

# Scale and Orientation-Based Background Weighted Histogram for Human Tracking

Khadija Laaroussi · Abderrahim Saaidi · Mohamed Masrar · Khalid Satori

Received: 11 April 2016/Revised: 29 June 2016/Accepted: 4 July 2016/Published online: 15 July 2016  
© 3D Research Center, Kwangwoon University and Springer-Verlag Berlin Heidelberg 2016

**Abstract** The Mean Shift procedure is a popular object tracking algorithm since it is fast, easy to implement and performs well in a range of conditions. However, classic Mean Shift tracking algorithm fixes the size and orientation of the tracking window, which limits the performance when the target's orientation and scale change. In this paper, we present a new human tracking algorithm based on Mean Shift technique in order to estimate the position, scale and orientation changes of the target. This work combines moment features of the weight image with background information to design a robust tracking algorithm entitled Scale and Orientation-based Background Weighted Histogram (SOBWH). The experimental

results show that the proposed approach SOBWH presents a good compromise between tracking precision and calculation time, also they validate its robustness, especially to large background variation, scale and orientation changes and similar background scenes.

**Keywords** Human tracking · Mean Shift algorithm · Scale and orientation estimation · Moment features · Background information

## 1 Introduction

Object tracking in image sequences is, in recent decades, a very active research area in computer vision that arises in many applications such as human computer interaction, security and surveillance, traffic control and smart rooms. Tracking is to estimate the target locations in each image of a video sequence. The localization process is based on the object recognition from a set of visual features such as color, shape, speed, etc. Although, many tracking algorithms have been proposed in the literature [4, 12, 13, 17, 28]. Among them the Mean Shift algorithm [4] turns out to be robust and accurate compared other algorithms [17, 28]. This method was proposed by Fukunaga et al. [5], and it was used for the first time in 1997 in the context of image segmentation. Then, it was adopted by Comaniciu et al. [4] for real-time tracking of

---

K. Laaroussi (✉) · A. Saaidi · M. Masrar · K. Satori  
LIAN, Department of Mathematics and Computer  
Science, Faculty of Science Dhar El Mahraz, Sidi  
Mohamed Ben Abdellah University, P.O 1796,  
Atlas, Fez, Morocco  
e-mail: khadija.laaroussi3@usmba.ac.ma

A. Saaidi  
e-mail: abderrahim.saaidi@usmba.ac.ma

M. Masrar  
e-mail: masrar.m@hotmail.com

K. Satori  
e-mail: khalidsatorim3i@yahoo.fr

A. Saaidi  
LSI, Laboratory of Engineering Sciences,  
Polydisciplinary Faculty of Taza, Sidi Mohamed Ben  
Abdellah University, P.O 1223, Taza, Morocco

deformable objects in an image sequence using the color density of the target. The tracking is carried out from its initial position in the first image. The target is modeled as an ellipse, on which calculates its color distribution. The initial color distribution is referenced as a model, and then compared with that of the candidates to determine the most probable position in the next frame. Originally, Mean Shift algorithm, iteratively seeks to maximize a similarity measure between the target model and target candidates.

Despite the great success achieved by the Mean Shift algorithm [4] in many situations (noise, illumination change, rotation, partial occlusion), its performance suffers from the use of a fixed size window, thus a window size that works at one scale is not suitable for another as the target moves towards and away from the camera. If the kernel size is chosen too large, the tracking window will contain many background pixels as well as the foreground object pixels, which often leads the tracking algorithm to converge to an area between multiple modes, rather than converging to just one of the modes. If the kernel size is chosen too small, the localization becomes poor since some pixels on the object are not included in the search window and the similarity function often has many local maxima. Furthermore, other common problems to all tracking algorithms are, first, the background influence because the target can't be represented accurately without any background information and if the correlation between target and background is high, the localization accuracy of the object will be decreased. Second, the appearance of distractor elements, this problem arises when the tracked scene or the background includes similar objects to the target from the viewpoint of appearance. These objects can be confused with the target and may distract the tracking algorithm, hence their qualification distractor. In summary, all these limitations represent a major challenge causing the drift of the tracking algorithm and which increase the tracking difficulties.

The main objective of this work is to propose a new human tracking algorithm based on Mean Shift technique [4]. Moreover, this work aims to develop a technique cope well with the difficulties discussed in the paragraph above, operating under a minimum control over the object and its environment and without a priori knowledge of the appearance model. The proposed work represents another tracking algorithm compared to our previous work presented in [11].

The main contributions of this paper are as follows. First, we have used moment features of the weight image [2] to estimate the position, scale and orientation changes of the target. Therefore, considering that the weight image derived from the target model and the target candidate represents the probability that a pixel belongs to the target. Hence, the target scale is calculated using the zeroth order moment and the Bhattacharyya coefficient [10], then the width, height and orientation were estimated by the second order moments and the estimated area. Second, a representation model of background named Background-Weighted Histogram (BWH) [19] has been incorporated into the target representation. The proposed approach combines the moment features of the weight image [2] with BWH [19] to generate a robust tracking algorithm named Scale and Orientation-based Background Weighted Histogram (SOBWH). Furthermore, this new approach SOBWH used to effectively exploit the main features of the background, so that tracking can't be confused with stable objects in the background. In addition, by learning the background with a weighting mechanism [19], we can hide the intensity values related to the background and help to discriminate between the target and its background, which improves the position, scale and orientation estimation in the presence of distractor elements. Our experimental results validate the performance of the implemented method SOBWH under complex tracking scenarios namely, large background variation, scale and orientation changes, target and camera motion, distinction between the object and its background, partial occlusion, illumination change, appearance of distractor elements, rotation and deformation.

The rest of the paper is organized as follows: Sect. 2 reviews some related work. Section 3 briefly introduces the traditional Mean Shift tracking algorithm. Section 4 presents the Mean Shift tracking algorithm for scale and orientation changes of the target. Section 5 investigates the proposed method. The experiment results and discussion are as followed in Sect. 6 and Sect. 7 concludes the paper.

## 2 Related work

The works presented in this section are among the tracking methods based on the appearance of an object, especially those that are based on scale

adaptation and orientation changes of the target. However, the researches to address these problems can be classified into three categories [29].

The first category is the incremental heuristics methods. In the work by Comaniciu et al. [4], Mean Shift algorithm is performed at three different scales (the previous scale, a slightly larger scale and a slightly smaller scale) and for each different window size, the similarity measure based on the Bhattacharyya coefficient [10] is computed for comparison. The window size yielding the largest Bhattacharyya coefficient is chosen as the updated scale. This principle has been used by many authors [1, 21]. However, this approach does not work well with the increase of the object size since the smaller windows usually have higher similarity and therefore the scale is often underestimated. Collins [3] extended the Mean Shift algorithm by adapting Lindeberg's theory of feature scale selection based on local maxima of differential scale-space filters. Then, Mean Shift is applied in the spatial and scale dimensions to the weight image. It uses blob tracking and a scale kernel to accurately capture the target's variation in scale. However, this method requires recomputing Gaussian kernels at every Mean Shift iteration, which is computationally intensive, also it is can't handle the rotation changes of the target. One more similar work is adaptive pyramid Mean Shift tracking which is proposed by Li et al. [16]. Different thresholds for the scale incremental are applied and this scale selection method can track objects of varying sizes while the complexity remains unchanged. However, the tracking process may be interrupted when illumination change seriously.

The second category is spatial moment methods. Bradski [2] modified the Mean Shift algorithm and produced the Continuously Adaptive Mean Shift (CAMSHIFT) algorithm for face tracking, which only consider the moment of the weight image determined by target model to estimate the scale and orientation of the target. Zivkovic et al. [32] proposed a similar algorithm that estimates the scale and orientation changes named Expectation–Maximisation Shift (EM-Shift). They concurrently estimated the target position and the covariance matrix that describes the target shape by calculating the spatial moment of weight image of the target candidate. Yang et al. [27] introduced a new similarity measure that estimates the scale by comparing the second moments of the target model and the target candidate. A scale and orientation

adaptive mean shift tracking (SOAMST) algorithm is proposed by Ning et al. [20], in which the moment features are used to determine the scale and orientation changes of the target. The second moments are computed from an image of weights that are proportional to the probability that a pixel belongs to the target model. This method showed that the estimated area and the second order moment can adaptively estimate the changes of the target with high accuracy. In summary, all these methods rely much on the weight image to estimate the scale and orientation changes of the target, which is sensitive to illumination and cluttered backgrounds, thus, they fail when the object has similar color with its background.

The third category is feature mapping-based methods, where the local features between two consecutive frames contain the information about scale and orientation changes. Peng et al. [22] proposed an automatic bandwidth selection to exploit a feature mapping strategy to estimate the scale variation based on backward tracking and object centroid registration. A scale invariant feature transform (SIFT) based Mean Shift algorithm for object tracking is presented in [31] to obtain a reliable estimation of the location and target scale. However, this method brings in the local key points detection and feature matching, which affects the tracking efficiency and it is not well suitable for non-rigid target tracking. Similarly, [14, 30] rely on support features for scale estimation after the Mean Shift algorithm solves for position. Liang et al. [14] search for the target boundary by correlating the image with four templates. The boundaries positions directly determine the target scale. Zhao et al. [30] exploit affine structure to recover the target relative scale from feature point correspondences between consecutive frames. In summary, the methods depending on feature matching are able to robustly estimate the scale, but they can't be seamlessly integrated to the Mean Shift framework. Moreover, estimating scale from feature correspondences takes times, which requires presence of well-localized features that can be detected with high repeatability, and it has difficulties dealing with a non-rigid or a deformable object.

Other attempts were made to study different representation methods. Yu et al. [29] proposed a three-dimensional Mean Shift tracking algorithm, which combines the multi-scale model and background weighted spatial histogram to design a novel

Mean Shift algorithm which directly estimates the target position and scale in three-dimensional image space. Hu et al. [9] proposed an enhanced Mean Shift tracking algorithm using joint spatial-color feature and a similarity function in order to estimate the scale and orientation of the target. Recently, Vojir et al. [26] address the problem of scale adaptation and present a novel theoretically justified scale estimation mechanism which relies solely on the Mean Shift procedure for the Hellinger distance. Finally, all these methods presented above didn't consider the influence of large background variation, thus, they are not discriminate enough when the object and its background have similar colors.

### 3 Traditional Mean Shift tracking algorithm

Mean Shift tracking technique is to find the position of the target model in the current image from its color distribution which is usually approximated by a normalized histogram [4]. The color object to be tracked is supposed to have a density function  $q$ , and that of the target candidate centered at a point  $y$  the density  $p(y)$ . The problem is to find the point  $y$  whose density associated  $p(y)$  is closest to  $q$ . In [4], the appearance of the target is represented by an  $m$ -bin RGB weighted histogram computed from an ellipse region containing the object to track. Let  $\{z_i^*\}_{i=1\dots n}$  be the normalized pixel positions in the target region, which is supposed to be centered at the origin point. The target model  $\hat{q}$  and the probability of the feature  $s = 1\dots m$  in this model  $\hat{q}_s$  are computed as:

$$\begin{cases} \hat{q} = \{\hat{q}_s\}_{s=1\dots m} \\ \hat{q}_s = F \sum_{i=1}^n k\left(\|z_i^*\|^2\right) \delta[b(z_i^*) - s] \end{cases} \quad (1)$$

Where  $m$  is the size of the histogram (i.e. the number of classes used),  $\delta$  is the Kronecker function [4, 15], where the sum of delta functions for  $s = 1\dots m$  is 1. The function  $b: \mathbb{R}^2 \rightarrow \{1\dots m\}$  associates each pixel in the position  $z_i^*$  the index  $b(z_i^*)$  of its bin in  $m$ -histogram.  $k(z)$ <sup>1</sup> is an isotropic kernel function, monotonous and decreasing, assigning a lower weight to distant coordinates from center of the target model.  $z_i$  is the distance between the pixel  $z_i$  and the window center. The normalization constant  $F$  is obtained by imposing  $\sum_{s=1}^m \hat{q}_s = 1$  where [4]:

$$F = \frac{1}{\sum_{i=1}^n k\left(\|z_i^*\|^2\right)} \quad (2)$$

Similarly, let  $\{z_i\}_{i=1\dots n_h}$  be the normalized pixel locations of the target candidate centered at  $y$  in the current frame. Using the same kernel profil  $k(z)$ , but with bandwidth  $h$ , the target candidate model  $\hat{p}(y)$  and the probability of the feature  $s = 1\dots m$  in this model  $\hat{p}_s(y)$  are given by:

$$\begin{cases} \hat{p}(y) = \{\hat{p}_s(y)\}_{s=1\dots m} \\ \hat{p}_s(y) = F_h \sum_{i=1}^{n_h} k\left(\left\|\frac{y - z_i}{h}\right\|^2\right) \delta[b(z_i) - s] \end{cases} \quad (3)$$

Where  $F_h$  is the normalization constant defined by:

$$F_h = \frac{1}{\sum_{i=1}^{n_h} k\left(\left\|\frac{y - z_i}{h}\right\|^2\right)} \quad (4)$$

Where  $h$  is the scale of the target candidate (i.e. the number of pixels considered in the localization process). The Bhattacharyya coefficient [10] is used as a measure of similarity between two distributions  $\hat{q}_s$  and  $\hat{p}_s(y)$  as follows:

$$\rho[\hat{p}(y), \hat{q}] = \sum_{s=1}^m \sqrt{\hat{p}_s(y) \hat{q}_s} \quad (5)$$

The Bhattacharyya distance [10] is then defined by:

$$d[\hat{p}(y), \hat{q}] = \sqrt{1 - \rho[\hat{p}(y), \hat{q}]} \quad (6)$$

To find the location corresponding to the target in the current image, the distance (6) should be minimized as a function of  $y$ , which is equivalent to maximize the Bhattacharyya coefficient (5). This maximization can be performed efficiently using the Mean Shift iterations. The search for the new position of the target in the current frame starts at the estimated target location  $\hat{y}_0$  in the previous frame. Thus, the probabilities  $\{\hat{p}_s(\hat{y}_0)\}_{s=1\dots m}$  of the target candidates at the position  $\hat{y}_0$  in the current image should be calculated first. Using Taylor expansion around  $\hat{p}_s(y_0)$ , the linear approximation of the Bhattacharyya coefficient (5) is obtained as:<sup>1</sup>

<sup>1</sup> The profile of a kernel  $K$  is defined as a function  $k: [0 \ \infty[ \rightarrow \mathbb{R}$  such that  $K(z) = k(\|z\|^2)$  [4].

$$\rho[\hat{p}(y), \hat{q}] \approx \frac{1}{2} \sum_{s=1}^m \sqrt{\hat{p}_s(y_0) \hat{q}_s} + \frac{1}{2} F_h \sum_{i=1}^{n_h} \omega_i k \left( \left\| \frac{y - z_i}{h} \right\|^2 \right) \tag{7}$$

Where

$$\omega_i = \sum_{s=1}^m \sqrt{\frac{\hat{q}_s}{\hat{p}_s(\hat{y}_0)}} \delta[b(z_i) - s] \tag{8}$$

Since the first term in (7) is independent of  $y$ , to minimize the distance in (6) is equivalent to maximize the second term in (7). In this procedure, the estimated target moves from the current location  $\hat{y}_0$  to the new location  $\hat{y}_1$  according to the equation [4]:

$$\hat{y}_1 = \frac{\sum_{i=1}^{n_h} z_i \omega_i g \left( \left\| \frac{\hat{y}_0 - z_i}{h} \right\|^2 \right)}{\sum_{i=1}^{n_h} \omega_i g \left( \left\| \frac{\hat{y}_0 - z_i}{h} \right\|^2 \right)} \tag{9}$$

When we choose kernel  $g$  with the Epanechnikov profile [24], Eq. (9) is reduced to:

$$\hat{y}_1 = \frac{\sum_{i=1}^{n_h} z_i \omega_i}{\sum_{i=1}^{n_h} \omega_i} \tag{10}$$

Where  $g(z) = -k'(z)$  [4]. By using Eq. (10), the Mean Shift algorithm finds in the new frame the most similar region to the object with limited iterations and the new location  $\hat{y}_1$  is the center of the target region in the current frame. From Eq. (10) it can be observed that the key parameters in the Mean Shift algorithm are the weights  $\omega_i$ .

#### 4 Mean Shift tracking for scale and orientation changes of the target

The enlarging or shrinking of the target is usually a gradual process in consecutive frames. In this section, we analyze how to calculate adaptively the scale and orientation changes of the target under the Mean Shift framework.

##### 4.1 Estimating the target area

Based on the analysis presented in [2, 20], we can consider the weight image in the Mean Shift

algorithm as a density distribution function of the target, where the weight value of a pixel reflects the possibility that it belongs to the target. Thus, the scale and orientation of the target can be well estimated by using this density distribution function together with the moment features of the weight image. By definition, the zeroth image moment can be considered as the weighted area of the target in the target candidate region as follows:

$$M_{00} = \sum_{i=1}^n \omega(z_i) \tag{11}$$

To continuously track the object, some adjustment parameters are needed to adapt the tracking to scale change. However, to keep the target in the searching region we increase the size of the target candidate by increasing the width and length of the searching area based on the values from previous frames. On the other hand, using the Eq. (11), the Bhattacharyya coefficient  $\rho$  can be used to adjust  $M_{00}$  in estimating the target area, which is denoted by  $B$ . We use the following equation to estimate it:

$$B = e(\rho) M_{00} \tag{12}$$

Where  $e(\rho)$  is a monotonically increasing Gaussian function with respect to the Bhattacharyya coefficient  $\rho (0 \leq \rho \leq 1)$ . The authors in [20] have been demonstrated that  $M_{00}$  always greater than the real target area and it will monotonically approach to the real target area with  $\rho$  increasing.

Thus, we require that  $e(\rho)$  should be monotonically increase and reach maximum 1 when  $\rho$  is 1. The Gaussian function  $e(\rho)$  is defined as follows:

$$e(\rho) = \exp\left(\frac{\rho - 1}{\sigma}\right) \tag{13}$$

From Eqs. (12) and (13) we can see that when the target model approaches the target candidate,  $e(\rho)$  gets close to 1 and in this case it is more reliable to use  $M_{00}$  as the estimation of the target area.

When  $\rho$  decreases, the similarity between target candidate and target model diminishes,

$M_{00}$  will be much bigger than the target area but  $e(\rho)$  is less than 1 so that  $B$  can avoid being biased too much from the real target area. Note that when  $\rho$  approaches to 0, the tracked target gets lost and  $e(\rho)$  will be very small, so that  $B$  is close to 0. However, by

experimentation the parameter  $\sigma$  is taken between 1 and 2 [20].

#### 4.2 Estimating the width, height and orientation of the target

The moment features describe numeric quantities at some distance from a reference point or axis [23], they are commonly used in image analysis and pattern recognition tasks [23]. This is due essentially to their simplicity, the invariance and geometric meaning of the low order moment values. Similar to the CAMSHIFT algorithm [2], the first order moments of the weight image  $\{M_{10}, M_{01}\}$ , are used to locate the target center as follows:

$$M_{10} = \sum_{i=1}^{n_h} w_i z_{i,1}, M_{01} = \sum_{i=1}^{n_h} w_i z_{i,2} \tag{14}$$

Equally, the second order moments  $\{M_{02}, M_{11}, M_{20}\}$  [2] are calculated as follows:

$$M_{20} = \sum_{i=1}^{n_h} w_i z_{i,1}^2 \quad M_{02} = \sum_{i=1}^{n_h} w_i z_{i,2}^2 \tag{15}$$

$$M_{11} = \sum_{i=1}^{n_h} w_i z_{i,1} z_{i,2}$$

Where pair  $(z_{i,1}, z_{i,2})$  are the coordinates of pixel  $i$  in the candidate region. Comparing Eq. (10) with Eqs. (11) and (14), the new position  $\hat{y}_1$  of the target found in Eq. (9) is actually the ratio of the first order moment to the zeroth order moment:

$$y_1 = (\bar{z}_1, \bar{z}_2) = \left( \frac{M_{10}}{M_{00}}, \frac{M_{01}}{M_{00}} \right) \tag{16}$$

Where  $(\bar{z}_1, \bar{z}_2)$  represents the centroid of the target candidate region. According to [2], the second order moment describes the shape and orientation (long axis) of an object. By using Eqs. (10), (11), (15) and

(16), we can convert Eq. (9) to the second order moment as follows:

$$\begin{aligned} \mu_{20} &= M_{20}/M_{00} - \bar{z}_1^2 & \mu_{02} &= M_{02}/M_{00} - \bar{z}_2^2 \\ \mu_{11} &= M_{11}/M_{00} - \bar{z}_1 \bar{z}_2 \end{aligned} \tag{17}$$

Equation (17) can be rewritten as the following covariance matrix:

$$Cov = \begin{bmatrix} \mu_{20} & \mu_{11} \\ \mu_{11} & \mu_{02} \end{bmatrix} \tag{18}$$

This covariance matrix can be decomposed by using the singular value decomposition (SVD) [7] as follows:

$$\begin{aligned} Cov &= U \times S \times U^T \\ &= \begin{bmatrix} u_{11} & u_{12} \\ u_{21} & u_{22} \end{bmatrix} \times \begin{bmatrix} \lambda_1^2 & 0 \\ 0 & \lambda_2^2 \end{bmatrix} \times \begin{bmatrix} u_{11} & u_{12} \\ u_{21} & u_{22} \end{bmatrix}^T \end{aligned} \tag{19}$$

Where  $U = \begin{bmatrix} u_{11} & u_{12} \\ u_{21} & u_{22} \end{bmatrix}$  and  $S = \begin{bmatrix} \lambda_1^2 & 0 \\ 0 & \lambda_2^2 \end{bmatrix}$   $\lambda_1^2$  and  $\lambda_2^2$  are the eigenvalues of the covariance matrix [7]. The vectors  $(u_{11}, u_{21})^T$  and  $(u_{12}, u_{22})^T$  represent the orientation of the two main axes of the target. In the CAMSHIFT algorithm [2],  $\lambda_1$  and  $\lambda_1$  are directly used as the width and height of the target [18]. However, in [20] the authors proposed a new technique to estimate more accurately the width and height of the target. The target to be tracked is represented by an ellipse, for which the lengths of the semi-major axis and semi-minor axis are denoted by  $a$  and  $b$ , respectively. It has been shown in [18] that the ratio of  $\lambda_1$  and  $\lambda_2$  can well approximate the ratio of  $a$  to  $b$ , i.e.,  $\lambda_1/\lambda_2 \approx a/b$ . Therefore, the terms  $a$  and  $b$  can be defined as follows  $a = k\lambda_1$  and  $b = k\lambda_2$ , where  $k$  is a scale factor. Using the geometrical moments of the ellipse [18] and the target region  $B$  estimated by the Eq. (12), it can be written as  $\pi ab = \pi(k\lambda_1)(k\lambda_2) = B$ . Then, we derive easily that:

$$k = \sqrt{B/(\pi\lambda_1\lambda_2)} \tag{20}$$

$$a = \sqrt{\lambda_1 B / (\pi \lambda_2)} \quad b = \sqrt{\lambda_2 B / (\pi \lambda_1)} \quad (21)$$

Now the covariance matrix becomes:

$$Cov = \begin{bmatrix} u_{11} & u_{12} \\ u_{21} & u_{22} \end{bmatrix} \times \begin{bmatrix} a^2 & 0 \\ 0 & b^2 \end{bmatrix} \times \begin{bmatrix} u_{11} & u_{12} \\ u_{21} & u_{22} \end{bmatrix}^T \quad (22)$$

Where  $a^2$  and  $b^2$  are the eigenvalues of the covariance matrix  $Cov$  in Eq. (22), which represent the height and width of the ellipsoid region tracked. Once the position, scale and orientation of the target are estimated in the current frame, the location of the target candidate region needs to be determined in the next frame. The covariance matrix representing the size of the target candidate region in the next frame is defined as follows [20]:

$$Cov_2 = U \times \begin{bmatrix} (a + \Delta d)^2 & 0 \\ 0 & (b + \Delta d)^2 \end{bmatrix} \times U^T \quad (23)$$

Where  $\Delta d$  is the increment of the target candidate region in the next frame. The position of the initial target candidate region is defined by the following ellipse region:

$$(x - y_1) \times Cov_2^{-1} \times (x - y_1)^T \leq 1 \quad (24)$$

Where  $x$  are the points of the ellipse used in the next research region and  $y_1$  the position of the target in the next frame.

### 5 Proposed method

The background information plays a very important role for the reason that it is often presented in the selected target region. Since, there are two main reasons to use it. First, if some of the target's features are also present in the background, their relevance for the target localization is reduced. Second, in many applications it is difficult to exactly delineate the target without any background information, thus, for a better

target representation, the background model has been used to improve the target localization.

#### 5.1 Background-Weighted Histogram

A representation model of the background named Background-Weighted Histogram (BWH) has been proposed by Comaniciu et al. [4]. The original background model  $\hat{o}$  is modeled by the discrete representation (normalized histogram)  $\{\hat{o}_s\}_{s=1\dots m}$  (with  $\sum_{s=1}^m \hat{o}_s = 1$ ) in the feature space as follows:

$$\begin{cases} \hat{o} = \{\hat{o}_s\}_{s=1\dots m} \\ \hat{o}_s = F_1 \sum_{i=1}^n k(\|z_i^*\|^2) \delta[b(z_i^*) - s] \end{cases} \quad (25)$$

Where  $\hat{o}_s$  represents the probability of colors  $s = 1\dots m$  in this model and  $F_1$  is the normalization constant. Then, the weights coefficients were defined using the following formula [4]:

$$\left\{ \vartheta_s = \min\left(\frac{\hat{o}_s^*}{\hat{o}_s}, 1\right) \right\}_{s=1\dots m} \quad (26)$$

With  $\hat{o}_s^*$  is the minimal non-zero value of the representation. These weights are only used to define a transformation for the representations of the target model and candidates. The transformation diminishes the importance of those features which have low  $\vartheta_s$ , i.e., are more prominent in the background and less important for target representation. This representation is computed in an elliptical region around the target with a fixed three times of the target area [4]. Then, the transformed target model is obtained by the following formula:

$$\begin{cases} \hat{q}' = \{\hat{q}'_s\}_{s=1\dots m} \\ \hat{q}'_s = F' \vartheta_s \sum_{i=1}^n k(\|z_i^*\|^2) \delta[b(z_i^*) - s] \end{cases} \quad (27)$$

Where  $\hat{q}'_s$  represents the probability of colors  $s = 1\dots m$  in this model, and  $F'$  is the normalization constant expressed as:

$$F' = \frac{1}{\sum_{i=1}^n k\left(\|z_i^*\|^2\right) \sum_{s=1}^m \vartheta_s \delta[b(z_i^*) - s]} \quad (28)$$

Similarly, the transformed target candidate model is:

$$\begin{cases} \hat{p}'(y) = \{\hat{p}'_s(y)\}_{s=1 \dots m} \\ \hat{p}'_s(y) = F'_h \vartheta_s \sum_{i=1}^{n_h} k\left(\left\|\frac{y - z_i}{h}\right\|^2\right) \delta[b(z_i) - s] \end{cases} \quad (29)$$

Where now  $F'_h$  is given by:

$$F'_h = \frac{1}{\sum_{i=1}^n k\left(\left\|\frac{y - z_i}{h}\right\|^2\right) \sum_{s=1}^m \vartheta_s \delta[b(z_i) - s]} \quad (30)$$

The new weight formula  $\omega'_i$  calculated by the BWH algorithm is:

$$\omega'_i = \sum_{s=1}^m \sqrt{\frac{\hat{q}'_s}{\hat{p}'_s(y)}} \delta[b(z_i) - s] \quad (31)$$

### 5.2 Corrected Background-Weighted Histogram

A Corrected Background-Weighted Histogram (CBWH) algorithm is proposed by Ning et al. [19]. Rather than both transforming the target model and the target candidate, it just transforms the target model. The authors proved that the weights assigned to the pixels in the target candidate region by BWH are proportional to those without information from the background. Since, CBWH reduce the prominent background features only in the target model, then, the target candidate model still uses the original model as follows:

$$\begin{cases} \hat{p}(y) = \{\hat{p}_s(y)\}_{s=1 \dots m} \\ \hat{p}_s(y) = F_2 \sum_{i=1}^{n_h} k\left(\left\|\frac{y - z_i}{h}\right\|^2\right) \delta[b(z_i) - s] \end{cases} \quad (32)$$

Where  $\hat{p}_s(y)$  represents the probability of colors  $s = 1 \dots m$  in this model and  $F_2$  is the normalization constant defined by:

$$F_2 = \frac{1}{\sum_{i=1}^{n_h} k\left(\left\|\frac{y - z_i}{h}\right\|^2\right) \sum_{s=1}^m \delta[b(z_i) - s]} \quad (33)$$

The new weight formula  $\omega''_i$  computed by the CBWH in the target candidate region, is given by:

$$\omega''_i = \sum_{s=1}^m \sqrt{\frac{q'_s}{\hat{p}_s(y)}} \delta[b(z_i) - s] \quad (34)$$

Motivated by the significant advances made by the CBWH algorithm [19], we adopted this technique in this work in order to exploit the key features of the background.

### 5.3 Proposed tracking algorithm

The appearance model of the object is a key element affecting performance for many objects tracking system. Estimation techniques based on the kernel densities have been proposed in [4, 6] for constructing statistical representations of the object appearance. Their advantage is to provide a relatively flexible and generic description of the appearance and allow a rapid implementation and real-time in some conditions. This work presents a new appearance model for human tracking based on Mean Shift algorithm [4]. It aims to solve the main tracking challenges related to the scale and orientation changes of the target in the presence of distractor elements where the only knowledge available on the target is its position in the first frame of the images sequence. Two areas are well developed. Firstly, the moment features of the weight image [2] are used to estimate the position, scale and orientation changes of the target. In this context, we support that the weight image derived from the target model and the target candidate represents the probability that a pixel belongs to the target. Thus, the target scale is calculated using the zeroth order moment and the Bhattacharyya coefficient [10], then the width, height and orientation were estimated by the second order moments and the estimated area. Second, the Background-Weighted Histogram (BWH) [19] has been incorporated into the target representation in order to derive a simple representation of the background and to reduce the interference of the main features of the background in the target localization. Finally, these two concepts are combined to generate a robust tracking algorithm entitled Scale and Orientation-based Background Weighted Histogram (SOBWH).

The entire of the proposed algorithm for human tracking SOBWH is summarized as follows.



---

**Algorithm:** Scale and Orientation-based Background Weighted Histogram for Human Tracking
 

---

There is the estimation distribution of the target model  $\{\hat{q}_s\}_{s=1..m}$  according to (1) and the estimation of its initial position  $\hat{y}_0$  in the previous frame. You must then:

1. Calculate the Background-Weighted Histogram  $\{\hat{o}_s\}_{s=1..m}$  and  $\{\mathbf{v}_s\}_{s=1..m}$  by equations (25) and (26) respectively, then calculate the transformed target model  $\{\hat{q}'_s\}_{s=1..m}$  according to (27).
  2. Initialize the position of the target candidate  $\hat{\mathbf{p}}(\mathbf{y}_0)$  in the current frame with  $\hat{y}_0$ , and set the iteration number  $\mathbf{k} = \mathbf{0}$ , then, calculate the target candidate distribution  $\hat{\mathbf{p}}(\mathbf{y}_0)$  according to (32) in the current frame and evaluate the Bhattacharyya coefficient  $\rho[\hat{\mathbf{p}}(\hat{\mathbf{y}}_0), \hat{q}'] = \sum_{s=1}^m \sqrt{\hat{\mathbf{p}}_s(\hat{\mathbf{y}}_0) \hat{q}'_s}$ .
  3. Calculate the new weights formula  $\{\omega_i''\}_{i=1..n_h}$  according to (34).
  4. Find the new location  $\hat{y}_1$  of the target candidate according to (10).
  5. Compute  $\hat{\mathbf{p}}_s(\mathbf{y}_1)_{s=1..m}$ , and evaluate  $\rho[\hat{\mathbf{p}}(\hat{\mathbf{y}}_1), \hat{q}'] = \sum_{s=1}^m \sqrt{\hat{\mathbf{p}}_s(\hat{\mathbf{y}}_1) \hat{q}'_s}$ .
  6. While  $\rho[\hat{\mathbf{p}}(\hat{\mathbf{y}}_1), \hat{q}'] < \rho[\hat{\mathbf{p}}(\hat{\mathbf{y}}_0), \hat{q}']$   
 Do  $\hat{y}_1 \leftarrow \frac{1}{2}(\hat{\mathbf{y}}_0 + \hat{\mathbf{y}}_1)$   
 Evaluate  $\rho[\hat{\mathbf{p}}(\hat{\mathbf{y}}_1), \hat{q}']$
  7. Let  $\mathbf{d} \leftarrow \|\hat{\mathbf{y}}_1 - \hat{\mathbf{y}}_0\|$   
 If  $\mathbf{d} < \varepsilon$  or  $\mathbf{k} \geq N_{max}$   
 Stop iteration and go to step 10.  
 Otherwise  
 $\hat{\mathbf{y}}_0 \leftarrow \hat{\mathbf{y}}_1$ ,  $\mathbf{k} \leftarrow \mathbf{k} + 1$  and go to step 3.
  8. Calculate the weight images  $\omega_i''$  for target scale change according to (34), then estimate target area  $\mathbf{B}$  for tracking using equation (12).
  9. Analyze the moment features in Mean Shift algorithm according to (14) and (15).
  10. Estimate the width, height and orientation from the target candidate model using equation (22).
  11. Estimate the initial target candidate model for next frame using equation (24).
  12. Set the next frame as the current frame with initial location  $\hat{\mathbf{y}}_0$  and go to step 1.
- 

The stopping criterion threshold  $\varepsilon$ , used in step 7 is derived by constraining the vectors  $\hat{y}_0$  and  $\hat{y}_1$  to be within the same pixel in original image coordinates. The default value is  $\varepsilon = 0.1$  [4]. From real time constraints, we also limit the number of Mean Shift iterations to  $N_{max} = 20$  [4]. At each iteration, the vector Mean Shift is calculated as the similarity measure between the histograms of the target model and target candidates is increased. This process is repeated until the convergence is achieved, in practice, the average number of iterations is much smaller, about 4.

The implementation of the proposed algorithm for human tracking SOBWH is much simpler than as presented above. The role of step 6 is only to avoid potential numerical problems in the Mean Shift maximization. In practice, we only iterate by computing the weights in step 3, then the new position is calculated in step 4, and testing the size of the kernel shift in step 7. The Bhattacharyya coefficient is computed only after the algorithm completion to evaluate the similarity between the target model and the chosen candidate. An obvious advantage of the Mean Shift tracking system compared to the standard matching is the removal of extensive research. The estimation of the object state is thus carried out in a limited number of iterations. The complexity of the classic Mean Shift algorithm is very low because it requires very simple operations, where, the time complexity of it is given by  $O(Tn^2)$  where  $T$  is the number of iterations and  $n$  is the number of data points in the data set.

## 6 Experimental results

In this section, we perform tests confirming the robustness of the proposed method SOBWH for human tracking based on Mean Shift algorithm [4] that aims to solve the main tracking challenges related to the scale and orientation changes of the target in the presence of distractor elements. The experiments are done using a PC (Intel® Core™ 2 Duo CPU T5800 with 2.00 GHz CPU and 3 GB RAM). The tracking initialization is defined by manually marking a bounding box in the first frame. Once the object to be tracked is localized, our proposed approach SOBWH is used to iteratively find the best object matched region for every subsequent frame. For quantitative performance evaluation of all the methods presented, we calculate the localization errors based on the Object Centroid Position Error formula (OCPE) which is approximated by the distance between the center of the tracking results and that of the ground truth that appears in [25].

$$\text{OCPE} = \frac{1}{N_{rg}} \sum_{\exists ig(t_i) \Delta r(t_i)} \sqrt{(xg_i - xr_i)^2 + (yg_i - yr_i)^2} \quad (35)$$

Where  $N_{rg}$  is the total number of images for each sequence,  $(xg_i, yg_i)$  and  $(xr_i, yr_i)$  are the positions of the ground truth and the results obtained by different

methods considered at frame  $i$  respectively. Ideally, an optimal approach is expected to have a small error. The tracking results obtained by our proposed approach SOBWH are compared with the same existing tracking algorithms in the literature, namely the Adaptive Scale Mean Shift algorithm [4], the EM-Shift algorithm [32] and the SOAMST algorithm [20]. All these algorithms are based on the use of representation by appearance model for object tracking, and they aim to address the scale and orientation changes of the target under the Mean Shift framework. For all these methods, we selected RGB color space as a feature space and it was quantized into  $s = 16 \times 16 \times 16$  bins. Note that the shape of our tracking algorithm is approximated by an ellipse and its appearance by a normalized histogram based on features derived from a fusion procedure of background information and moment features of the weight image. In order to obtain a better view about how the different algorithms perform, we draw the tracking results separately in different images. We used four Benchmark sequences to validate the performance of the proposed approach SOBWH: Human9 (305 frames), Diving (231 frames), David (471 frames) and Walking Woman (597 frames). All the datasets and the ground truth are available in [8]. The mean

position error of the target localization computed by (35), the average iteration numbers and the average processing time for each approach are presented in Tables 1, 2 and 3 respectively.

In the first experiment, we have used the “Human9” sequence with 305 frames of  $320 \times 240$  pixels and which presents a person walking in a road. This sequence has many critical challenges such as illumination change, gradual diminution of the scale, deformation, motion blur and fast motion of the target. In these condition the final qualitative tracking results for all the algorithms considered are shown in Fig. 1. Due to the fast motion and drastic changes in scale and background all these algorithms Adaptive Scale, EM-Shift and SOAMST lose the target very quickly after the frame 72 (Fig. 1a–c), however, our proposed approach SOBWH is able to track the human successfully in the whole sequence Fig. 1d. For quantitative evaluation, Tables 1, 2 and 3 summarize the results obtained, thus, the proposed approach SOBWH achieves higher target localization accuracy than other algorithms Adaptive Scale, EM-Shift and SOAMST, it is also needs less number of iterations, which means that the proposed method converges faster and requires less computation time. The average error for target localization per image is presented in

**Table 1** Average errors for target localization in four sequences and for four different methods considered

Video sequence	Adaptive scale approach	EM-Shift approach	SOAMST approach	Our approach SOBWH
Human9	60.4832	74.7009	68.2669	<b>11.5550</b>
Diving	22.9877	52.3882	24.9141	<b>22.0695</b>
David	44.2561	65.5511	31.5872	<b>23.3873</b>
Walking woman	77.5726	19.0786	15.9103	<b>6.1252</b>
Mean	51.3249	52.9297	38.5881	<b>15.7842</b>

**Table 2** Average number of iterations in four sequences and for four different methods considered

Video sequence	Adaptive scale approach	EM-Shift approach	SOAMST approach	Our approach SOBWH
Human9	15	5	6.4196	<b>6.9016</b>
Diving	15	5	3.9434	<b>3.6739</b>
David	15	5	4.7770	<b>4.5669</b>
Walking woman	15	5	4.4103	<b>4.2224</b>
Mean	15	5	4.8875	<b>4.8412</b>

**Table 3** Average processing time in four sequences and for four different methods considered

Video sequence	Adaptive scale approach	EM-Shift approach	SOAMST approach	Our approach SOBWH
Human9	1.1944	0.8246	3.2619	<b>0.5388</b>
Diving	1.0652	0.7786	1.6906	<b>0.6340</b>
David	1.7390	0.8899	1.0595	<b>0.3638</b>
Walking woman	1.1572	0.8847	0.5771	<b>0.1992</b>
Mean	1.2889	0.8444	1.6472	<b>0.4339</b>

Fig. 2. The experimental results show that the proposed approach SOBWH reduces the background interference in target localization and helps to discriminate between the target and its background, which improves the tracking accuracy in the presence of disruptive elements and scale changes compared to other algorithms Adaptive Scale, EM-Shift and SOAMST.

The second experiment is “Diving” sequence with 231 frames of  $400 \times 224$  pixels. The difficulties of the sequence are: the ability to take into account the fast target motion as well as those of the camera, different viewpoints of the target, scale variation and in-plane rotation. Qualitative tracking results are shown in Fig. 3. Because of important deformation and target rotation, both Adaptive Scale and EM-Shift algorithms lose the target from frame 168 (Fig. 3a, b). As shown in (Fig. 3c, d) SOAMST algorithm and our approach SOBWH achieve to track the target more efficiency over the whole sequence, however, the proposed method SOBWH estimates more accurately the target scale change even if it moves at different speeds during the sequence. Refer to the statistics results presented in Tables 1, 2 and 3 the proposed approach SOBWH achieves higher target localization accuracy than other algorithms Adaptive Scale, EM-Shift and SOAMST. The average error for target localization per image is presented in Fig. 4. These experimental results confirm the robustness of the proposed method while having an appearance model invariant to rotation, scale variation and able to discriminate between the target and its background.

The third experiment is “David” sequence with 471 images of  $320 \times 240$  pixels. The sequence shows a person entering and moving into a room and the target to be tracked is his face. This sequence involves many critical challenges such as illumination and scale changes, occlusion, in and out plane rotation, motion

blur and deformation. Figure 5 presents final qualitative tracking results of all methods. Because of fast motion and drastic background change, neither adaptive scale algorithm nor EM-Shift algorithm achieves good tracking results (Fig. 5a, b). However, Fig. 5d indicates that the proposed method SOBWH successfully tracks the face over the whole sequence and estimates more accurately the scale change compared to SOMAST algorithm. From statistics results presented in Tables 1, 2 and 3 the proposed approach achieves higher target localization accuracy than the other referenced algorithms. The average error for target localization per image is presented in Fig. 6. Based on these experimental results, the proposed approach SOBWH is less sensitive to major illumination and background changes compared to other algorithms Adaptive Scale, EM-Shift and SOAMST.

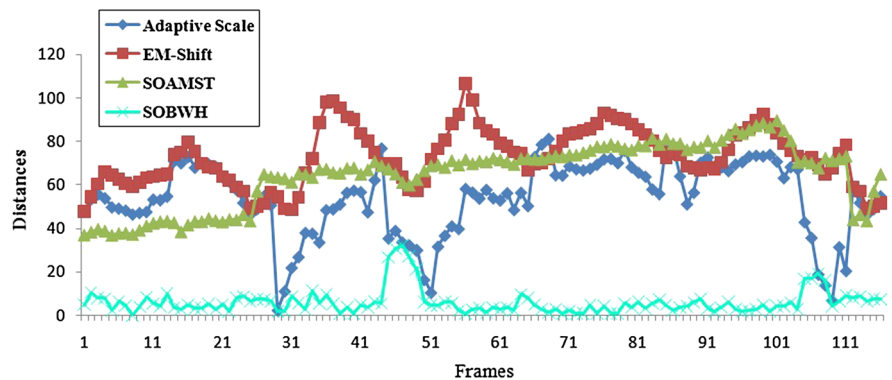
In the fourth experiment, we have used a long sequence with 597 frames “Walking Woman” of  $352 \times 288$  pixels. It is complex and challenging since the woman moves quickly and has obvious illumination variation, occlusion, scale and orientation changes, deformation, motion blur and target rotation outside the image plane. Final qualitative tracking results are shown in Fig. 7. Because the woman is part occluded by the first car whose color is similar to that of the woman’s shirt, both algorithms Adaptive Scale and EM-Shift derive the target from image 135 (Fig. 7a, b) and they will not take it after. But when the woman moves quickly such as in frames 369 and 587, the estimated target scale and orientation by SOMAST are not as accurate as those by our proposed method. Therefore, the proposed algorithm SOBWH able to correctly track the woman in the whole sequence Fig. 7d. Tables 1, 2 and 3 show that the proposed approach achieves higher target localization accuracy than the other algorithms Adaptive Scale, EM-Shift and SOAMST. The average error for target

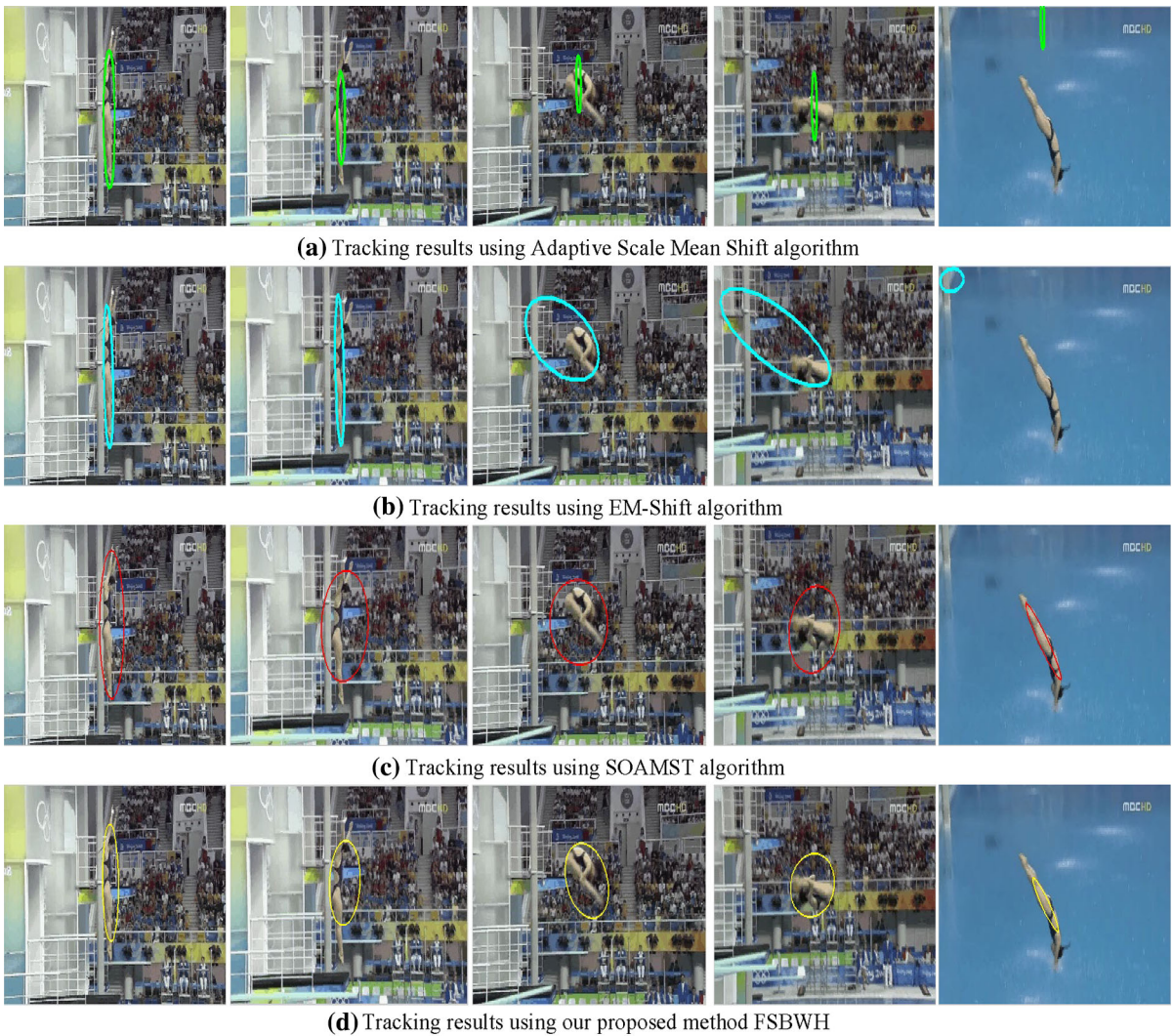


**Fig. 1** Tracking results from “Human9” sequence for frames 6, 72, 108, 154 and 287 are displayed. **a** Tracking results using Adaptive Scale Mean Shift algorithm, **b** tracking results using

EM-Shift algorithm, **c** tracking results using SOAMST algorithm, **d** tracking results using our proposed method SOBWH

**Fig. 2** Average error for target localization per image by “Human9” sequence

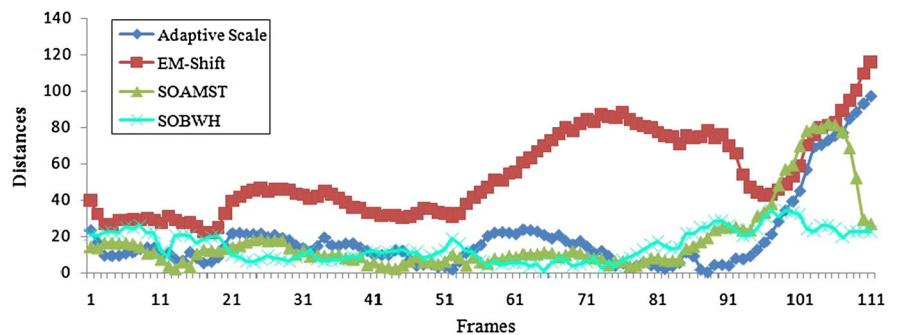




**Fig. 3** Tracking results from “Diving” sequence for frames 2, 26, 120, 168 and 214 are displayed. **a** Tracking results using Adaptive Scale Mean Shift algorithm, **b** tracking results using

EM-Shift algorithm, **c** tracking results using SOAMST algorithm, **d** tracking results using our proposed method SOBWH

**Fig. 4** Average error for target localization per image by “Diving” sequence

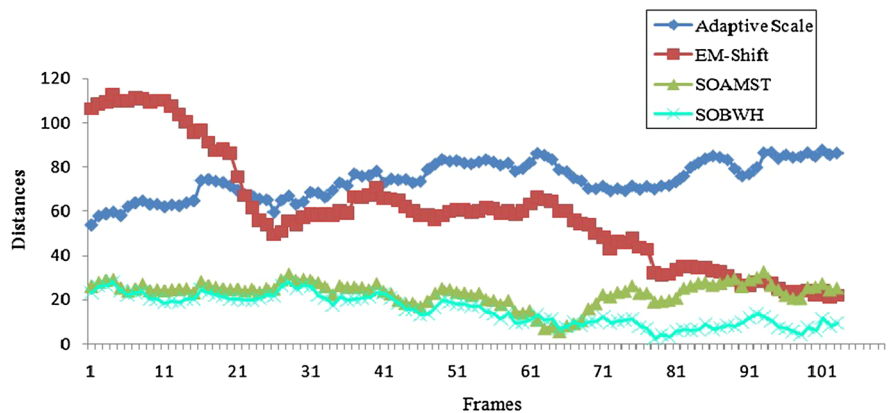


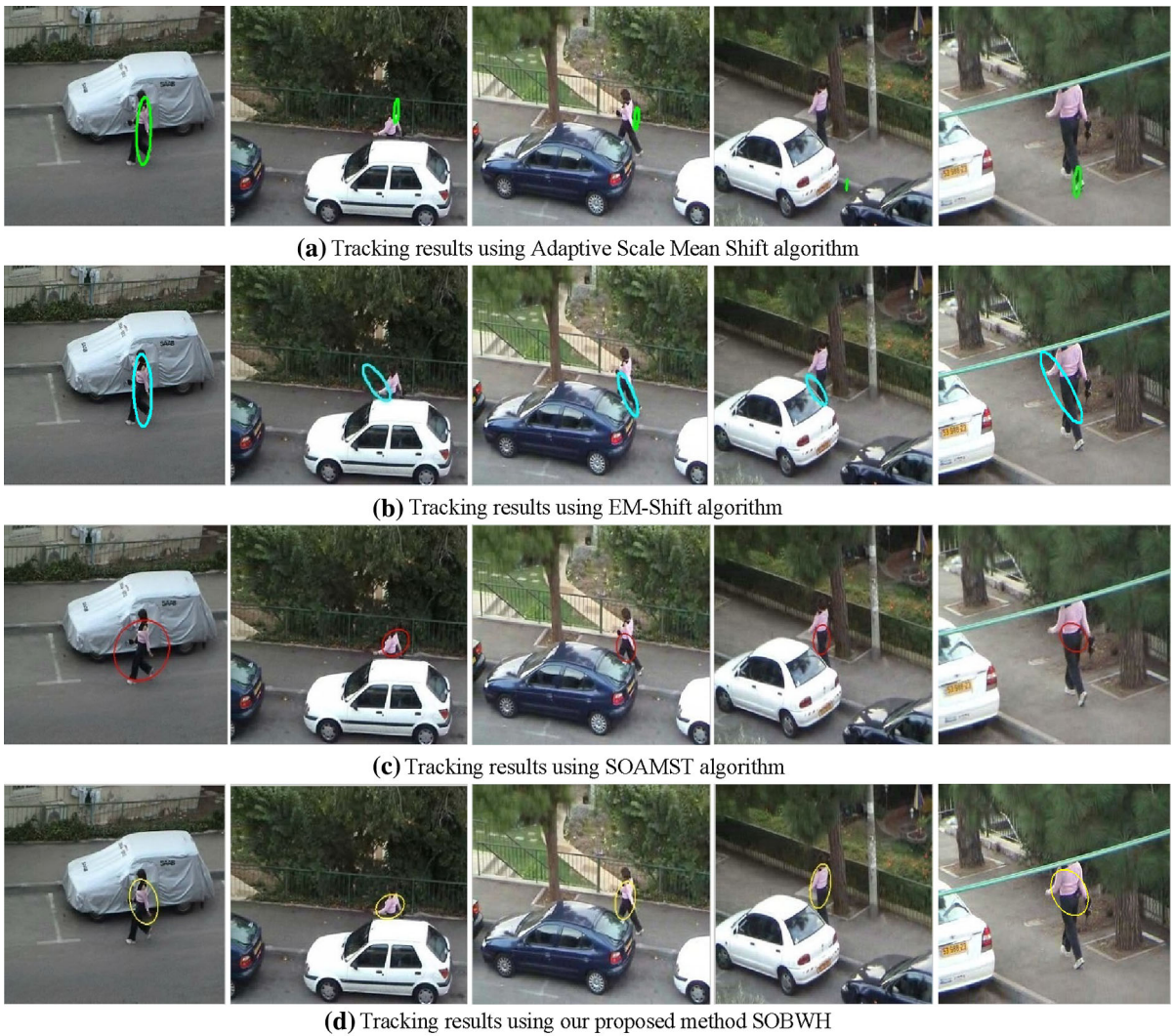


**Fig. 5** Tracking results from “David” sequence for frames 2, 9, 296, 439 and 471 are displayed. **a** Tracking results using Adaptive Scale Mean Shift algorithm, **b** tracking results using

EM-Shift algorithm, **c** tracking results using SOAMST algorithm, **d** tracking results using our proposed method SOBWH

**Fig. 6** Average error for target localization per image by “David” sequence

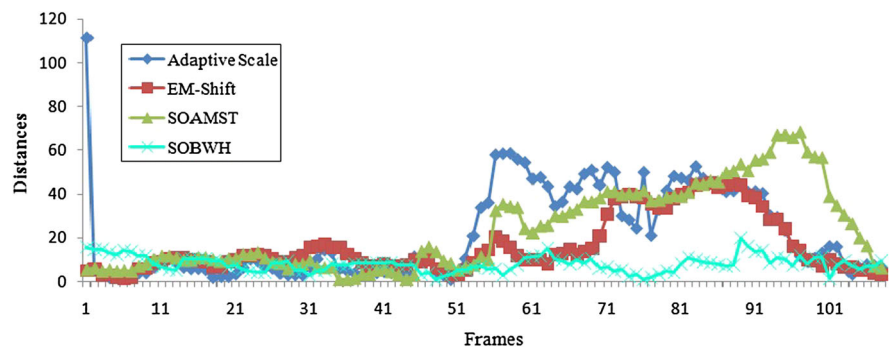




**Fig. 7** Tracking results from “Walking Woman” sequence for frames 8, 135, 194, 369 and 587 are displayed. **a** Tracking results using Adaptive Scale Mean Shift algorithm, **b** tracking

results using EM-Shift algorithm, **c** tracking results using SOAMST algorithm, **d** tracking results using our proposed method SOBWH

**Fig. 8** Average error for target localization per image by “Walking Woman” sequence



localization per image is presented in Fig. 8. Based on these experimental results, the proposed approach SOBWH allows to highlight the background features in target area, while having an appearance model invariant to partial occlusion, scale change and able to discriminate between the target and its background.

In summary, by calculating the average error for each method (Table 1), we arrived that the proposed approach SOBWH provides minimal localization accuracy, followed by SOAMST algorithm, Adaptive Scale algorithm and EM-Shift algorithm respectively. It also requires less number of iterations (Table 2) which means that the proposed method converges faster and requires less computation time (Table 3). However, our qualitative and quantitative experiments show that the use of background information [19] with the moment features of the weight image [2] can effectively track the position, scale and orientation changes of the target despite the disruptive factors of the scene, as well as, introduces new information in tracking objects such as scale and orientation of the target.

## 7 Conclusion

In this paper, we proposed a human tracking algorithm based on Mean Shift technique. The novelty of this work is to combine moment features of the weight image and background weighted histogram to design a robust tracking algorithm Scale and Orientation-based Background Weighted Histogram SOBWH in order to estimate effectively the position, scale and orientation changes of the target in the presence of distractor elements and which is used to exploit the main background features, so that an object can be effectively discriminated against the background of the image. The qualitative results and quantitative analysis validate the effectiveness and robustness of the proposed approach SOBWH for human tracking, so that it presents considerable advantages compared to other algorithms in terms of accuracy and computation time. Finally, comparative evaluations confirm the pertinence of the developed ideas in this paper, by demonstrating that we have exceeded a variety of recent algorithms in the literature, namely SOAMST, Adaptive Scale and EM-Shift in various difficult tracking scenarios such as scale and orientation changes, partial occlusion, motion and zoom camera,

complex background change, deformation, rotation and illumination variation. These advantages allow us to work freely in the tracking domain with unconstrained environments.

## References

1. Boussetouane, F., Dib, L., & Snoussi, H. (2013). Improved mean shift integrating texture and color features for robust real time object tracking. *The Visual Computer*, 29(3), 155–170.
2. Bradski, G. (1998). Computer vision face tracking for use in a perceptual user interface. *Intel Technology Journal*, 2(2), 1–15.
3. Collins, R. T. (2003). Mean-shift blob tracking through scale space. In *Proceedings of IEEE conference on computer vision and pattern recognition*, vol. 2 (pp. 234–240).
4. Comaniciu, D., Ramesh, V., & Meer, P. (2003). Kernel-based object tracking. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(5), 564–577.
5. Fukunaga, K., & Hostetler, L. D. (1975). The estimation of the gradient of a density function, with applications in pattern recognition. *IEEE Transactions on Information Theory*, 21(1), 32–40.
6. Georgescu, B., & Meer, P. (2004). Point matching under large image deformations and illumination changes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26, 674–688.
7. Horn, R. A., & Johnson, C. R. (1991). *Topics in matrix analysis*. Cambridge: Cambridge University Press.
8. <http://www.visual-tracking.net/>.
9. Hu, J. S., Juan, C. W., & Wang, J. J. (2008). A spatial-color mean shift object tracking algorithm with scale and orientation estimation. *Pattern Recognition Letters*, 29(16), 2165–2173.
10. Kailath, T. (1967). The divergence and Bhattacharyya distance measures in signal selection. *IEEE Transactions on Communication Technology*, 15, 52–60.
11. Laaroussi, K., Saaidi, A., Masrar, M., Satori, K. (2016). Human tracking based on appearance model. In *Proceedings of the mediterranean conference on information & communication technologies*, vol. 1 (pp. 297–305). Springer International Publishing.
12. Laaroussi, K., Saaidi, A., Masrar, M., & Satori, K. (2014). Video-surveillance system for tracking moving people using color interest points. *World Applied Sciences Journal*, 32(2), 289–301.
13. Laaroussi, K., Saaidi, A., & Satori, K. (2014). People tracking using color control points and skin color. *Journal of emerging technologies in web intelligence*, 6(1), 94–100.
14. Liang, D., Huang, Q., Jiang, S., Yao, H., Gao, W. (2007). Mean shift blob tracking with adaptive feature selection and scale adaptation. In *Proceedings international conference image processing*, vol. 3 (pp. III-369).
15. Liang, Z. P., & Lauterbur, P. C. (2000). *Principles of magnetic resonance imaging*. Bellingham: SPIE Optical Engineering Press.



16. Li, S., Chang, H., & Zhu, C. (2010). Adaptive pyramid mean shift for global real time visual tracking. *Image and Vision Computing*, 28(3), 424–437.
17. Li, X., Hu, W., Shen, C., Zhang, Z., Dick, A., & Hengel, A. V. D. (2013). A survey of appearance models in visual object tracking. *ACM Transactions on Intelligent Systems and Technology*, 4(4), 58.
18. Mukundan, R., & Ramakrishnan, K. R. (1998). *Moment functions in image analysis: theory and applications, vol. 100*. Singapore: World Scientific.
19. Ning, J., Zhang, L., Zhang, D., & Wu, C. (2010). Robust mean shift tracking with corrected background-weighted histogram. *IET Computer Vision*, 6(1), 62–69.
20. Ning, J., Zhang, L., Zhang, D., & Wu, C. (2012). Scale and orientation adaptive mean shift tracking. *IET Computer Vision*, 6(1), 52–61.
21. Parameswaran, V., Ramesh, V., Zoghiani, I. (2006). Tunable kernels for tracking. In *Proceedings IEEE computer society conference on computer vision and pattern recognition* (pp. 2179–2186).
22. Peng, N. S., Yang, J., Liu, Z., & Zhang, F. C. (2005). Automatic selection of kernel-bandwidth for mean shift object tracking. *Journal of Software*, 16(9), 1542–1550.
23. Prokop, R. J., & Reeves, A. P. (1992). A survey of moment based techniques for unoccluded object representation and recognition. *Graphical Models and Image Processing*, 54(5), 438–460.
24. Qingchang, G., Xiaojuan, C., Hongxia, C. (2007). Mean shift of variable window based on the Epanechnikov Kernel. In *Proceeding international conference on mechatronics and automation* (pp. 2314–2319).
25. Senior, A., Hampapur, A., Tian, Y., Brown, L., Pankanti, S., & Bolle, R. (2006). Appearance models for occlusion handling. *Image and Vision Computing*, 24(11), 1233–1243.
26. Vojir, T., Noskova, J., & Matas, J. (2014). Robust scale-adaptive mean shift for tracking. *Pattern Recognition Letters*, 49, 250–258.
27. Yang, C., Duraiswami, R., Davis, L. (2005). Efficient Mean Shift tracking via a new similarity measure. In *Proceeding computer vision and pattern recognition, vol. 1* (pp. 176–183).
28. Yilmaz, A., Javed, O., & Shah, M. (2006). Object tracking: a survey. *ACM Computer Survey*, 38(4), 13.
29. Yu, W., Tian, X., Hou, Z., Zha, Y., & Yang, Y. (2015). Multi-scale mean shift tracking. *Computer Vision, IET*, 9(1), 110–123.
30. Zhao, C., Knight, A., Reid, I. (2008). Target tracking using mean shift and affine structure. In *Proceeding of the 19th international conference in pattern recognition* (pp. 1–5).
31. Zhou, H., Yuan, Y., & Shi, C. (2009). Object tracking using SIFT features and mean shift. *Computer Vision and Image Understanding*, 113(3), 345–352.
32. Zivkovic, Z., Kröse, B. (2004). An EM-like algorithm for color histogram based object tracking. In *Proceeding IEEE conference computer vision and pattern recognition, vol. 1* (pp. 798–803).