# Efficient Mean-Shift Tracking Using an Improved Weighted Histogram Scheme

**Dejun Wang[1], Kai Chen[1],Weiping Sun[1], Shengsheng Yu[1], and Hanbing Wang[2]**
[1] School of Computer Science and Technology Huazhong University of Science and Technology
Wuhan, 430074, The People's Republic of China
[e-mail: dejunw123@gmail.com] [e-mail: kchen@hust.edu.cn] [e-mail: wpsun@hust.edu.cn] [e-mail: ssyu@mail.hust.edu.cn]
[2] Wuhan Mechanical Technology College
[e-mail:maoziwang05@qq.com]
*Corresponding author: Kai Chen

*Received January 27, 2014; revised April 4, 2014; accepted May 5, 2014; published June 27, 2014*

## Abstract

An improved Mean-Shift (MS) tracker called joint CB-LBWH, which uses a combined weighted-histogram scheme of CBWH (Corrected Background-Weighted Histogram) and LBWH (likelihood-based Background-Weighted Histogram), is presented. Joint CB-LBWH is based on the notion that target representation employs both feature saliency and confidence to form a compound weighted histogram criterion. As the more prominent and confident features mean more significant for tracking the target, the tuned histogram by joint CB-LBWH can reduce the interference of background in target localization effectively. Comparative experimental results show that the proposed joint CB-LBWH scheme can significantly improve the efficiency and robustness of MS tracker when heavy occlusions and complex scenes exist.

***Keywords:*** Target tracking, Mean-Shift, weighted histogram, target representation

This work is supported by the National Natural Science Foundation of China (NSFC) under Grant 61300140.

**http://dx.doi.org/10.3837/tiis.2014.06.009**

## 1. Introduction

Real-time object tracking has been extensively studied over many years, since it is an important step in many computer vision tasks such as human-computer interaction [1], medical imaging, robotics [2], and video surveillance [3]. Object tracking is often difficult partly due to the complex application environment with accompanying imaging noise, illumination changes, occlusions, moving cameras, changes of viewpoint, and so on. Some other fundamental problems of object tracking are due to the changes in object appearance and background. Many tracking algorithms have been proposed to overcome these challenges such as: the Mean−Shift (MS) tracker [4-6], the covariance tracker [7-9], the particle filter tracker[10-12], sparse representation-based trackers [13-15], and multiple trackers [16-18].

Target representation is one important component in typical visual trackers, but target representations usually suffer from interference from background information. To handle this problem in object tracking, extensive techniques have been presented. These target representation techniques can be classified into two main categories: pixel-oriented [19-20] and representation-oriented [4,21-23]. Some trackers may adopt the two kinds of techniques at the same time (*e.g.* [24]). The first category of techniques is applied at pixel-level so that background pixels included in the target region can been discarded to better separate the target from the background. In [19], Chen *et al*. proposed an on-line data fusion method to label pixels by combining spatial and temporal data through a dynamic Bayesian network (DBN) [25], In [20], AdaBoost is used over lots of corresponding tuned features generated from seed features to perform global color feature selection, and then a pixel-wise tracker is generated by using an object mask. The pixel-oriented techniques, however, are intuitively time-consuming and may still result in false classifications. The latter improves target representation by reducing background interference to better fit the target appearance. Comaniciu *et al*. [4] proposed a background-weighted histogram (BWH) to tune target representation. A simple representation of the background features has been exploited to select salient components from the target model and target candidate model. It decreases background interference from prominent background features in the target, and candidate, models. However, Ning *et al*. [21] demonstrated that the BWH transformation formula is actually incorrect and then proposed a corrected background-weighted histogram (CBWH) to transform only the target model but not the target candidate model: this actually reduces the relevance of background information in target localization. Wang *et al*. [23] proposed a new fusion strategy to unify all weight calculation methods within a mean-shift framework. Then they derived a new weight calculation method from the fusion strategy, which incorporates the local background [26] information in the form of target against background (TAB) formulation. Although this method performs relatively well if similar background colors are present, it may fail to track objects in challenging image sequences with drastic background appearance changes and partial occlusion. The main reason is that the background saliency is ignored by this method. In [24], Ning *et al*. extracted only those pixels corresponding to the main local binary pattern (LBP) [27] features and further proposed a joint color-texture histogram to model the target in the mean-shift algorithm, the advantage of which is the spatial information of the target and LBP texture features have been combined with the color features. However, using LBP features in mean-shift tracking remains problematic which requires further investigation. In [22], Pu *et al*. proposed a novel texture descriptor, called the Completed Local Ternary Pattern (CLTP), which is more discriminative and less sensitive to

noise in near-uniform image regions such as cheeks and foreheads. They combined CLTP and color into a joint color-CLTP histogram and used the new and distinctive target representation to perform mean-shift tracking. The joint color-CLTP histogram has achieved better robustness and tracking efficiency. While combining texture features with a color histogram is quite efficient, these methods also have some disadvantages. Tracked objects or scenes can be complex. Therefore, imposing texture constraints on the appearance of objects or a scene may not be discriminative enough under certain circumstances.

In [28], Collins *et al*. calculated a likelihood of color being found in the foreground region with respect to background to facilitate on-line feature selection. Motivated by Collins *et al*.'s work, we introduce the likelihood feature and propose a novel weighted-histogram scheme, designated the likelihood-based background-weighted histogram (LBWH) scheme, to achieve better object histogram representation. As opposed to the corrected background-weighted histogram (CBWH), the likelihood-based background-weighted histogram (LBWH) takes account of feature confidence and is robust and less sensitive to foreground changes of scene. Then we use a combined weighted-histogram scheme of CBWH and LBWH, called the joint CB-LBWH, for object tracking. Feature saliency and confidence are both used as weighted-histogram criteria. As the joint CB-LBWH tracker does not make assumptions about target appearance and is easy to compute, it is more robust and efficient especially in cases involving heavy occlusions and complex scenes.

Among the various tracking algorithms, mean-shift tracking algorithms have been extensively used in object tracking and have recently been the focus of much research due to their low complexity and robustness [21-22,29-32]. In this paper, mean-shift is also used to track the target rather than the gradient descent [33-34] method. When changes in appearance and pose arise, it tends to remain robust due to its limited search region.

The rest of the paper is organized as follows. In Section 2 we review the original BWH and CBWH scheme. In Section 3 we describe the likelihood-based background-weighted histogram (LBWH) and joint CB-LBWH in detail (in particular we show its relevance). In Section 4 we present experimental results and analyses. And finally, the salient conclusions are drawn.

## 2. Related work

In the conventional mean-shift tracking algorithm, the object is represented by a kernel-weighted color histogram because of its robustness to scale, rotation, and partial occlusion. However, it is not always discriminative enough when the tracked object has similar color features to its background. Tracking success or failure may depend primarily on object representation, thus the mean-shift algorithm is prone to failure. For a better target representation, the background model has been used to improve the color histogram.

### 2.1 BWH

Assume that we have an original background model $b = \{b_u\}_{u=1,\ldots,m}$ (with $\sum_{u=1}^{m} b_u = 1$) and its minimal non-zero entry $b^*$ of the background model in an image. The background window of the target surrounds it as a rectangular ring with a fixed area three times that of the target area, so that $b = \{b_u\}_{u=1,\ldots,m}$ is the representation of the background window and $b^*$ is the discrete density of the less salient feature. The target representation task is to get the model discriminative enough against the background which best finds the location of target in the

image. This is accomplished by a background-weighted histogram (BWH) procedure [4] in which the goodness of the target model is dependent upon the feature saliency information in the background's color histogram:

$$\left\{\tau_u = \min\left(\frac{b^*}{b_u},1\right)\right\}_{u=1\ldots m}. \tag{0}$$

The $\tau_u$ is the weight coefficient within the range of the prescribed histogram bin $u$ in the quantized feature space. Then lower $\tau_u$ values are more prominent in the background and less important for target representation: it is therefore used to transform the representations of both target model and target candidate model. A simple representation of the background features has been exploited to select salient components from the target model and target candidate model. The target model can then be obtained from:

$$q' = \{q'_u\}_{u=1,\ldots,m}; q'_u = c'_1\tau_u\sum_{i=1}^{n} k\left(\left\|\frac{x_i}{h}\right\|\right)^2 \delta\left[b(x_i)-u\right], \tag{0}$$

where $q'_u$ represents the density of feature $u$ in target model $q'$, $k(x)$ is an isotropic kernel profile, $m$ is the number of feature bins, $x_i$ ($i=1,\ldots,n$) is the pixel position in the target region centered at the original position, $\delta$ is the Kronecker delta function [35], $b(x_i)$ maps the pixel to the histogram bin index, $h$ is the bandwidth, and $c'_1$ is a normalization constant defined by:

$$c'_1 = \frac{1}{\sum_{i=1}^{n} k\left(\left\|\frac{x_i}{h}\right\|\right)^2 \sum_{u=1}^{m}\tau_u\delta\left[b(x_i)-u\right]}. \tag{0}$$

Similarly, using the same kernel profile $k(x)$, the target candidate model of the candidate region could be obtained thus:

$$p' = \{p'_u\}_{u=1,\ldots,m}; p'_u = c'_2\tau_u\sum_{i=1}^{n_h} k\left(\left\|\frac{y-x_i}{h}\right\|\right)^2 \delta\left[b(x_i)-u\right], \tag{0}$$

where $p'_u$ represents the density of feature $u$ in target candidate model $p'$, $x_i$ ($i=1,\ldots,n$) is the pixel position in the target candidate region centered at $y$, and $c'_2$ is a normalization constant defined by:

$$c'_2 = \frac{1}{\sum_{i=1}^{n_h} k\left(\left\|\frac{y-x_i}{h}\right\|\right)^2 \sum_{u=1}^{m}\tau_u\delta\left[b(x_i)-u\right]}. \tag{0}$$

### 2.2 CBWH

A corrected BWH (CBWH) algorithm is proposed by Ning *et al*. [21]. Rather than both transforming the target model and the target candidate model, it just transforms the target model to be more discriminative. Ning *et al*. proved the aforementioned BWH transformation result, in practice, as being identical to the usual target representation under the mean-shift tracking framework. That is to say, the aforementioned BWH transformation cannot reduce

the effects of prominent background features in the target candidate region for target localization. Therefore, in the CBWH algorithm the target candidate model still uses the original model as follows:

$$p = \{p_u\}_{u=1,\ldots,m}; p_u = c_2 \sum_{i=1}^{n_h} k\left(\left\|\frac{y - x_i}{h}\right\|\right)^2 \delta\left[b(x_i) - u\right], \tag{0}$$

where $p_u$ represents the density of feature $u$ in original target candidate model $p$, and $c_2$ is normalization constant defined by:

$$c_2 = \frac{1}{\sum_{i=1}^{n_h} k\left(\left\|\frac{y - x_i}{h}\right\|\right)^2 \sum_{u=1}^{m} \delta\left[b(x_i) - u\right]}. \tag{0}$$

Ning *et al*.'s corrected BWH (CBWH) scheme achieves the goal of BWH which uses the salient background features to enhance discrimination of a target model against the background. Notwithstanding the demonstrated success of CBWH, no attempts have been made to directly exploit the background confidence information, which also remains critical for object tracking like background saliency information.

## 3. The proposed weighted histogram schemes

An object to be tracked must be accurately represented as far as possible. It is not inevitable that some background pixels are present in object representation; It is also not inevitable that the background region contains some of the target features. Thus we need to justify the corresponding feature histogram in object representation.

### 3.1 Feature likelihood ratio

We first determine the region of interest, its surrounding neighborhood of three times the target area is defined as the direct background window. Given a feature space, joint simple normalized histograms $h_{fg}$ and $h_{bg}$ of the specified feature space are obtained. We follow the idea of Collins *et al*. [28] to yield a set of tuned likelihood values $L(i)$ as defined by Eq. (8):

$$L(i) = \log \frac{\max\left(h_{fg}(i), \delta\right)}{\max\left(h_{bg}(i), \delta\right)}, \tag{0}$$

Original image Likelihood image

**Fig. 1.** A typical feature likelihood image.

where $\delta$ is a small positive constant that avoids division by zero ($\delta$ is set to 0.001 in this work), and $i$ is the feature bin. Since $h_{fg}$ and $h_{bg}$ have been normalized, $h_{fg}(i)$ and $h_{bg}(i)$ imply the discrete class conditional probability densities of target and background respectively. It represents the log of the ratio of the *a posteriori* probabilities $\ln\left(p(C_1 \mid X)/P(C_2 \mid X)\right)$ for the two classes, also known as the log odds. In this paper, it indicates the probability that each feature belongs to the foreground. The results of calculating the log likelihood ratio of the feature are divided into three types of cases: firstly, it non-linearly transforms the feature(s), distinctive among the object region, into positive values; secondly, distinctive feature(s) among the background become negative values; and thirdly, it collapses towards zero for shared feature values present in both object and background. At the same time, it actually forms a new likelihood feature [28]. A typical feature likelihood image is shown in **Fig. 1**.

### 3.2 LBWH

We would like distributions of target and background to ideally be approximated by their histogram as much as possible. Differing from Comaniciu *et al*. [4] this background model is a more complex representation of the background features; a likelihood-based background model (LB) is derived and used to select the most confident components from the target model. This scheme is called the likelihood-based background-weighted histogram (LBWH) scheme.

After $L(i)$ is calculated as Eq. (8), it bestows a measure of confidence in the evaluation of the foreground for each feature that appears in a histogram bin. However, $L(i)$ is unbounded and lacks any sense of probability of occurrence. To overcome these drawbacks, the weight $\pi(i)$ for each feature is obtained by using a sigmoid M-estimator as shown in Eq. (9):

$$\pi(i) = \max\left(\frac{1}{1+\exp\left(-\dfrac{L(i)-m}{n}\right)}, 0.01\right), \tag{0}$$

where *m*, *n* are constants chosen based on how much we want to reduce interference inside the target region and obtain more robust target model.

The new background model LB is given by $\hat{b} = \left\{\hat{b}_u\right\}_{u=1,\ldots,m}$ (with $\sum_{u=1}^{m} \hat{b}_u = 1$) and is obtained by normalizing $\pi(i)$. Denoting $\tilde{b}$ by its maximal non-zero value in $\left\{\hat{b}_u\right\}_{u=1,\ldots,m}$, we then define a transformation for the representation of the target model and a weight coefficient histogram is calculated from Eq. (10):

$$\left\{\hat{\tau}_u = \max\left(\frac{\tilde{b}}{b_u}, 1\right)\right\}_{u=1\ldots m}. \tag{0}$$

The transformation reduces the weights of those features with low $\hat{\tau}_u$, that is, the confident features belonging to the background. The new target model is then defined by Eq.(1):

$$\hat{q}_u = \hat{c}_1 \hat{\tau}_u \sum_{i=1}^{n} k\left(\left\|\frac{x_i}{h}\right\|\right)^2 \delta\left[b(x_i) - u\right], \tag{1}$$

where $\hat{c}_1^{'}$ is the normalization constant and expressed by Eq.(2):

$$\hat{c}_1^{'} = \frac{1}{\sum_{i=1}^{n} k\left(\left\|\frac{x_i}{h}\right\|\right)^2 \sum_{u=1}^{m} \hat{\tau}_u \delta\left[b(x_i) - u\right]}. \tag{2}$$

### 3.3 Joint CB-LBWH

Ning *et al.* used a corrected BWH (CBWH) scheme to down-weight the salient background features to fix only the target model but not the target candidate model. Although the idea of CBWH is reasonable, they only take into account the saliency of probability mass among the background region and ignore the foreground information. Since the likelihood feature encodes foreground and background information, the proposed likelihood-based background-weighted histogram scheme (LBWH) works well due to foreground and background information being exploited. However, if two sample points from the foreground region have the same feature confidence, the saliency of the background region is not always the same. CBWH down-weights salient background features to suppress background interference and focuses on discriminative features from the target region; LBWH computes the likelihood feature using both the background and foreground feature histograms, and transforms the likelihood value to suppress background interference. They are partly complementary for the purposes of adjusting target histograms, so it is feasible to fuse two types of different background features. We believe that combining CBWH and LBWH schemes could make a tracker more robust due to fusing two types of different background features, so we have a combined weighted histogram scheme comprising CBWH and LBWH, which is called the joint CB-LBWH.

In this joint CB-LBWH, we consider both feature confidence and saliency of background. Using Eq. (1) and Eq. (10), for each feature bin in the target region, we assign new weight coefficient to get the new target histogram model. The new weights are given by the product of the aforementioned two weights as in Eq.(3):

$$\left\{\omega_u = \tau_u \times \hat{\tau}_u\right\}_{u=1\ldots m}. \tag{3}$$

Then the new tuned target model becomes that shown in Eq.(4):

$$\hat{q}_u = C\omega_u \sum_{i=1}^{n} k\left(\left\|\frac{x_i}{h}\right\|\right)^2 \delta\left[b(x_i) - u\right], \tag{4}$$

where $C$ is the normalization constant given by Eq.(5):

$$C = \frac{1}{\sum_{i=1}^{n} k\left(\left\|\frac{x_i}{h}\right\|\right)^2 \omega(x_i)}. \tag{5}$$

Similarly in [21], $\omega_u$ is only used to transform the target model but not the target candidate model.

An example of the target model image of CBWH, LBWH, and joint CB-LBWH is shown in **Fig. 2**. The corresponding non-zero weights of the features therein are shown in **Fig. 3**. For the original image shown in **Fig. 2**, we compute the Bhattacharyya similarities [29] between the target model and its surrounding background region by CBWH, LBWH, and joint CB-LBWH. CBWH, LBWH, and joint CB-LBWH have Bhattacharyya similarities [29] of 0.04, 0.07, and 0.02 respectively, which implies that joint CB-LBWH can best distinguish target from

background. Since the weight of each feature in the target model in joint CB-LBWH is determined by the feature confidence and saliency of the background region, the joint CB-LBWH scheme is more robust and efficient. After the joint CB-LBWH has run its course, the mean-shift tracker searches the target in the current frame using the tuned target model.

(A) original image (B) CBWH (C) LBWH (D) joint CB-LBWH

**Fig. 2.** An example of the target model image.

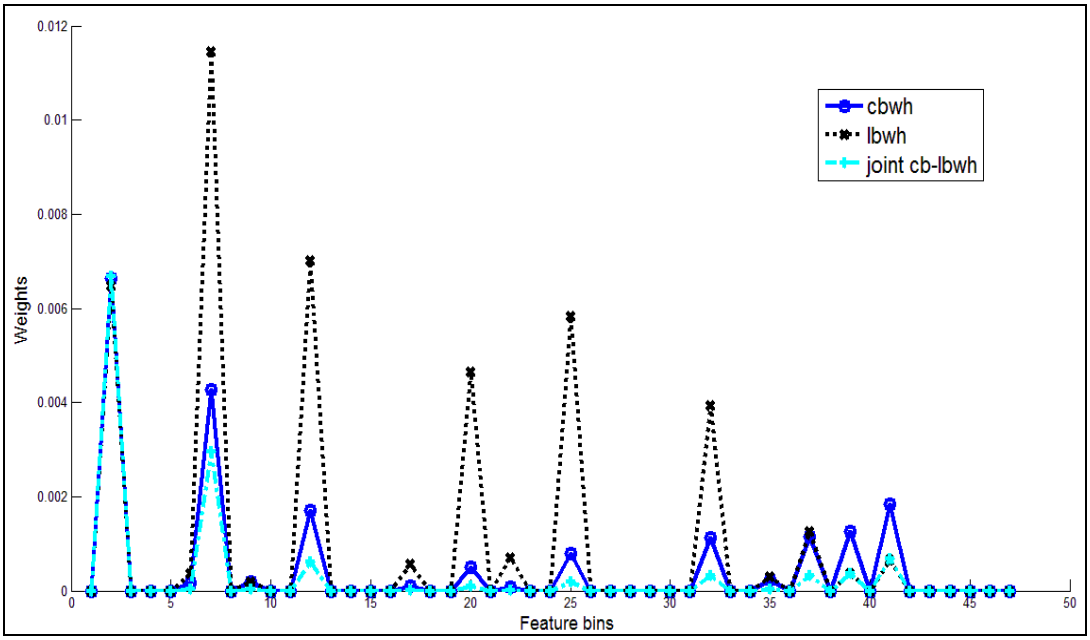



**Fig. 3.** Corresponding non-zero weights of the features in **Fig. 2**.

## 4. Experimental work

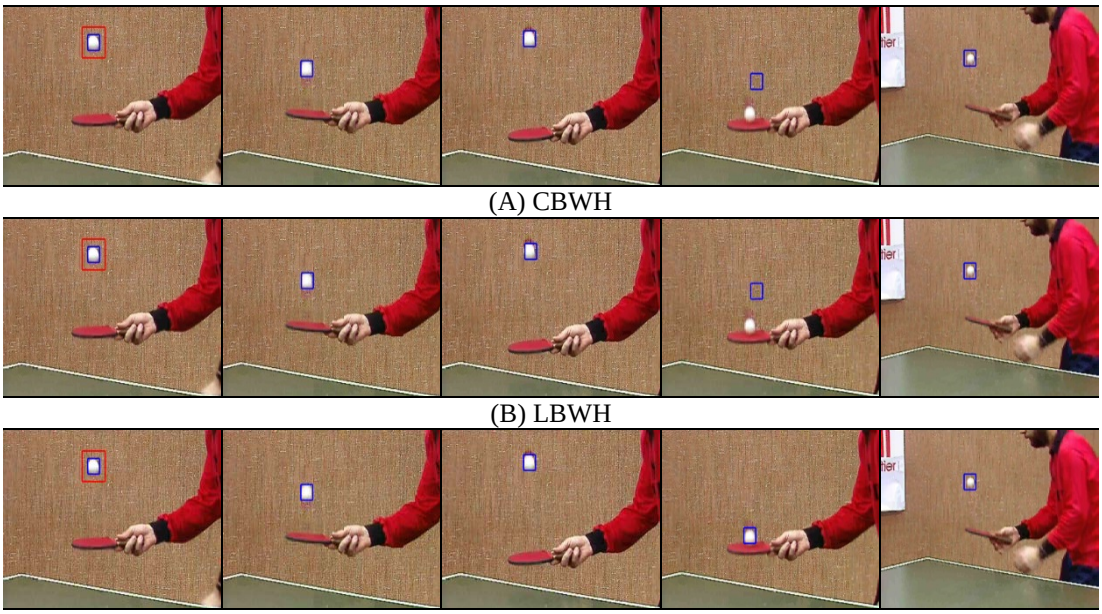
To verify the effectiveness of the proposed algorithm, several video sequences are used to track a moving object. The experiments are done using an AMD Athlon Dual-Core 2.3 GHz processor, running MS-Windows™ 7.0, with 2.0 GB of RAM; and using MATLAB™ R2011a. We select an Epanechnikov kernel profile for the mean-shift tracking algorithm. The Bhattacharyya function is used as a proximity measurement function. The histograms are computed under an RGB color feature space and are quantized into 16 × 16 × 16 bins. We take 20 iterative steps as the mean-shift algorithm termination criterion.

In the experiments, we compare the proposed scheme (joint CB-LBWH) with the corrected background-weighted histogram (CBWH) and likelihood-based background-weighted histogram (LBWH) scheme.

**Table 1.** Average error and average number of iterations

| Video sequences | CBWH | | LBWH | | Joint CB-LBWH | |
|---|---|---|---|---|---|---|
| | Average error | Average number of iterations | Average error | Average number of iterations | Average error | Average number of iterations |
| Ping-Pong ball | 3.38 | 3.23 | 3.41 | 3.56 | 2.61 | 3.10 |
| Table tennis player | 2.96 | 3.47 | 2.29 | 3.62 | 3.07 | 2.95 |
| Walking woman | 35.68 | 4.93 | 19.97 | 4.86 | 11.48 | 4.38 |
| ThreePast Shop2cor | 17.84 | 3.88 | 17.43 | 3.56 | 15.63 | 3.48 |
| Ping-Pong ball (24 × 28) | 4.39 | 5.71 | 5.23 | 6.79 | 3.50 | 5.06 |
| Ping-Pong ball (29 × 32) | 2.87 | 5.04 | 3.97 | 5.42 | 2.55 | 3.42 |

We measure the mean position error distance of the target localization and average iteration numbers of each scheme in **Table 1**. The first row of **Table 1** shows the average distance and average iteration numbers of each scheme on the benchmark "Ping-Pong ball sequence" which is used in [21]. The sequence has 52 frames each measuring 352 × 240 pixels. The target to be tracked is the moving ball. Since there are distinctive color differences between the target and the background, the CBWH and LBWH models locate the target well except in the case where the ball touches the bat in Frame 26. In a video sequence, if sudden background changes arise, they tend to contaminate the target region due to the presence of marginal pixels: in the end the tracker may lose its target. To suppress only the salient background features or the confident features belonging to the background is inefficient and will not readily distinguish target from background. Based on suppressing the aforementioned two types of background features, the joint CB-LBWH remains sufficient discriminative power between target and background. Thus the joint CB-LBWH still successfully locks onto the target in Frame 26. Tracking results from the three schemes are shown in **Fig. 4** which shows that the joint CB-LBWH is more discriminatory and more robust in the face of sudden background change. **Table 1** lists the number of iterations in each of the three schemes. The average number of iterations is 3.1 for joint CB-LBWH, 3.23 for CBWH, and 3.96 for LBWH. The joint CB-LBWH method requires fewer computations.



(A) CBWH

(B) LBWH

(C) joint CB-LBWH

**Fig. 4.** Tracking results from the Ping-Pong ball sequence for Frames 1, 13, 22, 26, and 45.

In the second "tennis player sequence" video (as used in [21]), one man plays table tennis in the room. The head of the player is labelled with a rectangle measuring 18 × 25 pixels (inner blue rectangle) in Frame 1 as the tracked target. As the man plays, his head keeps changing position and angle. Since the color histogram is robust to rotation invariance and the video sequence has distinctive color differences between target and background, all of the three methods locate the target. By introducing the likelihood feature to fix the target model, the proposed likelihood-based background-weighted histogram scheme (LBWH) work well due to foreground and background information being exploited, especially in the case where the appearance of the target changes and the appearance of the background seldom changes. From the second row of **Table 1**, as far as target localization accuracy is concerned, the joint CB-LBWH method performs slightly worse than CBWH. LBWH performs best since the foreground features of the target model are well exploited in target foreground change scenes. We believe that the improper use of background color saliency information for the joint CB-LBWH method may have weakened the role of the likelihood feature for this video sequence. For brevity, we only show experimental results from the proposed joint CB-LBWH method in **Fig. 5**. The three methods: CBWH, LBWH, and joint CB-LBWH have an average number of iterations of 3.47, 3.62, and 2.95, respectively. The joint CB-LBWH scheme needs the fewest average number of iterations: **Fig. 6** shows the number of iterations needed by the methods.

**Fig. 5.** Tracking results of joint CB-LBWH method used on table tennis player sequence for Frames 1, 16, 26, 40, and 54.

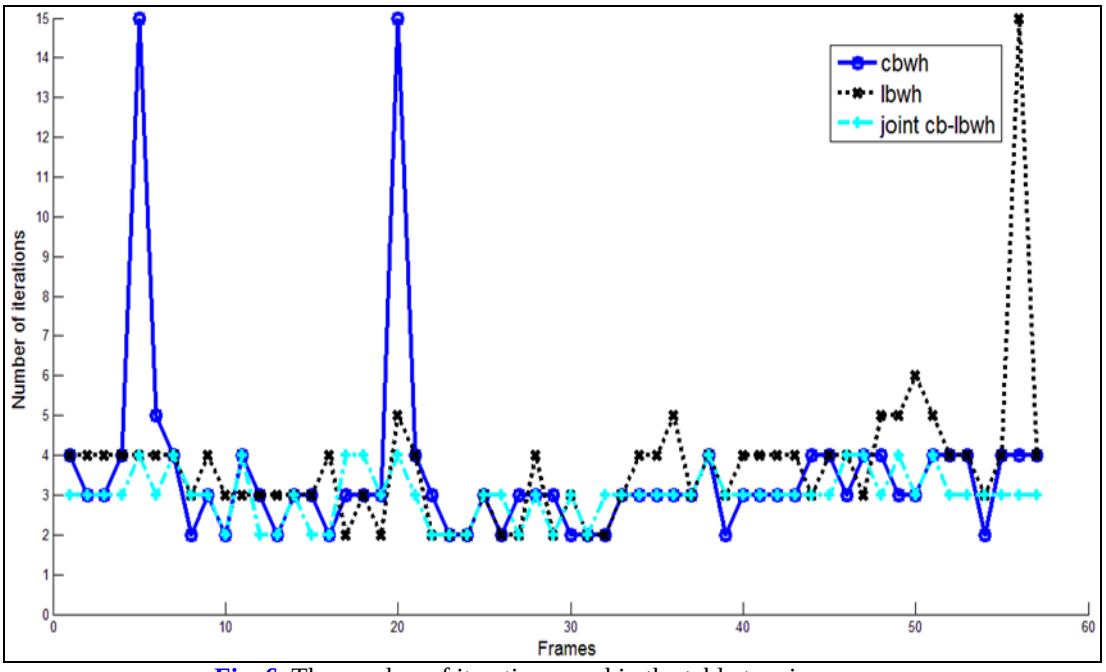
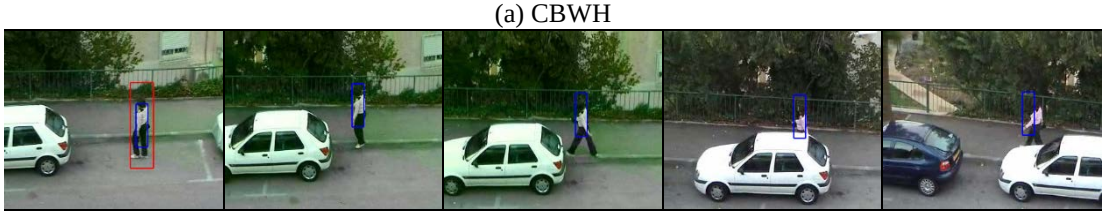


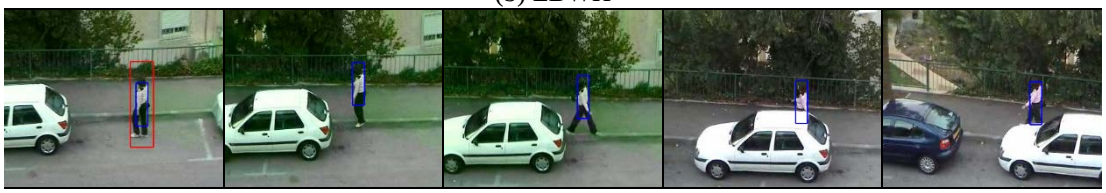

**Fig. 6.** The number of iterations used in the table tennis sequence.

The third experiment tests the methods on the "walking woman sequence" [36], where the tracked target is the walking woman. In this sequence, the walking woman is part-occluded by a car whose color is similar to that of the woman's shirt. At the same time, the background also undergoes significant changes. Tracking results of the three methods are shown in **Fig. 7**. Because of the changing background and heavy occlusion during tracking, the target and background are hard to distinguish by taking account of only background color features, thus CBWH locks the target early in the period but loses it before the end. Since the background likelihood features about the confident knowledge belonging to the foreground are used to suppress interference from background occlusion, LBWH is suitable for tracking in this case. The joint CB-LBWH model combines the two types of different background features so that it further enhances the power to discriminate between target and background. The joint CB-LBWH and LBWH methods could still track the target in the end, while the joint CB-LBWH method achieves higher target localization accuracy. As far as the average number of iterations is concerned, the joint CB-LBWH also performs best. This indicates that the joint CB-LBWH model is more discriminative and more robust against occlusion and can more efficiently integrate target and background information than CBWH and LBWH.

(a) CBWH



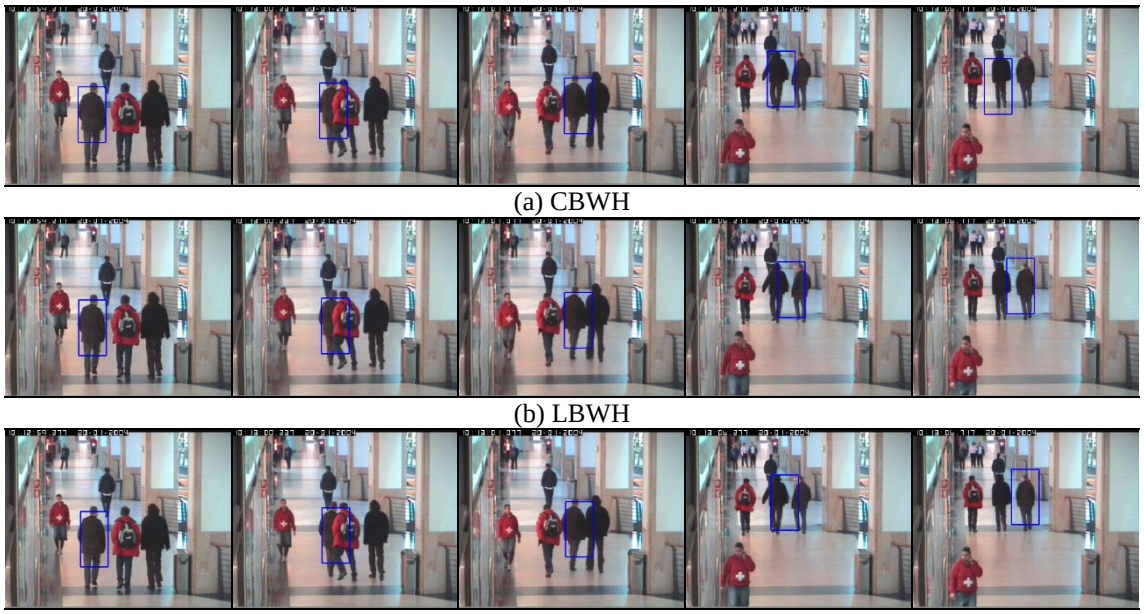
(b) LBWH



(c) joint CB-LBWH

**Fig. 7.** Tracking results from walking woman sequence for Frame 1, 23, 29, 53 and 87.

The fourth experiment uses a more complex "ThreePastShop2cor sequence" [37]; in the sequence, the tracked left-most person walks with two people together in a corridor and exchanges mutual positions with each of other two *en route*. The right-most person is very similar to the target person; meanwhile there are heavy occlusions and obvious illumination variations in the sequence. These are the main challenges posed by the sequence. Tracking results of the three methods are shown in **Fig. 8**. After many consecutive occlusion frames, since the feature confidence is ignored by CBWH, the CBWH tracker will tend to track the similar object close to the target. Meanwhile errors accumulate over time across the frames. Hence the CBWH tracker drifts from the target in Frame 190. For LBWH, since it incorporates feature confidence, which considers both foreground and background regions, to enhance the approximation of the target model, it is more discriminative and robust to similar background than CBWH; for the joint CB-LBWH, it is based on the other two methods and the approximation of target model is better and has higher target localization accuracy than LBWH if similar backgrounds is close to the target. From **Table 1** we can see that LBWH performs slightly better than CBWH and the joint CB-LBWH model is more accurate than other two methods. We compute the Bhattacharyya similarities between the target model and its surrounding background region by CBWH, LBWH, and joint CB-LBWH in Frame 1: CBWH, LBWH, and joint CB-LBWH have Bhattacharyya similarities of 0.2939, 0.2537, and 0.098 respectively, which implies that joint CB-LBWH can best distinguish target from background. As far as the average number of iterations is concerned, the joint CB-LBWH model outperforms the other two methods. The average number of iterations is 3.48 for joint CB-LBWH, 3.88 for CBWH, and 3.56 for LBWH.



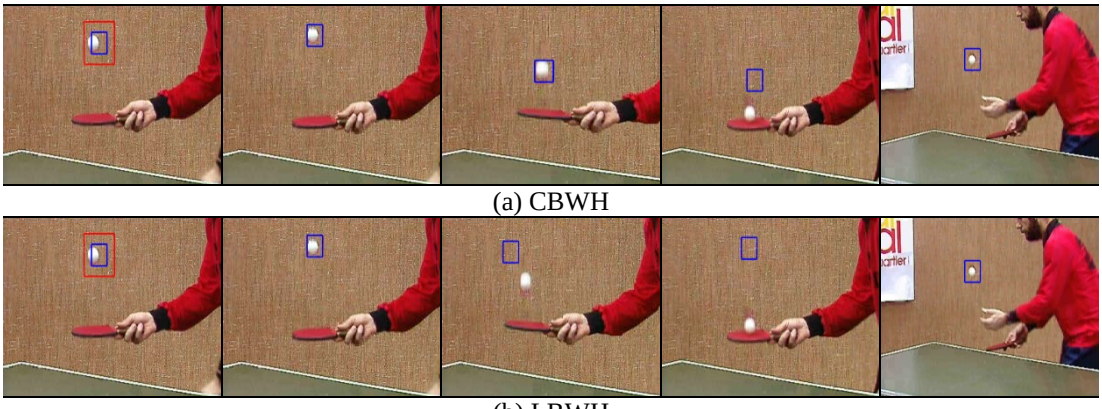
(a) CBWH

(b) LBWH

(c) joint CB-LBWH

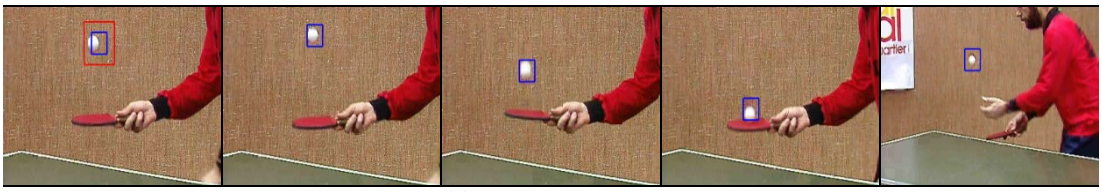
**Fig. 8.** Tracking results from the ThreePastShop2cor sequence for Frames 3, 27, 48, 178, and 190.

Additionally, we use the "Ping-Pong ball sequence" to test the robustness of each method for inaccurately labelled targets since in many real-tracking systems inaccurate target initialization often occurs [8]. In this experiment, the moving ball is inaccurately labelled with a hand-drawn rectangle measuring 24 × 28 pixels (inner blue rectangle) in Frame 1. Since the initial target region occupies only half of the ball and much background information, it is handicapped severely by such a deliberate inaccurate initialization. When the ball moves quickly, the CBWH and LBWH methods fail to track the target and then gradually recover the correct tracking. Since CBWH reduces the impact of features shared by the target and background, and decreases the relevance of the background to target localization, it reduces the dependency of mean-shift tracking on target initialization. Thus CBWH performs better than LBWH. Due to the proper fusing of feature saliency information and confidence information, the joint CB-LBWH model is less sensitive to bad target initialization, so the joint CB-LBWH method can robustly track the target and is more stable than the others. Tracking results from the three methods are shown in Fig. 9.



(a) CBWH

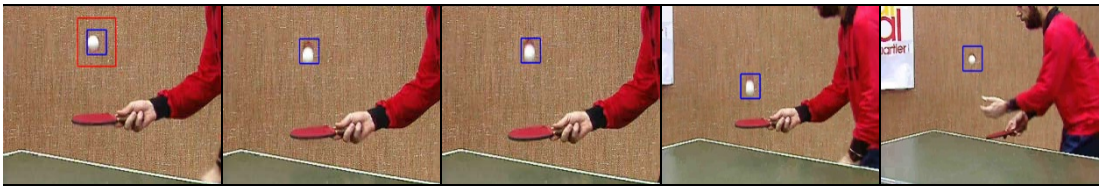
(b) LBWH



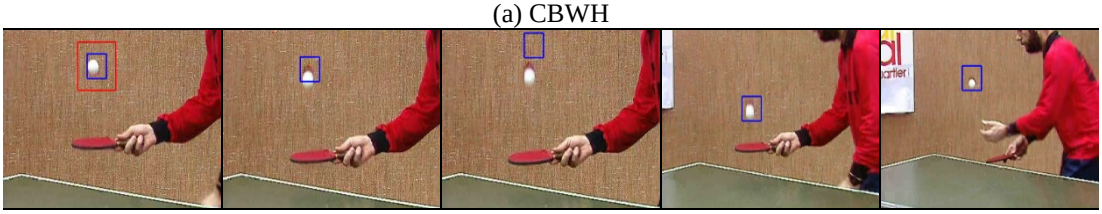
(c) joint CB-LBWH

**Fig. 9.** Tracking results from the Ping-Pong ball sequence with inaccurate initialization (size 24 × 28 pixels) for Frames 1, 2, 13, 26, and 50

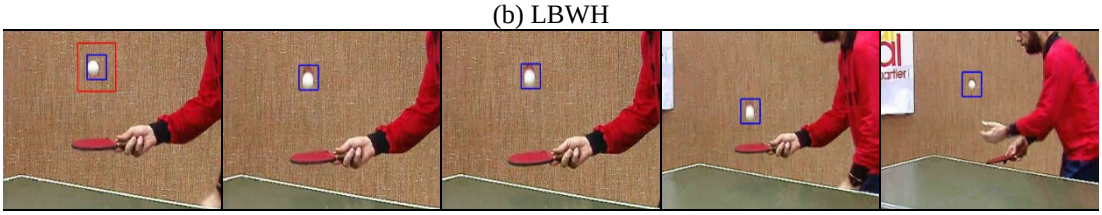
Another inaccurate initialization is tested on this sequence. We hand-label (inaccurately) the fast-moving ball with a rectangle measuring 29 × 32 pixels (inner blue rectangle) in Frame 1. Tracking results from selected frames from each method are shown in **Fig. 10** which shows that the proposed approach works best (*i.e.* the same as in the case of the first inaccurate initialization).



(a) CBWH



(b) LBWH



(c) joint CB-LBWH

**Fig. 10.** Tracking results from the Ping-Pong ball sequence with inaccurate initialization (size 29 × 32 pixels) for Frames 1, 8, 23, 39

**Table 1** lists the average number of mean shift iterations required by the three schemes for the six experiments. Compared with CBWH and LBWH, the joint CB-LBWH model needs the lowest average number of iterations. The main difference among all mean-shift-like tracking algorithms lies in the weight calculation [23]. For different mean-shift-like tracking algorithms, the same sample points in the target have different weight coefficients, that is to say, their importance is not the same. Rather than considering each single feature type, the joint CB-LBWH model efficiently extracts likelihood, and saliency, features and transforms them to suppress background interference and highlights the main target features in the search-region. Therefore the proposed joint CB-LBWH enhances the weight calculation during the tracking iterations, which improves the mean shift vector and the convergence of mean shift tracking becomes more rapid. In the end, the average number of iterations needed by the joint CB-LBWH method is smaller than the others. Meanwhile joint CB-LBWH yields

better tracking results.

## 5. Conclusions

The kernel-weighted histogram in the mean-shift tracker is improved upon by introducing feature confidence as measured by the statistical likelihood value calculated from a kernel density estimate of the background and foreground feature distribution. Based on the likelihood feature, a more complex background model is derived: the likelihood-based background model (LB). Then a novel weighted-histogram scheme is proposed: the likelihood-based background-weighted histogram (LBWH) scheme. The more prominent and confidently assigned features are deemed to have more significance for tracking the target. The tuned target histogram, with the weight values which are assigned by these features, can reduce background interference during target localization. Subsequently the CBWH and LBWH schemes are combined and the resulting scheme is designated the appellation: joint CB-LBWH. Finally, mean-shift tracking is performed. The major advantage of this scheme lies in that it can be applied to various complex objects under different environments. Our experiments demonstrate that the proposed joint CB-LBWH scheme significantly improves the efficiency and robustness of a mean-shift tracker in the presence of heavy occlusions and complex scenes.

## References

[1] G. R. Bradski, “Real time face and object tracking as a component of a perceptual user interface,” in *Proc. of the IEEE International Workshop on Applications of Computer Vision*, pp. 214–219, 1998. http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=732882

[2] N. P. Papanikolopoulos, P. K. Khosla, and T. Kanade, “Visual tracking of a moving target by a camera mounted on a robot: a combination of control and vision,” *IEEE Transactions on Robotics and Automation*, vol. 9, no. 1, pp. 14–35, 1993. Article (CrossRef Link)

[3] C. Stauffer and W. E. L. Grimson, “Learning Patterns of Activity Using Real-Time Tracking,” *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 22, no. 8, pp. 747–757, 2000. Article (CrossRef Link)

[4] D. Comaniciu, V. Ramesh, and P. Meer, “Kernel-based object tracking,” *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 25, no. 5, pp. 564–577, 2003. Article (CrossRef Link)

[5] A. Dulai and T. Stathaki, “Mean shift tracking through scale and occlusion,” *IET Signal Processing*, vol. 6, no. 5, pp. 534–540, 2012. Article (CrossRef Link)

[6] I. Leichter, “Mean Shift Trackers with Cross-Bin Metrics,” *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 34, no. 4, pp. 695–706, 2012. Article (CrossRef Link)

[7] F. Porikli, O. Tuzel, and P. Meer, “Covariance Tracking using Model Update Based on Lie Algebra,” in *Proc. of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR*, vol. 1, pp. 728–735, 2006. http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=1640826

[8] A. Romero, M. Gouiffés, and L. Lacassagne, “Covariance Descriptor Multiple Object Tracking and Re-identification with Colorspace Evaluation,” in *Proc. of the IEEE ACCV - Workshop on Detection and Tracking in Challenging Environments*, vol. 7729, pp. 400–411, 2013. http://link.springer.com/chapter/10.1007%2F978-3-642-37484-5_33#

[9] Y. W. Xuguang Zhang, Xiaoli Li, Ming Liang, “Covariance tracking with forgetting factor and random sampling,” *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, vol. 19, no. 3, pp. 547−558, 2011. Article (CrossRef Link)

[10] M. Isard and A. Blake, “CONDENSATION—Conditional Density Propagation for Visual Tracking,” *International Journal of Computer Vision*, vol. 29, no. 1, pp. 5–28, 1998.

Article (CrossRef Link)
[11] S. K. Zhou, R. Chellappa, and B. M., "Visual tracking and recognition using appearance-adaptive models in particle filters," *IEEE Trans. Image Processing*, vol. 11, pp. 1491–1506, 2004. Article (CrossRef Link)
[12] J. Kwon, K. M. Lee, and F. C. Park, "Visual tracking via geometric particle filtering on the affine group with optimal importance functions," in *Proc. of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR*, pp. 991–998, 2009. http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=5206501&tag=1
[13] X. Mei and H. Ling, "Robust visual tracking using L1 minimization," in *Proc. of the International Conference Computer Vision, ICCV*, pp. 1436–1443, 2009. http://ieeexplore.ieee.org/iel5/5453389/5459144/05459292.pdf?arnumber=5459292
[14] Y. Wu, E. Blasch, G. Chen, L. Bai, and H. Ling, "Multiple source data fusion via sparse representation for robust visual tracking," in *Proc. of the 14th International Conference on Information Fusion (FUSION)*. pp. 1–8, 2011. http://ieeexplore.ieee.org/iel5/5959904/5977431/05977451.pdf?arnumber=5977451
[15] X. Mei and H. Ling, "Robust visual tracking and vehicle classification via sparse representation," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 33, no. 11, pp. 2259–72, 2011. Article (CrossRef Link)
[16] B. Stenger, T. Woodley, and R. Cipolla, "Learning to track with multiple observers," in *Proc. of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR*, pp. 2647–2654, 2009. http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=5206634
[17] J. Kwon and K. M. Lee, "Visual tracking decomposition," in *Proc. of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR*, pp. 1269–1276, 2010. http://cv.snu.ac.kr/publication/conf/2010/VTD_CVPR2010.pdf
[18] Y. Gao, R. Ji, L. Zhang, and A. Hauptmann, "Symbiotic Tracker Ensemble Towards A Unified Tracking Framework," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. PP, no. 99, pp. 1, 2014. http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6720135
[19] J. Chen and Q. Ji, "Online Spatial-temporal Data Fusion for Robust Adaptive Tracking," in *Proc. of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR*, pp. 1–8, 2007. http://www.ecse.rpi.edu/~cvrl/chenj/track.pdf
[20] F. Wei, S. Chou, and C. Lin, "A region-based object tracking scheme using Adaboost-based feature selection," in *Proc. of the IEEE International Symposium on Circuits and Systems(ISCAS)*, pp.2753–2756, 2008. http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=4542027
[21] J. Ning, L. Zhang, D. Zhang, and C. Wu, "Robust mean-shift tracking with corrected background-weighted histogram," *IET Computer Vision*, vol. 6, no. 1, pp. 62 - 69, 2012. http://www4.comp.polyu.edu.hk/~cslzhang/paper/IET_CV_2010.pdf
[22] Xiao. Pu and Zhi. Z, "A More Robust Mean Shift Tracker on Joint Color-CLTP Histogram," *International Journal of Image, Graphics and Signal Processing (IJIGSP)*, vol.4, no.12, pp.34-42, 2012. http://www.mecs-press.org/ijigsp/ijigsp-v4-n12/v4n12-5.html
[23] L. Wang, C. Pan, and S. Xiang, "Mean-shift tracking algorithm with weight fusion strategy," in *Proc. of the 18th* IEEE *International Conference on Image Processing (ICIP)*, pp.*473 - 476*, 2011. http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6116554&tag=1
[24] J. Ning, L. Zhang, D. Zhang, and C. Wu, "Robust Object Tracking Using Joint Color-Texture Histogram," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 23, no. 07, pp. 1245–1263, 2009. Article (CrossRef Link)
[25] Z. Ghahramani, "Learning dynamic Bayesian networks," *Adaptive Processing of Sequences and Data Structures: Lecture Notes in Artificial Intelligence*, Springer-Verlag, Berlin, vol.1387, pp. 168–197, 1998. Article (CrossRef Link)
[26] L. Wang, H. Wu, and C. Pan, "Mean-Shift Object Tracking with a Novel Back-Projection Calculation Method," in *Proc. of the 9th Asian Conference on Computer Vision*, vol. 5994, pp. 83–92, 2010. http://link.springer.com/chapter/10.1007/978-3-642-12307-8_8
[27] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Transactions on Pattern Analysis and*

*Machine Intelligence*, vol. 24, no. 7. pp. 971–987, 2002. Article (CrossRef Link)

[28] R. T. Collins, "Online Selection of Discriminative Tracking Features," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 10, pp. 1631–1643, Oct. 2005. Article (CrossRef Link)

[29] D. Comaniciu, V. Ramesh, and P. Meer, "Real-time tracking of non-rigid objects using mean shift," in *Proc. of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR*, pp. 142–149, 2000. http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=854761&tag=1

[30] T. Vojir, J. Noskova, and J. Matas, "Robust Scale-Adaptive Mean-Shift for Tracking," *Image Analysis*, Springer Berlin Heidelberg, pp. 652-663, 2013. http://link.springer.com/chapter/10.1007%2F978-3-642-38886-6_61

[31] J. Ning, L. Zhang, D. Zhang, and C. Wu, "Scale and orientation adaptive mean shift tracking," *IET Computer Vision*, vol. 6, no. 1, pp. 52-61, 2012. Article (CrossRef Link)

[32] R. Duraiswami and L. Davis, "Efficient Mean-Shift Tracking via a New Similarity Measure," in *Proc. of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR*, vol. 1, pp. 176–183, 2005. http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=1467265

[33] B. D. Lucas and T. Kanade, "An Iterative Image Registration Technique with an Application to Stereo Vision," in *Proc. of the 7th International Joint Conference on Artificial Intelligence*, vol. 2, pp. 674–679, 1981. http://ijcai.org/Past%20Proceedings/IJCAI-81-VOL-2/PDF/017.pdf

[34] G. D. Hager and P. N. Belhumeur, "Efficient region tracking with parametric models of geometry and illumination," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 20, no. 10. pp. 1025–1039, 1998. Article (CrossRef Link)

[35] J. H. Heinbockel, *Introduction to Tensor Calculus and Continum Mechanics*, Trafford Publishing, Vol. 52, pp. 14–31, 2001. http://nebm.ist.utl.pt/repositorio/download/1090

[36] E. R. and I. S. A. Adam, "Robust fragments-based tracking using the integral histogram," in *Proc. of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR*, vol. 1, pp. 798–805, 2006. http://dl.acm.org/citation.cfm?id=1153412

[37] "CAVIAR Test Case Scenarios," http://www.homepages.inf.ed.ac.uk/rbf/ CAVIAR

**Dejun Wang** received the M.S. degree from Yunnan University, in 2005. Then he was a lecturer in the School of Computer at Hubei University of Technology. He is now pursuing his Ph.D. degree in the School of Computer Science & Technology at Huazhong University of Science and Technology. His research interests include video analysis, pattern recognition, and embedded system.

**Kai Chen** is a lecturer in the School of Computer Science &Technology at Huazhong University of Science and Technology. His research interests include pattern recognition, computer network application, computer network security and computer network protocol analysis.

**Weiping Sun** is an associate professor in the School of Computer Science & Technology at Huazhong University of Science and Technology. Her research interests include computer vision, computer network and storage, and pattern recognition.

**Shengsheng Yu** is a professor in the School of Computer Science & Technology at Huazhong University of Science and Technology. His research interests include discrete signal processing, computer network and storage, and communication.

**Hanbing Wang** is a lecturer in Wuhan Mechanical Technology College. He graduated from Huazhong University of Science and Technology as a Master of Software Engineering.