overrightarrow{y})$ are 0 to m components of vector $\overrightarrow{p}$ and $q_0, q_1...q_m$ are 0 to m components of vector $\overrightarrow{q}$. It is clear that similarity function is equal with cosine of angle between two vectors in m-dimension space. For two considered distribution, according to eq. (6), distance between target model and target candidate will be as:

$$d(\overrightarrow{y}) = \sqrt{1 - \rho\left(\overrightarrow{p}(\overrightarrow{y}), \overrightarrow{q}\right)} \quad (8)$$

Approaching the target candidate to reference model will increase the similarity criterion and will decrease defined distance [35].

4.2.5 Target localization

To find the object location in the current frame, distance criterion between $\overrightarrow{y}$ and target model should be minimized. Localization process begins from initial $\overrightarrow{y}$ in the previous frame at a neighborhood of the current window and in each step, distance criterion will calculate. Between all of the obtained $\overrightarrow{y}$, for it that distance is minimized, it will choose as a current target location. It is clear that if the object movement is very fast so that object stays out of window, it is not possible to find target location. Also implementation of this method will apply high computational cost.

4.3 Mean-shift algorithm

This algorithm is based on gradient and with use of histogram tries to find the area of image which has been moved in comparison to the previous frame. Because this method will track the target according to the rate of increasing gradient; therefore in comparison to the template matching method, this method is much faster and requires less computational amount, because ignores blindly searching [36]. Steps of mean-shift algorithm are as below.

1- Location and size of window will determine manually by operator.
2- Defines mean-shift vector (eq. (19))
3- Searching window will move according to the mean-shift vector and center of window will place at the end of vector.
4- In new window location, new mean-shift vector recalculate
5- Repeat step 3 until convergence is reached or the stop condition is met.

Mean-shift algorithm is a non-parametric recursive algorithm to find the local modes of PDF. This method uses Kernel Density Estimator (KDE) as an estimator of PDF.

Generally estimation of PDF carries out with parametric or non-parametric methods [37, 38]. One of the non-parametric methods in estimation of PDF is kernel method. The idea of kernel estimator was raised in 1956 by Rosenblatt [39] and in 1962 by Parzen [40]. Till now, study on the extension of kernel estimator is continuing.

If assume X_1, X_2, ..., X_n are n random sample of a distribution with PDF of f, therefore kernel estimator of $f(x)$ is $\hat{f}_h(x)$ as eq. (9) [41].

$$\hat{f}_h(x) = \hat{f}_h(X_1, X_2..., X_n x) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right) \quad (9)$$

where n is a number of distribution and real positive value of h controls the smoothing of kernel estimator and function $K(.)$ is necessary to satisfy the following conditions:

$$\int_{-\infty}^{+\infty} K(x)dx = 1 \tag{10}$$

$$\int_{-\infty}^{+\infty} xK(x)dx = 0 \tag{11}$$

$$\lim_{x\to\infty} |xK(x)| = 0 \tag{12}$$

Parzen showed that if $h \to 0$ then $nh \to \infty$. Therefore, estimator $\hat{f}_h$ is compatible at average [40]. Also $\hat{f}_h$ is a PDF that means

$$\int_{-\infty}^{+\infty} \hat{f}_h(x)dx = 1 \tag{13}$$

that all properties of continuity and derivative inherited from K.

Kernel estimator (eq. (9)) is related to smoothing parameter h. If h value chooses too small, then estimator $\hat{f}_h$ will be very wavy. While for large value of h, this estimator will be too smooth and will not show PDF behavior. Figure 5 shows curve of estimator function based on different values of h [41].

Mean-shift algorithm with finding maximum of similarity function will determine the target location. In order to optimize gradient method, in this algorithm samples of target histogram will convolve with the mask which generally is an isotopic kernel.

Therefore, target locating will find the local maximum of similarity function. Table 1 introduces some of the most important isotropic kernels. The first column of Table 1 shows the types of isotropic kernels. The second column is the kernel function of each isotropic kernel where $K(x)$ is a kernel function versus variable x and the third and fourth column are

shape and performance of isotropic kernels, respectively.The efficiency of estimator will evaluate with mean square error (MSE) [41].

Due to eq. (9), finding maximums of density function means determination the points with higher density (Fig. 6a). Figure 6b shows the movement of target vector toward the dense points to find the maximums of PDF.

To find these maximums, gradient of both sides of eq. (9) will calculate (eq. (14)).

$$\nabla \hat{f}_h(x) = \frac{1}{nh} \sum_{i=1}^{n} \nabla K(\frac{x-X_i}{h}) \tag{14}$$

With considering estimator function feature, we have.

$$\hat{f}_h(x) = \frac{1}{nh} \sum_{i=1}^{n} K(\|\frac{x-X_i}{h}\|^2) \tag{15}$$

Gradient formula can be expressed as follows:

$$\nabla \hat{f}_h(x) = \frac{2}{nh^3} \sum_{i=1}^{n} (x-X_i) K'(\|\frac{x-X_i}{h}\|^2) \tag{16}$$

With definition.

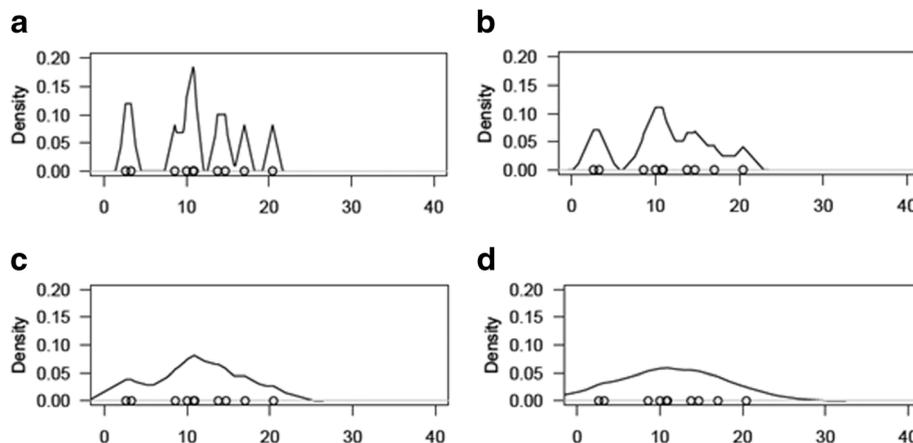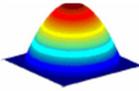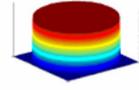$$g(x) = -K'(x) \tag{17}$$

We have:

Fig. 5 Comparison of the estimated function based on different values of h for Guassian kernel. **a** $h = 0.5$, (**b**) $h = 1$, (**c**) $h = 2$, (**d**) $h = 4$ [41]

Table 1 Different isotropic kernel and their performance [36].

Isotropic Kernel	Function	Shape	Performance
Epanechnikov	$K_E(x) = \begin{cases} c(1-\|x\|^2) & \|x\|\leq 1 \\ 0 & o\,w. \end{cases}$		1
Uniform	$K_U(x) = \begin{cases} c & \|x\|\leq 1 \\ 0 & o\,w. \end{cases}$		$(\frac{108}{125})^{\frac{1}{2}} \approx 0.9295$
Normal	$K_N(x) = \begin{cases} c.\exp(-\frac{1}{2}\|x\|^2) & \|x\|\leq 1 \\ 0 & o\,w. \end{cases}$		$(\frac{36\pi}{125})^{\frac{1}{2}} \approx 0.9512$
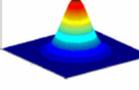
$$\nabla \hat{f}_h(x) = \frac{2}{nh^3} \sum_{i=1}^{n} (X_i - x).g(\|\frac{x-X_i}{h}\|^2)$$

$$= \frac{2}{nh^3} [\sum_{i=1}^{n} g(\|\frac{x-X_i}{h}\|2)][\frac{\sum_{i=1}^{n} X_i g(\|\frac{x-X_i}{h}\|^2)}{\sum_{i=1}^{n} g(\|\frac{x-X_i}{h}\|^2)} - x] \quad (18)$$

With definition *m(x)* as below.

$$m(x) = (\frac{\sum_{i=1}^{n} X_i g_i}{\sum_{i=1}^{n} g_i} - x) \quad (19)$$

Equation (18) will be as eq. (20).

$$\nabla \hat{f}_h(x) = \frac{1}{nh} [\sum_{i=1}^{n} g(\|\frac{x-X_i}{h}\|2)].m(x) \quad (20)$$

where *m(x)* is a vector that will guide the target toward the dense points. Therefore, *m(x)* known as mean-shift vector (Fig. 7) [36].

4.3.1 Mean-shift calculations

As explained in subsection 4.3, mean-shift algorithm uses color feature to find target location in subsequent frames. For usage of this feature, color histogram of object is calculated and weighted. Target localization will find with finding situation that has a most similarity to the target reference model. Therefore, a similarity criterion between target reference model $\overrightarrow{q}$ and target candidate $\overrightarrow{p}(\overrightarrow{y})$, $\rho(\overrightarrow{p}(\overrightarrow{y}), \overrightarrow{q})$ will define and measure in each step. Maximum of this function means the placement point of target. This algorithm will find the maximum points of function to reduce the calculations and processing time, with use of method based on gradient.

With definition of target reference model and target candidate as a monotonically decreasing kernel estimator and isotropic respectively, we have

$$q_u = C \sum_{i=1}^{n} K\left(\|\overrightarrow{x}_i\|^2\right) \delta[b(\overrightarrow{x}_i) - u] \quad (21)$$

$$p_u(\overrightarrow{y}) = C_h \sum_{i=1}^{n_h} K(\|\frac{\overrightarrow{y} - \overrightarrow{x}_i}{h}\|^2) \delta[b(\overrightarrow{x}_i) - u] \quad (22)$$

where C and C_h are normalized constants, n_h is number of pixels in the window, h is window size, $K(.)$ is a kernel estimator function, δ is a Dirac delta function and $b(\overrightarrow{x}_i)$ is area of histogram which pixel in $\overrightarrow{x}_i$ belongs to it. On the other hand
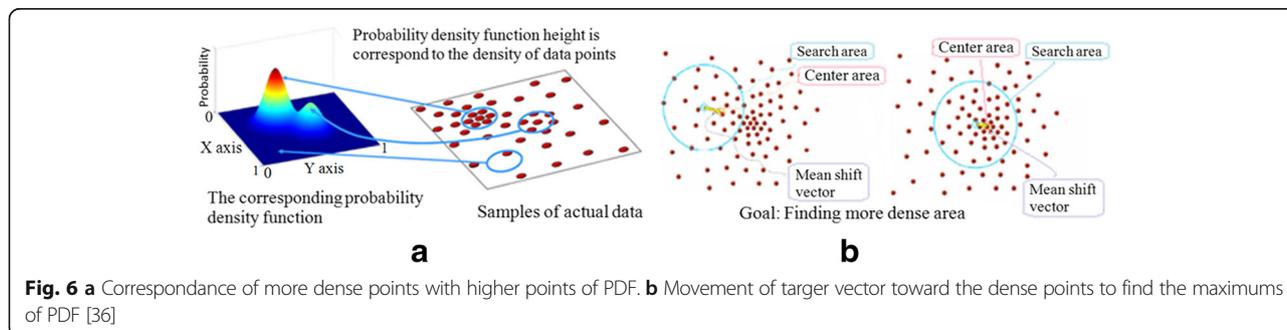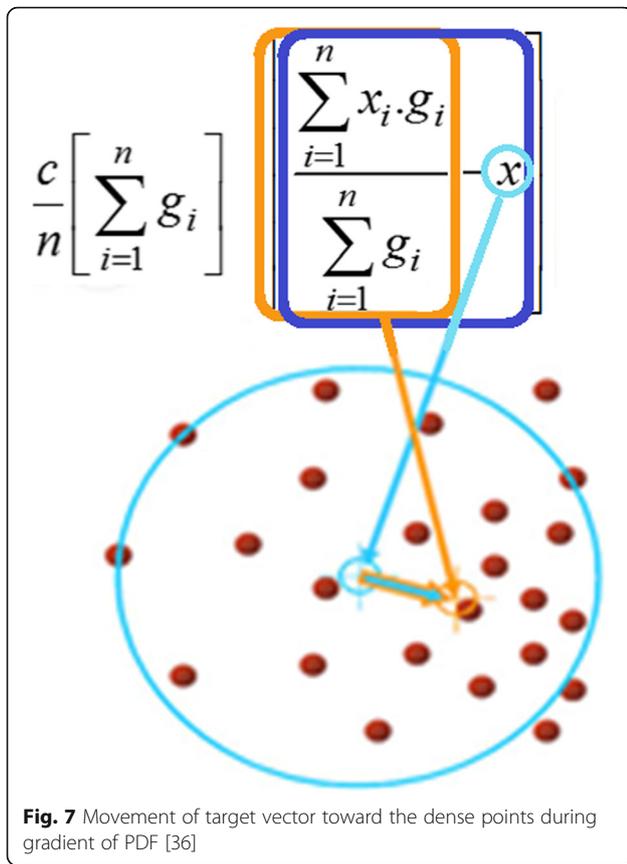

Fig. 6 a Correspondance of more dense points with higher points of PDF. **b** Movement of targer vector toward the dense points to find the maximums of PDF [36]

$$\frac{c}{n}\left[\sum_{i=1}^{n} g_i\right]\left\|\frac{\sum_{i=1}^{n} x_i \cdot g_i}{\sum_{i=1}^{n} g_i} - x\right\|$$

Fig. 7 Movement of target vector toward the dense points during gradient of PDF [36]

$$\delta(b(\vec{x}_i) - u) = \begin{cases} 1 & \text{if } \vec{x}_i \text{ belongs to the } u\text{-th histogram area} \\ 0 & \text{o.w.} \end{cases} \tag{23}$$

With considering two first sentences of Taylor series, similarity function around $p_u(\vec{y}_0)$ (probability density of target candidate initial position) will be

$$\rho(\vec{p}(\vec{y}), \vec{q}) \approx \sum_{u=1}^{m} \sqrt{p_u(\vec{y}_0) q_u}$$
$$+ \sum_{u=1}^{m} \left[\frac{1}{2} \times \frac{\sqrt{q_u}}{\sqrt{p_u(\vec{y}_0)}} (p_u(\vec{y}) - p_u(\vec{y}_0))\right] \tag{24}$$

After simplification of eq. (24), we have

$$\rho(\vec{p}(\vec{y}), \vec{q}) \approx \frac{1}{2} \sum_{u=1}^{m} \sqrt{p_u(\vec{y}_0) q_u} \tag{25}$$
$$+ \frac{1}{2} \sum_{u=1}^{m} p_u(\vec{y}) \sqrt{\frac{q_u}{p_u(\vec{y}_0)}}$$

With inserting eq. (21) and (22) into the eq. (25).

$$\rho(\vec{p}(\vec{y}), \vec{q}) \approx \frac{1}{2} \sum_{u=1}^{m} \sqrt{p_u(\vec{y}_0) q_u} \tag{26}$$
$$+ \frac{C_h}{2} \sum_{i=1}^{n_h} \omega_i K\left(\left\|\frac{\vec{y} - \vec{x}_i}{h}\right\|^2\right)$$

where ω_i is

$$\omega_i = \sum_{u=1}^{m} \sqrt{\frac{q_u}{p_u(\vec{y}_0)}} \delta[b(\vec{x}_i) - u] \tag{27}$$

First sentence of eq. (26) is a constant and independent of $\vec{y}$, therefore finding the maximums of similarity function is like as finding maximums of second sentence of eq. (26) which is density function estimator of current frame and is weighted with ω_i (eq. (27)). Modes of this function will obtain in the local neighborhood with mean-shift calculation [36]. In this process, isotropic kernel will move from target current position $\vec{y}_0$ to the new position $\vec{y}_1$ recursively as eq. (28).

$$\vec{y}_1 = \frac{\sum_{i=1}^{n_h} \vec{x}_i \omega_i g\left(\left\|\frac{\vec{y}_0 - \vec{x}_i}{h}\right\|^2\right)}{\sum_{i=1}^{n_h} \omega_i g\left(\left\|\frac{\vec{y}_0 - \vec{x}_i}{h}\right\|^2\right)} \tag{28}$$

where $g(.)$ is proportional to the derivative of estimator function. By taking an Epanechnikov estimation function, derivative of function is a fixed value; therefore, eq. (28) can be summarized as eq. (29) [35].

$$\vec{y}_1 = \frac{\sum_{i=1}^{n_h} \vec{x}_i \omega_i}{\sum_{i=1}^{n_h} \omega_i} \tag{29}$$

4.4 Extended mean-shift algorithm

Proposed tracking algorithm by Comaniciu et al. [36] will not adapt well with object shape variation and size of object. To overcome this problem, in mean-shift algorithm instead calculation local modes, with estimation of covariance matrix that includes local modes shape, this algorithm will extend.

In extended mean-shift algorithm, searching window will display by an ellipse with center of $\vec{\theta}$. As explained before, to increase the effect of central pixels of window and reduction effect of outlying pixels, the pixels are weighted. This can be carried out with putting the Guassian window on the pixels inside the ellipse. With consideration V as a covariance matrix, Guassian window in the position $\vec{x}_i$ will be as eq. (30). On the other

hand, target reference model vetor and target candidate vector are as eqs. (31) and (32), respectively.

$$N\left(\overrightarrow{x}_i, \overrightarrow{\theta}, V\right) = e^{\left(\overrightarrow{x}_i - \overrightarrow{\theta}\right)^T V^{-1} \left(\overrightarrow{x}_i - \overrightarrow{\theta}\right)} \qquad (30)$$

$$q_u = \sum_{i=1}^{N_v} N\left(\overrightarrow{x}_i, \overrightarrow{\theta}^*, V\right) \delta\left[b\left(\overrightarrow{x}_i\right) - u\right] \qquad 1 \le u \le m \qquad (31)$$

$$p_u\left(\overrightarrow{\theta}\right) = \sum_{i=1}^{N_v} N\left(\overrightarrow{x}_i, \overrightarrow{\theta}, V\right) \delta\left[b\left(\overrightarrow{x}_i\right) - u\right] \qquad (32)$$

where N_v is number of pixels inside the ellipse, i is the number of each pixel and $\overrightarrow{\theta}$ is center of ellipse that will determine by operator in target reference model. Also, $b\left(\overrightarrow{x}_i\right)$ determine the pixel area of belongingness with position $\overrightarrow{x}_i$ and $\delta\left(b\left(\overrightarrow{x}_i\right) - u\right)$ is a function that is one if $\overrightarrow{x}_i$ belongs to the area, otherwise is zero. The aim of tracking process is finding target candidate with the condition which has the most similarity to the target reference model. Similarity criterion is similarity function that is defined in eq. (33).

$$\rho\left(\overrightarrow{p}\left(\overrightarrow{\theta}\right), \overrightarrow{q}\right) = \sum_{u=1}^{m} \sqrt{p_u\left(\overrightarrow{\theta}\right)} \cdot \sqrt{q_u} \qquad (33)$$

With selection two first sentences of Taylor series for eq. (33), we have:

$$\rho\left(\overrightarrow{p}\left(\overrightarrow{\theta}\right), \overrightarrow{q}\right) \approx c_1 + c_2 \sum_{i=1}^{N_v} \omega_i N\left(\overrightarrow{x}_i, \overrightarrow{\theta}, V\right) \qquad (34)$$

where c_1 and c_2 are constant values and ω_i is as follow.

$$\omega_i = \sum_{u=1}^{m} \sqrt{\frac{q_u}{p_u\left(\overrightarrow{\theta}, V\right)}} \delta\left[b\left(\overrightarrow{x}_i\right) - u\right] \qquad (35)$$

In eq. (34), finding the maximum points of function in order to find the target with most similarity to the reference model is equal to maximization of eq. (36).

$$f\left(\overrightarrow{\theta}, V\right) = \sum_{i=1}^{N_v} \omega_i N\left(\overrightarrow{x}_i, \overrightarrow{\theta}, V\right) \qquad (36)$$

With consideration Gensen inequality [42]:

$$\log\left(f\left(\overrightarrow{\theta}, V\right)\right) \ge G\left(\overrightarrow{\theta}, a_1, \ldots, a_N\right) = \sum_{i=1}^{N_v} \log\left(\frac{\omega_i N(x_i, \theta, V)}{a_i}\right)^{a_i} \qquad (37)$$

where a_i's are optional coefficients that following equation holds from them.

$$\sum_{i=1}^{N_v} a_i = 1 \ , a_i \ge 0 \qquad (38)$$

With assumption estimation parameters as V_k and θ_k, two following essential steps will consider.

1) With assumption V_k and θ_k are constant, a_i coefficients calculate to maximize phrase G. Values of these coefficients are:

$$a_i = \frac{\omega_j N\left(\overrightarrow{x}_j, \overrightarrow{\theta}_k, V_k\right)}{\sum_{j=1}^{N_v} \omega_j N\left(\overrightarrow{x}_j, \overrightarrow{\theta}_k, V_k\right)} \qquad (39)$$

2) Having values of a_i coefficients, phrase g will maximize based on $V^{(k)}$ and $\theta^{(k)}$. Due to constant coefficients a_i, maximization will carry out with maximization of following function.

$$g\left(\overrightarrow{\theta}, V\right) = \sum_{i=1}^{N_v} a_i \log N\left(\overrightarrow{x}_i, \overrightarrow{\theta}, V\right) \qquad (40)$$

Therefore, with applying $\frac{\partial}{\partial \overrightarrow{\theta}} g\left(\overrightarrow{\theta}, V\right) = 0$, we have:

$$\overrightarrow{\theta}_{k+1} = \sum_{i=1}^{N_v} a_i \overrightarrow{x}_i = \frac{\sum_{i=1}^{N_v} \overrightarrow{x}_i \omega_i N\left(\overrightarrow{x}_i, \overrightarrow{\theta}_k, V_k\right)}{\sum_{i=1}^{N_v} \omega_i N\left(\overrightarrow{x}_i, \overrightarrow{\theta}_k, V_k\right)} \qquad (41)$$

Also for covariance function:

$$V_{k+1} = 2 \sum_{i=1}^{N_v} a_i \left(\overrightarrow{x}_i - \overrightarrow{\theta}_k\right) \left(\overrightarrow{x}_i - \overrightarrow{\theta}_k\right)^T \qquad (42)$$

Phrase V in eq. (42) has a role of bandwidth for estimator function [43]. Therefore, changing values of V can change the size of the search window. Figure 8 shows the performance of this algorithm in comparison to mean-shift algorithm. As shown in Fig. 8, this algorithm in each step adapts its dimension toward the target; therefore, it has robustness to reshaping and size and rotation of object. It should be mentioned that amount of calculations in comparison to the mean-shift algorithm, will not increase too much. Therefore, this algorithm can be used as a proper method for object tracking based on color feature.

4.5 Deficiencies of mean-shift algorithm

Mean-shift algorithm, despite a significant advantage in some circumstances it loses its capabilities and target-tracking stops [44, 45]. Because this method uses the color feature to describe the object, therefore under complex conditions, that color information of object is not able to
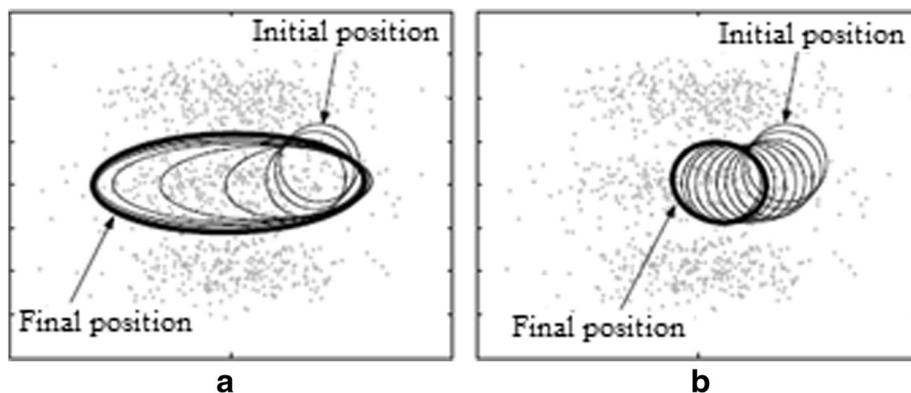
Fig. 8 Performance of two algorithms in target searching. **a** Mean-shift. **b** Extended mean-shift [55]

describe it; or this information does not give a precise description of the object position, this algorithm loses its efficiency. Usually this complex condition occurs when the color of target object is too close to the background color or under conditions that contrast of image is low so that color histogram cannot describe the precise object. On the other hand, this algorithm loses information about the movement of the object or location information. Under these conditions, the accuracy of the algorithm will drop sharply and target localization will be difficult.

Figure 9 shows a sample of object-tracking performance of mean-shift algorithm under complicated conditions. In this figure, algorithm start to track the football player in the playground, when the player move to the shadow of playground, because of the loss of differentiated attributes of target with background, the algorithm cannot correctly track the object and usually the target

position will obtain with errors. This error will continue to increase until the target miss.

Figure 10 shows an image of cyclist that moves toward a tree. Movement is such that distinction of object color and background color is lost; therefore, mean-shift algorithm stops the object tracking.

In this paper, a method will propose that with providing spatial information of object for mean-shift algorithm, this method is robust under complicated dynamic conditions and situations that color information will not present precise description of object.

5 Methods/experimental

For presentation of robust tracking method and elimination of algorithm limitations under complicated dynamic condition, in this section with providing object movement information for mean-shift algorithm, a combined algorithm is proposed.

Fig. 9 Performance of mean-shift algorithm in object tracking

Fig. 10 Performance of mean-shift algorithm in object tracking (from collection of PETS2001 databases [48])

Usage of object movement information could compensate lacks of spatial information in mean-shift algorithm. Also, with providing this information, algorithm will be robust under conditions where object color information is not enough for tracking. In facts, color and object movement information could be used as a complement of each other.

For providing object movement information, background subtraction method is used. Output image of Gaussian Mixture Model (GMM) background subtraction algorithm is a binary image which moving points have value 1 and fix points have value 0 [46]. Output image is not an ideal image and usually includes fake points (points where they have taken a wrong label 1). To obtain better results for output image includes movement information; post processing operations should apply to them.

5.1 Post-processing operation
5.1.1 Reduction of shade effect
Shade effect could label the pixel as a moving pixel and degrade the performance of algorithm. Therefore, it is necessary before subtraction operation, possibly shade effect reduces. To overcome the shade effect, with use of color intensity and brightness intensity simultaneously, shade pixels will determine.

5.1.2 Noise reduction and connected component analysis
In order to eliminate noise of image, in this paper median filter is used. Also to eliminate small and fake points of image and filling holes and empty region of

image, series of morphological operations is applied on the image. For elimination of fake points of image, morphological opening filter is used and for filling holes, morphological closing filter is applied on the image. Then, for unification of separate points of image, connected component analysis is used [47].

Figure 11 shows an effect of the post processing operation on the output image of applied background subtraction method before and after post processing operations. As it is clear from Fig. 11, post processing operation will eliminate fake points and fill empty regions and will highlight moving regions of the image.

5.2 Description of proposed algorithm
The obtained image from background subtraction introduces with b_i. Applying series of post processing operations will lead to the optimized binary image which includes moving areas presented with b_i. According to eq. (43), center of search window will move toward the moving region of image which is similar to mean-shift vector (eq. (19)).

$$\overrightarrow{y}_{k+1} = \frac{\sum_{i=1}^{n_h} \hat{b}_i \overrightarrow{x}_i}{\sum_{i=1}^{n_h} \hat{b}_i} \quad (43)$$

When object moves, center of window will move toward the moving area center. If the object in the next frame is

Fig. 11 Effect of post processing operation (**a**) The original image, (**b**) differentiated image, (**c**) differentiated image after post processing

fixed (it is not moved), output image has all zero value, and then target window position will not change (eq. (44)).

$$\overrightarrow{y}_{k+1} = \frac{\sum\limits_{i=1}^{n_h} \overrightarrow{x}_i}{n_h} = \overrightarrow{y}_k \quad (44)$$

To eliminate shortcomings of mean-shift algorithm and make it robust under complicated dynamic conditions, usage of spatial information with mean-shift algorithm simultaneously is proposed (eq. (41)).

Therefore, according to the eq. (45), target window center will calculate with consideration of color information and object movement information simultaneously in the next frame.

$$\overrightarrow{y}_{k+1} = \sum\limits_{i=1}^{n_h} a_i \overrightarrow{x}_i + \left(\frac{\sum\limits_{i=1}^{n_h} b_i \overrightarrow{x}_i}{\sum\limits_{i=1}^{n_h} b_i} - \overrightarrow{y}_k \right) \quad (45)$$

In the proposed eq. (45), first sentence is same as mean-shift equation and second sentence includes spatial information to compensate shortcomings of mean-shift algorithm.

At eq. (45), proposed method not only provide spatial information in conditions that color information is not enough and mean-shift algorithm will lost the target, but also will move the object position vector toward the target.

5.3 Explanation of proposed algorithm operation

Proposed algorithm should design such that if mean-shift algorithm works well, operation of algorithm continues and under condition that mean-shift algorithm could not track the target, help to improve algorithm. Therefore to explain the proposed method, there are general conditions as below.

1. Mean-shift algorithm tracks the target correctly

In this condition, according to Fig. 12, correction vector that obtained from binary image includes spatial information, will guide the algorithm toward the denser points (target).

2. Mean-shift algorithm is not capable to track the target
As shown in Fig. 12, when mean-shift algorithm track the target correctly, proposed algorithm is effective at convergence of algorithm, but as explained before, under condition that color histogram is not capable to have a discriminable description of object, mean-shift algorithm will lose its performance and gradually mean-shift vector will be away from the main target. Therefore, tracking error will increase significantly.

As shown in Fig. 13, under this condition, second sentence of eq. (45) (correction vector) will guide the position of search window toward the main target and algorithm will converge rapidly.

Also in this paper, to propose a method which could be robust against object size and shape variation, extended mean-shift algorithm as explained in the subsection 4.4 will replace with mean-shift algorithm.

This method will improve the performance of algorithm and increase its precision in different positions. In this method, instead of only local modes position estimation, with calculation of covariance matrix, local mode shapes will estimate and update in each step. On the other hand, implementation of method will not impose significant computational load to the system. As extended mean-shift algorithm in subsection 4.4 is explained, finding target position will be according to eq. (46).

$$\overrightarrow{\theta}_{k+1} = \sum\limits_{i=1}^{N_v} a_i \overrightarrow{x}_i + \left(\frac{\sum\limits_{i=1}^{N_v} b_i \overrightarrow{x}_i}{\sum\limits_{i=1}^{N_v} b_i} - \overrightarrow{\theta}_k \right) \quad (46)$$

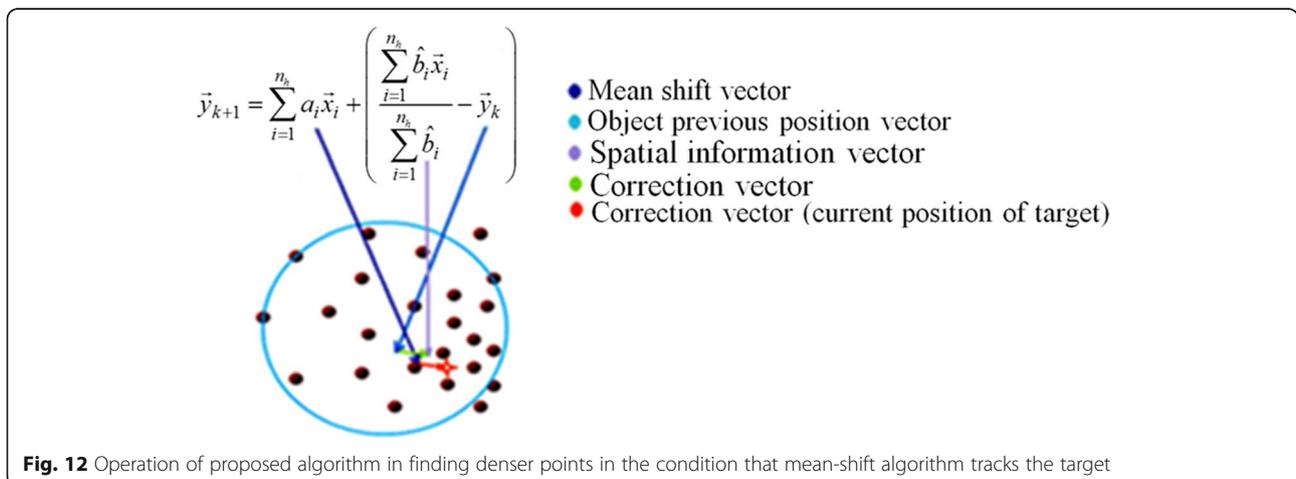where $\overrightarrow{\theta}_{k+1}$ is the position of $k+1$ from elliptic center

Fig. 12 Operation of proposed algorithm in finding denser points in the condition that mean-shift algorithm tracks the target
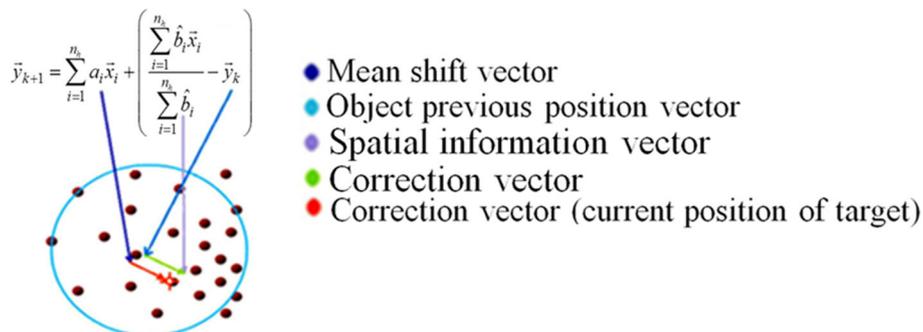
Fig. 13 Operation of proposed algorithm in finding denser points in the condition that mean-shift algorithm has lost target tracking

in the current frame. Updating covariance matrix is according to eq. (42).

Details of proposed algorithm are as below.

1. Target reference model q_u will determine
2. Evaluate value of similarity function with consideration search window position in the current frame and calculation of target candidate model p_u
3. Use of background subtraction algorithm GMM to create binary image which includes implemented movement information and post processing operation series to create optimize image
4. Weight values ω_i will calculate according to eq. (27)
5. With use of ω_i, a_i coefficients will calculate.
6. According to eq. (46), target candidate position $\vec{\theta}$ will determine
7. With use of $\vec{\theta}$ according to eq. (42) value V will calculate

8. In the case of convergence conditions, algorithm will stop, otherwise with substitution $k \leftarrow k + 1$ algorithm will repeat from step 2.

To avoid falling algorithm in the loop, usually a stop condition will consider. Therefore, if the number of repetitions is more than the defined limit, algorithm will stop.

Figure 14 shows a general schematic of proposed method. As shown in this figure, proposed algorithm is based on the initial parameters of mean-shift algorithm and spatial information. In order to evaluate proposed method, tracking error criterion is used which is Euclidean distance between resulting target position and real position.

5.3.1 Tracking parameters

In this paper, Epanechnikov kernel estimator function is used in the space of feature based on RGB color histogram. In order to reduce amount of computations, color space is
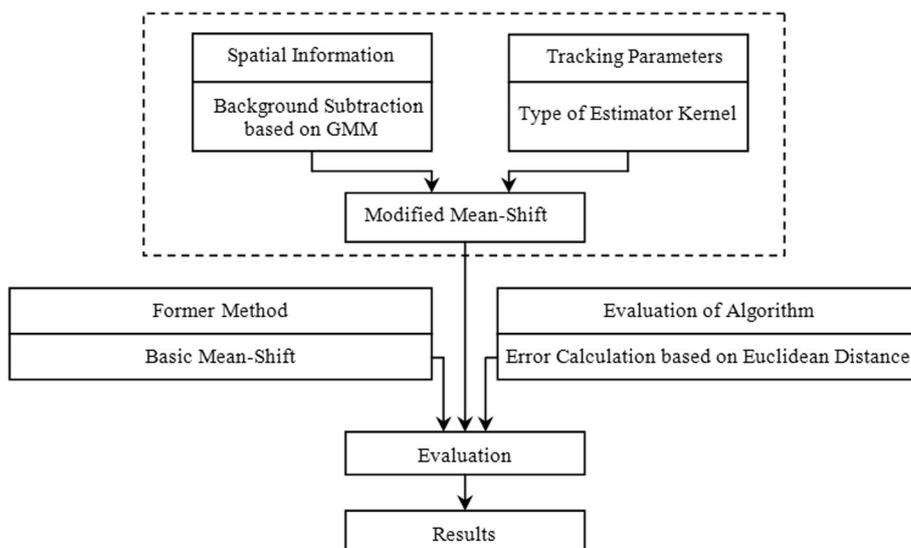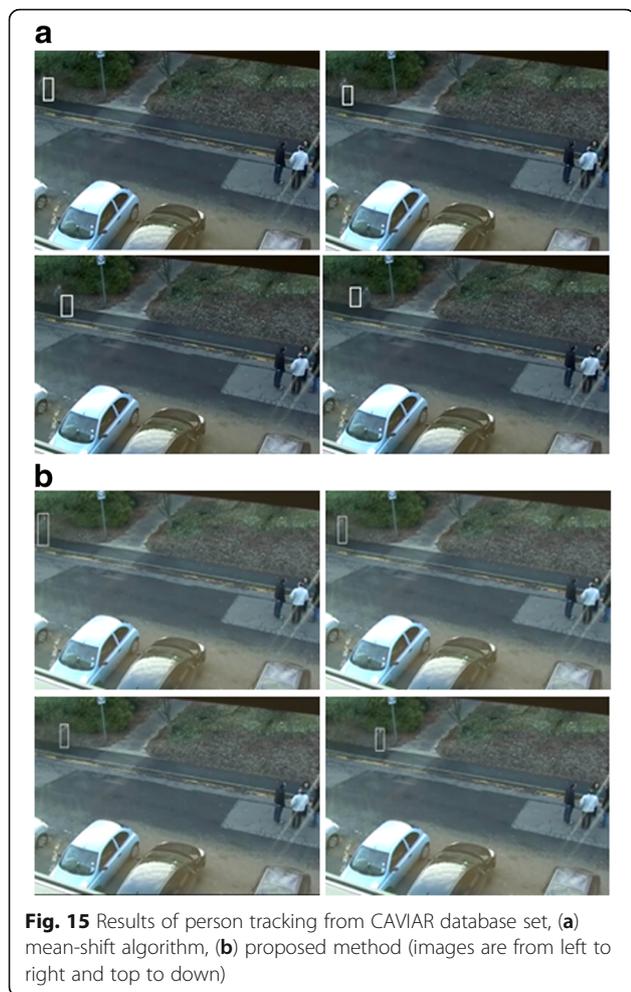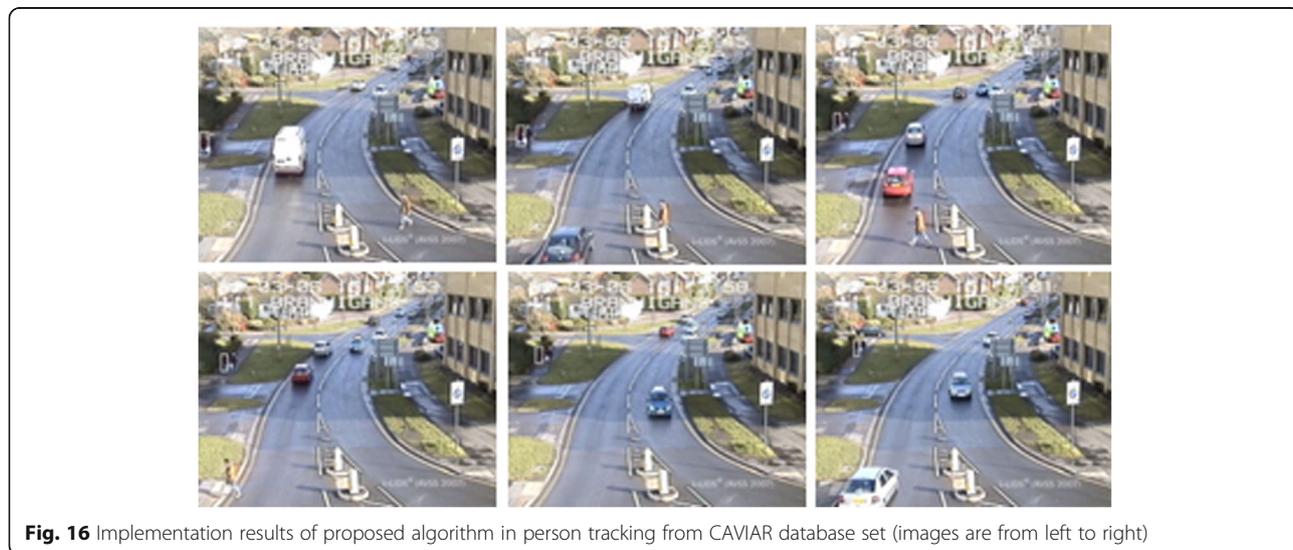

Fig. 14 General scheme of proposed method and evaluation of algorithm to achieve results

Fig. 15 Results of person tracking from CAVIAR database set, (**a**) mean-shift algorithm, (**b**) proposed method (images are from left to right and top to down)

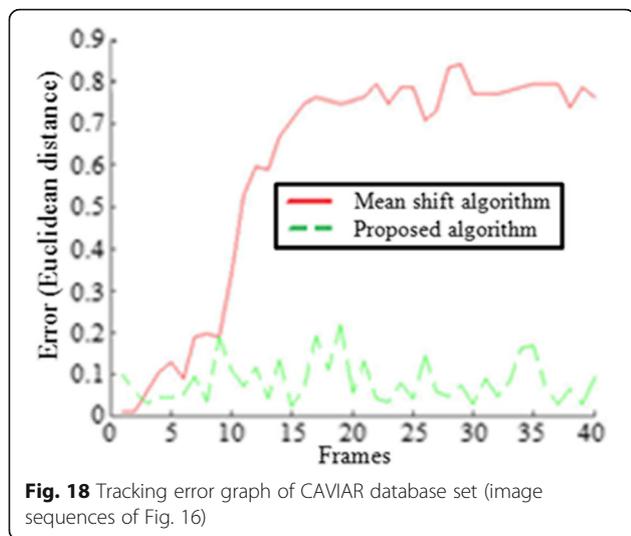**Fig. 17** Tracking error graph of CAVIAR database set (image sequences of Fig. 15)
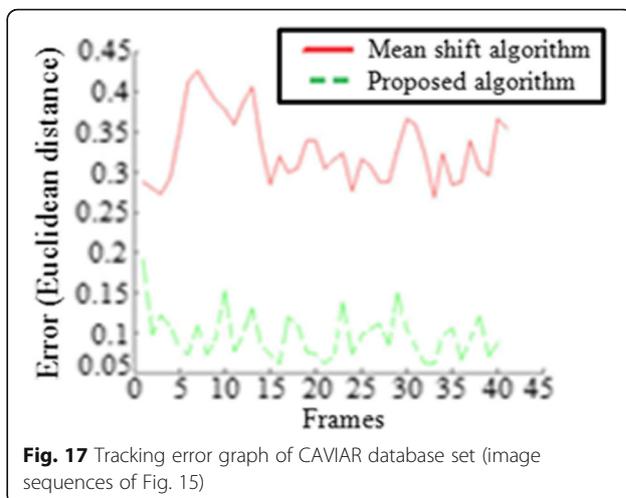
quantized into the $8 \times 8 \times 8$ space. Therefore, color histogram vector will reduce to the vector with length of 512.

Also in order to adapt search window size with object size variations, after establishing convergence condition, similarity distance criterion will calculate for three values smaller, larger and equal window size. Among them, the window size that has a minimum amount of similarity function will choose as a size of window. (Usually window size variations will be ±10% of window size). Also, in the proposed method, convergence condition is that similarity distance value in each step be less than threshold value 0.01.

At continuation, results of implementation for proposed algorithm on the number of video image sequences is presented specially under complicated condition on the MATLAB software environment.

Fig. 16 Implementation results of proposed algorithm in person tracking from CAVIAR database set (images are from left to right)

Fig. 18 Tracking error graph of CAVIAR database set (image sequences of Fig. 16)

6 Results and discussion

In this section, results of proposed method will introduce and finally some recently related method with review and discuss. For evaluation of proposed method, results of mean-shift algorithm are presented. Proposed method is evaluate on the set of standard video database CAVIAR [48] under complex dynamic condition which discrimination of target color with background has been reduced significantly. Some samples of CAVIAR database are shown in Figs. 15 and 16.

According to Fig. 15a, mean-shift algorithm has low performance in tracking while precision of proposed algorithm (Fig. 15b) is more than mean-shift algorithm in target tracking. For quantitate comparison of proposed method with mean-shift algorithm, error of these methods is shown in Figs. 17 and 18 for image sequences of Figs. 15 and 16, respectively. Tracking error

is calculated with Euclidean distance between obtained position from algorithm and real position.

As shown in Figs. 16 and 18, proposed algorithm based on mean-shift has good results in target tracking in comparison to initial mean-shift algorithm. According to the evaluation results of proposed method on the other image sequences, there were circumstances which in that condition, proposed algorithm had error in target tracking but not more than mean-shift algorithm.

Figure 19 includes a video image of vehicle, which in these images target during movement does not have a fix shape. As explained in subsection 4.4 extended mean-shift algorithm is more robust than mean-shift algorithm under object deformation. Therefore, in the proposed method, mean-shift is replaced with extended mean-shift. Figure 19 shows the performance of the proposed method in vehicle tracking.

In Fig. 20, the tracking error graph for proposed methods and men shift algorithm is shown.

According to Fig. 20, proposed algorithm based on mean-shift does not have proper performance. The reason is weak performance of mean-shift algorithm under object deformation. With respect to that mean-shift algorithm tracks the target based on color histogram of target region; therefore, target deformation will create essential variation in the histogram and mean-shift will have serious error and coverage of algorithm to find target position takes a long time and algorithm will stop before finding the target. Figure 21 shows an essential variation of histogram when the image is deformed.

As shown in Fig. 21, region of histogram that is related to the object is changed essentially with object deformation; therefore, extended mean-shift algorithm could be a better option to improve this situation. Significant reduction in tracking error for proposed algorithm based on extended mean-shift in Fig. 20 confirms it correctly.
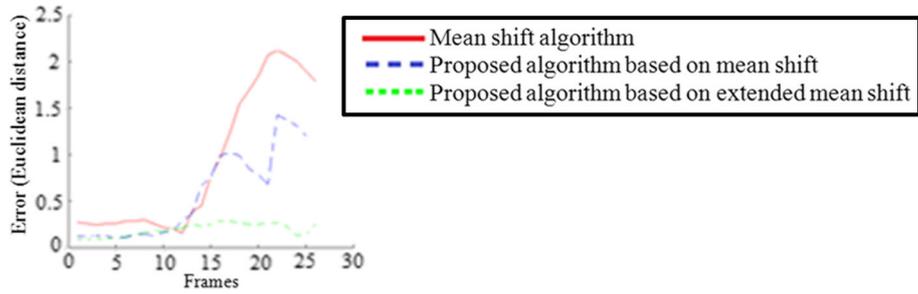

Fig. 19 Vehicle tracking results of proposed method based on extended mean-shift algorithm (images are from left to right)

Fig. 20 Comparison of methods based on tracking error in object deformation for image sequences of Fig. 19
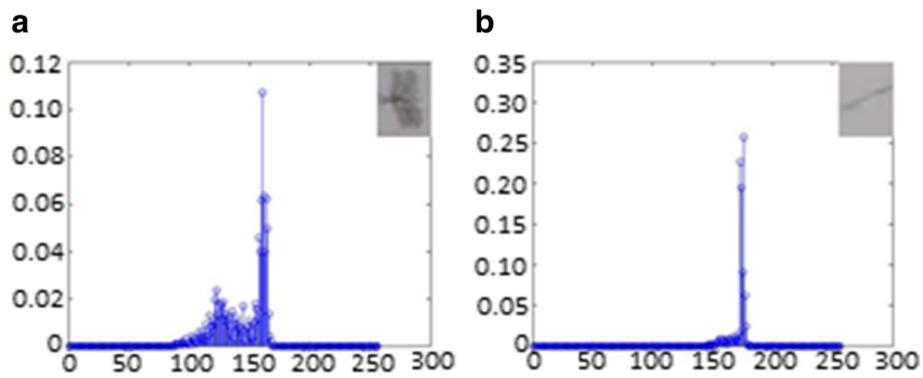

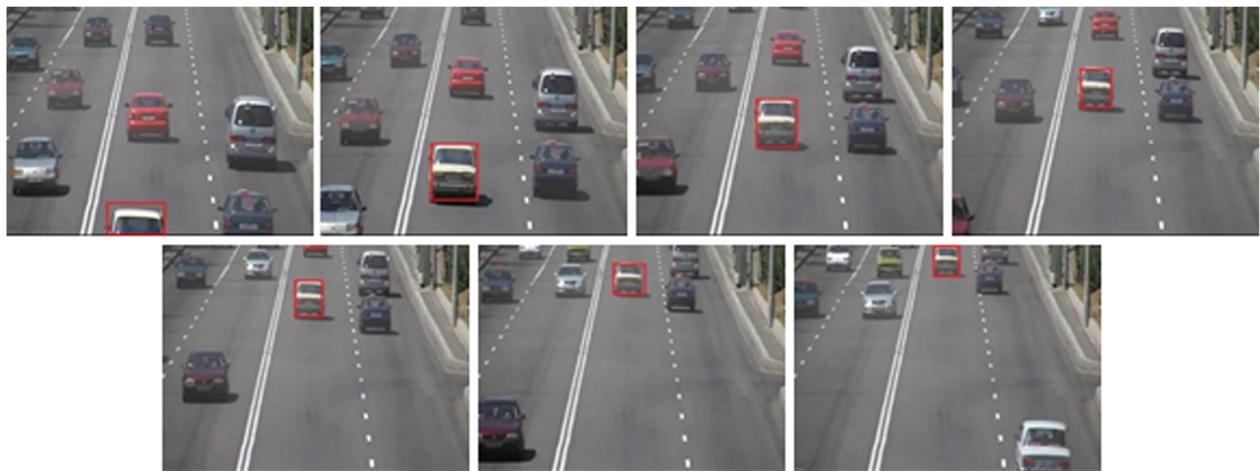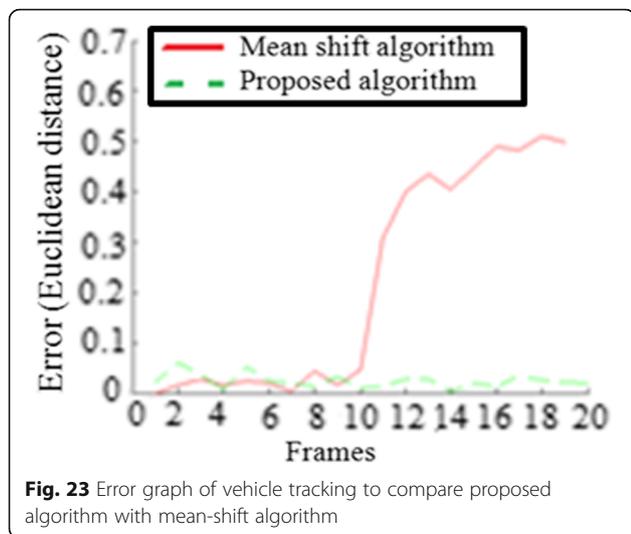**Fig. 21** Object histogram variation based on deformation, (**a**) histogram of original image (**b**) histogram of deformed image

Fig. 22 Operation of proposed algorithm in vehicle tracking at congestion (images are from left to right)

Fig. 23 Error graph of vehicle tracking to compare proposed algorithm with mean-shift algorithm

New scenario is organized based on extended mean-shift algorithm; therefore tracking process will carry out according to eq. (46). First of all, performance of proposed method will evaluate on the sequences of images for the busy road. Mean-shift algorithm has a significant error in vehicle tracking. Figure 22 shows some frames of proposed algorithm in vehicle tracking. Error graph of proposed method and mean-shift algorithm is shown in Fig. 23.

As shown in Fig. 23, object position based on proposed method is tracked correctly. When the color information of target in comparison to background of target is the same, error for mean-shift algorithm has a significant jump and mean-shift algorithm is not capable to track the target while proposed algorithm has a less error and is found the target correctly.

For verification of algorithm under normal condition (condition which color information is enough and mean-shift algorithm track the target well), proposed

method is tested on the some images sequence from PETS2001 standard database [48] that results are shown in Fig. 24.

In Fig. 24, target is a vehicle that tracking is based on the area which has a distinct color from background.

In a new scenario, performance of proposed algorithm will evaluate under low contrast condition. In this situation, tracking in the night will consider which background light is low and complex dynamically situations in the image are expected. To evaluate performance of proposed algorithm based on mean-shift and extended mean-shift, in Fig. 25, different conditions in tested.

In Fig. 26, tracking error graph of methods is shown and as it is clear in the low contrast condition, proposed method based on extended mean-shift algorithm has a lowest error and best precision in target tracking.

Since mean-shift algorithm is a method based on colors, therefore under the condition that color information of object could not have a precise description of the object or object color information to be close to the background color, this algorithm stops object tracking. In order to address the limitation of detection methods based on color histogram, some strategies are proposed that includes hybrid algorithms.

Nummiaro et al. [49] use the particle filter to estimate object location in mean-shift algorithm, which has led to the good resistance of algorithm in the occlusion condition. Although in this algorithm, there is still a problem of the loss of color information under complicated condition, this method has highly complicated computations. Therefore, in order to reduce the amount of computations related to particle filter, Fa-Liang et al. [50] proposed the usage of two different movement patterns of particles to estimate the object location. Combination of color and texture information simultaneously for precise object tracking based on mean-shift algorithm is proposed by Ning et al. [45]. Although in this method, color information will amplify but still the
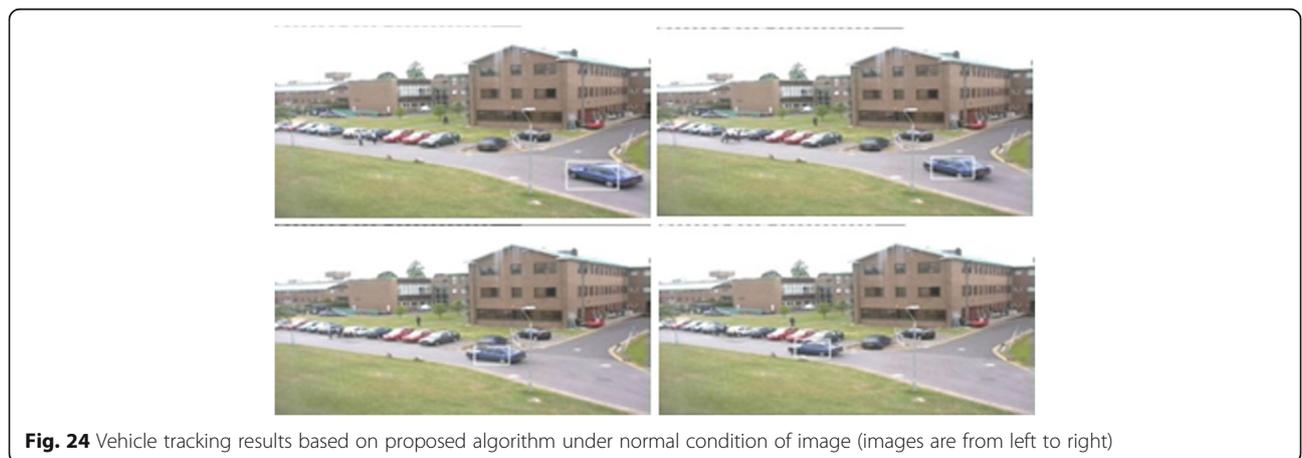

Fig. 24 Vehicle tracking results based on proposed algorithm under normal condition of image (images are from left to right)

Fig. 25 Target tracking results under low contrast of background (images are from left to right)

problem of information loss during the object movement has not been studied.

Chen et al. [51] proposed usage of Kalman filter in mean-shift algorithm. In this method, first with Kalman filter, object location will estimate, then exact location of target with use of mean-shift algorithm will determine. Although in this method use of Kalman filter will provide the location of object but using only the color information as a key feature will reduce accuracy of algorithm under complicated condition of image and makes a problem for this method. Also Zheng et al. [52] describe the object with two color and tissue features to prevent problems caused by the absence of color information and brightness fluctuations. Xuguang et al. [53] first propose a model based on oriented histogram and then provides a tracking algorithm based on mean-shift for gray images. Ju et al. [54] proposed tracking algorithm based on fuzzy histogram to reduce noise interference of mean-shift algorithm. This method under low contrast condition lost its performance and needs high computations.

7 Conclusions

In this paper, a method for robustness of mean-shift algorithm under dynamically complex condition is proposed. Results of simulation for proposed algorithm on the video images sequences have shown the success of proposed method with precise object tracking. Also implementation of proposed method on the sequences of images which the shape of object changes significantly during consecutive frames is shown that proposed algorithm with use of mean-shift algorithm has a low precision. To overcome this limitation, method which is adaptive with object deformation should be replaced with mean-shift algorithm. Therefore, usage of extended mean-shift algorithm is proposed. The results in the normal and complicated conditions that color information is lost and under low contrast condition proposed method with use of extended mean-shift algorithm showed vehicle tracking is carried out properly and error of tracking is less. Therefore, proposed method can be used as a precise tracking method in the normal and complicated conditions.
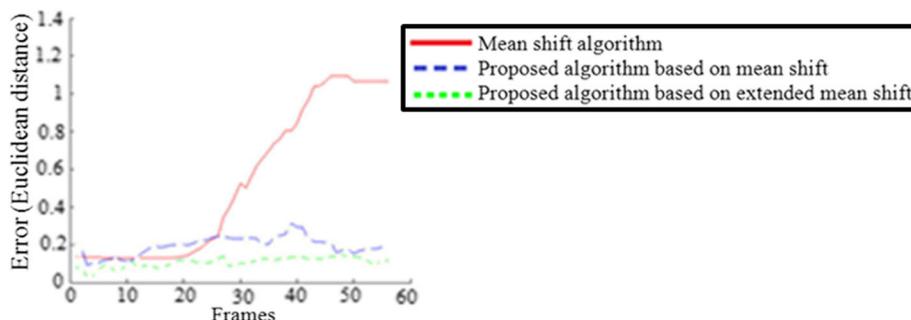

Fig. 26 Comparison of different algorithms in object tracking under low contrast condition based on tracking error

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

References
1. H Cho, SY Hwang, High-performance on-road vehicle detection with non-biased Cascade classifier by weight-balanced training. EURASIP J Image Video Process **16**, 1 (2015). https://doi.org/10.1186/s13640-015-0074-5
2. X Zhuang, W Kang, Q Wu, Real-time vehicle detection with foreground-based Cascade classifier. IET Image Process. **10**(4), 289–296 (2016)
3. T Tang, S Zhou, Z Deng, H Zou, L Lei, Vehicle detection in aerial images based on region Convolutional neural networks and hard negative example mining. J Sensors **17**(2), 336 (2017). https://doi.org/10.3390/s17020336
4. AC Chachich, A Pau, K Kenedy, E Olejniczak, J Hackney, Q Sun, E Mireles, Traffic sensor using a color vision method. Transp Sensors Controls **29**, 156–165 (1996)
5. Z Sun, G Bebis, R Miller, On road Vehicle Detection Using Optical Sensors: A Review, The 7th International IEEE Conference on Intelligent Transportation Systems (Institute of Electrical and Electronics Engineers (IEEE), Washington, 2004)
6. J Sochor, Fully Automated Real-Time Vehicles Detection and Tracking with Lanes Analysis, 18th Central European Seminar on Computer Graphics (CESCG) (Technical University Wien, Smolenice, Slovakia, 2014)
7. M-T Yang, R-K Jhang, J-S Hou, Traffic flow estimation and vehicle-type classification using vision-based spatial–temporal profile analysis. IET Comput. Vis. **7**(5), 394–404 (2013)
8. A Jazayeri, H Cai, JY Zheng, M Tuceryan, Vehicle detection and tracking in car video based on motion model. IEEE Trans Intell Transp Syst **12**(2), 583–595 (2011)
9. C-C Chiu, M-Y Ku, C-Y Wang, Automatic traffic surveillance system for vision-based vehicle recognition and tracking. J. Inf. Sci. Eng. **26**, 611–629 (2010)
10. W Zhang, QMJ Wu, G Wang, X You, Tracking and pairing vehicle headlight in night scenes. IEEE Trans. Intell. Transp. Syst. **13**(1), 140–153 (2012)
11. R O'Malley, E Jones, M Glavin, Rear-lamp vehicle detection and tracking in low-exposure color video for night conditions. IEEE Trans. Intell. Transp. Syst. **11**(2), 453–462 (2010)
12. G Salvi, An Automated Nighttime Vehicle Counting and Detection System for Traffic Surveillance, International Conference on Computational Science and Computational Intelligence (Conference Publishing Services (CPS), Las Vegas, 2014)
13. C Yan, Y Zhang, J Xu, et al., A highly parallel framework for HEVC coding unit partitioning tree decision on many-core processors. IEEE Signal Process Letters **21**(5), 573–576 (2014)
14. C Yan, Y Zhang, J Xu, et al., Efficient parallel framework for HEVC motion estimation on many-Core processors. IEEE Trans Circuits Syst Video Technol **24**(12), 2077–2089 (2014)
15. L Dan, J Qian, Research on Moving Object Detecting and Shadow Removal. 1st International Conference on Information Science and Engineering (Institute of Electrical and Electronics Engineers (IEEE), Nanjing, 2009)
16. T Kodama, T Yamaguchi, H Harada, A method of Object Tracking Based on Particle Filter and Optical Flow. ICCAS-SICE, p. 2685–2690 (Institute of Electrical and Electronics Engineers (IEEE), Fukuoka, 2009)
17. D Gao, J Zhou, Adaptive Background Estimation for Real-Time Traffic Monitoring, IEEE Intelligent Transportation Systems, p. 330–333 (Institute of Electrical and Electronics Engineers (IEEE), Oakland (CA), 2001)
18. T Kawanishi, T Kurozumi, K Kashino, S Takagi, A Fast Template Matching Algorithm with Adaptive Skipping using Inner-sub Template's Distances. Proceedings of the IEEE International Conference on Pattern Recognition, p. 654–657. (Institute of Electrical and Electronics Engineers (IEEE), England, 2004)
19. H Carrillo, J Villarreal, M Sotaquira, MA Goelkel, R Gutierrez, A Computer Aided Tool for the Assessment of Human Sperm Morphology. Proceedings of the 7th IEEE International Conference on Bioinformatics and Bioengineering, p. 1152–1157. (Institute of Electrical and Electronics Engineers (IEEE), Boston, 2007)
20. W Long, Y-H Yang, Stationary background generation: An alternative to the difference of two images. Pattern Recogn. **23**(2), 1351–1359 (1990)
21. K-P Karmann, A Brandt, R Gerl, Moving Object Segmentation Based on Adaptive Reference Images. In Proceeding 5th European Signal Processing Conference, p. 951–954. (Elsevier Science Publishers, Barcelona, 1990)
22. PG Michalopoulos, Vehicle detection video through image processing: The auto scope system. IEEE Trans. Veh. Technol. **40**(1), 21–29 (1991)
23. J Kan, J Tang, K Li, X Du, *Background Modeling Method Based on Improved Multi-Gaussian Distribution* (International Conference on Computer Application and System Modeling (ICCASM), Taiyuan, 2010), pp. 22–24
24. S Gupte, O Masoud, RFK Martin, NP Papanikolopoulos, Detection and classification of vehicles. IEEE Trans Intell Transp Syst **3**(1), 37–47 (2002)
25. SHA Musavi, BS Chowdhry, J Bhatti, Object Tracking based on Active Contour Modeling, International Conference on Wireless Communications, Vehicular Technology, Information Theory and Aerospace & Electronic Systems (VITAE) (Institute of Electrical and Electronics Engineers (IEEE) Aalborg, Denmark, 2014)
26. M Kass, A Witkin, D terzopoulos, Snakes: Active Contour Models. Proceeding of European Conference on Automation (Kluwer Academic Publishers, Birmingham, 1987)
27. M Rousson, N Paragios, Shape Priors for Level Set Representation. Proceeding of European Conference on Computer Vision (Springer-Verlag Berlin Heidelberg, Copenhagen, 2002)
28. V Caselles, R Kimmel, G Sapiro, Geodesic active contours. Int. J. Comput. Vis. **22**(1), 61–79 (1997)
29. B Han, L Davis, Object Tracking by Adaptive Feature Extraction. International Conference on Image Processing (Institute of Electrical and Electronics Engineers (IEEE), Singapore, 2004.
30. I Kim, MM Khan, TW Awan, et al., Multi-target tracking using color information. Int J Comput Commun Eng **3**(1), 11–15 (2014)
31. M Mason, Z Duric, *Using Histograms to Detect and Track Objects in Color Video* (Applied Imagery Pattern Recognition Workshop, Washington DC, 2001)
32. P Withagen, K Schutte, F Groen, *Likelihood-Based Object Detection and Object Tracking Using Color Histograms and EM* (International Conference on Image Processing, Rochester, 2002), pp. 22–25
33. A Yilmaz, O Javed, M Shah, Object tracking: A survey. J ACM Comput Surv (CSUR) **38**(4), 1–45 (2006)
34. Q Wang, RK Ward, Fast image/video contrast enhancement based on weighted Thresholded histogram equalization. IEEE Trans. Consum. Electron. **53**(2), 757–764 (2007)
35. D Comaniciu, V Ramesh, P Meer, Kernel-based object tracking. IEEE Trans. Pattern Anal. Mach. Intell. **25**(5), 564–577 (2003)

36. D Comaniciu, P Meer, Mean-shift: A robust approach toward feature space analysis. IEEE Trans. Pattern Anal. Mach. Intell. **24**(5), 603–619 (2002)

37. MR Leadbetter, GS Watson, On the estimation of the probability density. Ann Math Stat **34**(2), 408–491 (1963)

38. VA Epanechnikov, Non-parametric estimation of a multivariate probability density. Theory Probability Appl **14**(1), 153–158 (1969)

39. M Rosenblatt, Remarks on some non-parametric estimation of a density function. Ann Math Stat **27**(3), 832–837 (1956)

40. E Parzen, On estimation of a probability density function and mode. Ann Math Stat **33**(3), 1065–1076 (1962)

41. W Zucchini, Applied smoothing techniques Part 1: Kernel density estimation (Temple University Press, Philadephia, 2003)

42. TM Cover, JA Thomas, *Elements of Information Theory* (Wiley, New York, 1991)

43. MP Wand, MC Jones, Kernel Smoothing, (Chapman & Hall/CRC, London; New York, 1995).

44. L Wei, L Yi-ning, S Nan, Mean-shift tracking algorithm based on background optimization. J Comput Appl **29**(4), 1015–1017 (2009)

45. J Ning, L Zhang, D Zhang, C Wu, Robust object tracking using joint color-texture histogram. Int. J. Pattern Recognit. Artif. Intell. **23**(7), 1245–1263 (2009)

46. C Stauffer, WEL Grimson, *Adaptive Background Mixture Models for Real-Time Tracking* (IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Colorado, 1999), pp. 246–252

47. RC Gonzalez, RE Woods, Digital image processing, Third Edition, (Prentice-Hall, New Jersey, 2008)

48. http://datasets.visionbib.com. Accessed 21 Mar 2017.

49. K Nummiaro, E Koller-Meier, LV Gool, An adaptive color-based particle filter. Image Vis. Comput. **21**, 99–110 (2003)

50. C Fa-Liang, Z Yao, C Zhen-Xue, et al., Non-rigid object tracking algorithm based on mean shift and adaptive prediction. Control Decis **24**(12), 1821–1825 (2009)

51. L Chen, W Li, W Yin, *Joint Feature Points Correspondences and Color Similarity for Robust Object Tracking* (International Conference on Multimedia Technology (ICMT), Hangzhou, 2011), pp. 403–407

52. L Yuan-Zheng, L Zhao-Yang, G Quan-Xue, L Jing, Particle filter and mean shift tracking method based on multi-feature fusion. J. Electron. Inf. Technol. **32**(2), 411–415 (2010)

53. Z Xuguang, Z Enliang, W Yanjie, A New, Algorithm for tracking gray object based on mean-shift. Opt Tech **33**(2), 226–229 (2007)

54. M-Y Ju, C-S Ouyang, H-S Chang, *Mean-Shift Tracking Using Fuzzy Color Histogram* (International Conference on Machine Learning and Cybernetics, Qingdao, 2010), pp. 2904–2908

55. Z Zivkovic, B Krose, An EM-like Algorithm for Color-histogram-based Object Tracking. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, (CVPR) (Institute of Electrical and Electronics Engineers (IEEE), Washington, 2004), pp. 798–803