www.ietdl.org

Published in IET Computer Vision
Received on 1st August 2009
Revised on 27th June 2010
doi: 10.1049/iet-cvi.2009.0075

ISSN 1751-9632

Robust mean-shift tracking with corrected background-weighted histogram

J. Ning[1,2,3] L. Zhang[2] D. Zhang[2] C. Wu[1]

[1]The State Key Laboratory of Integrated Service Networks, Xidian University, Xi'an, People's Republic of China
[2]Department of Computing, The Hong Kong Polytechnic University, Kowloon, Hong Kong, People's Republic of China
[3]College of Information Engineering, Northwest A&F University, Yangling, People's Republic of China
E-mail: cslzhang@comp.polyu.edu.hk

Abstract: The background-weighted histogram (BWH) algorithm proposed by Comaniciu *et al.* attempts to reduce the interference of background in target localisation in mean-shift tracking. However, the authors prove that the weights assigned to pixels in the target candidate region by BWH are proportional to those without background information, that is, BWH does not introduce any new information because the mean-shift iteration formula is invariant to the scale transformation of weights. Then a corrected BWH (CBWH) formula is proposed by transforming only the target model but not the target candidate model. The CBWH scheme can effectively reduce background's interference in target localisation. The experimental results show that CBWH can lead to faster convergence and more accurate localisation than the usual target representation in mean-shift tracking. Even if the target is not well initialised, the proposed algorithm can still robustly track the object, which is hard to achieve by the conventional target representation.

1 Introduction

Object tracking is an important task in computer vision. Many algorithms [1] have been proposed to solve the various problems arisen from noises, clutters and occlusions in the appearance model of the target to be tracked. Among various object-tracking methods, the mean-shift tracking algorithm [2–4] is a popular one because of its simplicity and efficiency. Mean shift is a non-parametric density estimator that iteratively computes the nearest mode of a sample distribution [5]. After it was introduced to the field of computer vision [6], mean shift has been adopted to solve various problems, such as image filtering, segmentation [7–11] and object tracking [2, 3, 12–18].

In the mean-shift tracking algorithm, the colour histogram is used to represent the target because of its robustness to scaling, rotation and partial occlusion [2, 3, 19]. However, the mean-shift algorithm is prone to local minima when some of the target features are present in the background. Therefore Comaniciu *et al.* [3] further proposed the background-weighted histogram (BWH) to decrease background interference in target representation. The strategy of BWH is to derive a simple representation of the background features and to use it to select the salient components from the target model and target candidate model. More specifically, BWH attempts to decrease the probability of prominent background features in the target model and candidate model and thus reduce the background's interference in target localisation. Such an idea is reasonable and intuitive, and some works have been proposed to follow this idea [20–22]. In [20], the object is partitioned into a number of fragments and then the target model of each fragment is enhanced by using BWH. Different from the original BWH transformation, the weights of background features are derived from the differences between the fragment and background colours. In [21], the target is represented by combining BWH and adaptive kernel density estimation, which extends the searching range of the mean-shift algorithm. In addition, Allen *et al.* [22] proposed a parallel implementation of mean-shift algorithm with adaptive scale and BWH and demonstrated the efficiency of their technique in a single instruction multiple data computer. All the above BWH-based methods aim to decrease the distraction of background in target location to enhance mean-shift tracking. Unfortunately, all of them do not notice that the BWH transformation formula proposed in [3] is actually incorrect, which will be proved in this paper.

In this paper, we demonstrate that the BWH algorithm will simultaneously decrease the probability of prominent background features in the target model and target candidate model. Thus, BWH is equivalent to a scale transformation of the weights obtained by the usual target representation method in the target candidate region. Meanwhile, the mean-shift iteration formula is invariant to the scale transformation of weights. Therefore the mean-shift tracking with BWH in [3, 20–22] is exactly the same as the mean-shift tracking with usual target representation.

Based on the mean-shift iteration formula, the key to effectively exploit the background information is to decrease the weights of prominent background features. To this end, we propose to transform only the target model but not the target candidate model. A new formula for computing the pixel weights in the target candidate region is then derived. The proposed corrected background-weighted histogram (CBWH) can truly achieve what the original BWH method wants: reduce the interference of background in target localisation. An important advantage of the proposed CBWH method is that it can work robustly even if the target model contains much background information. Thus, it reduces greatly the sensitivity of mean-shift tracking to target initialisation. In the experiments, we can see that even when the initial target is not well selected, the proposed CBWH algorithm can still correctly track the object, which is hard to achieve by the usual target representation.

The rest of the paper is organised as follows. Section 2 introduces briefly the mean-shift algorithm and the BWH method. Section 3 proves that the BWH method is equivalent to the conventional mean-shift tracking method, and then the CBWH algorithm is presented. Section 4 presents experiments to test the proposed CBWH method. Section 5 concludes the paper.

2 Mean-shift tracking and BWH

2.1 Target representation

In object tracking, a target is usually defined as a rectangle or an ellipsoidal region in the frame and the colour histogram is used to represent the target. Denote by $\{\mathbf{x}_i^*\}_{i=1,\ldots,n}$ the normalised pixels in the target region, which has n pixels. The probability of a feature u, which is actually one of the m colour histogram bins, in the target model is computed as [2, 3]

$$\hat{q} = \{\hat{q}_u\}_{u=1,\ldots,m}; \quad \hat{q}_u = C\sum_{i=1}^{n} k(||\mathbf{x}_i^*||^2)\delta[b(\mathbf{x}_i^*) - u] \tag{1}$$

where $\hat{q}$ is the target model, $\hat{q}_u$ is the probability of the u^{th} element of $\hat{q}$, δ is the Kronecker delta function, $b(\mathbf{x}_i^*)$ associates the pixel $\mathbf{x}_i^*$ to the histogram bin, $k(\mathbf{x})$ is an isotropic kernel profile and constant C is $C = 1/\Sigma_{i=1}^{n}(||\mathbf{x}_i^*||^2)$.

Similarly, the probability of the feature $u = 1, 2, \ldots, m$ in the target candidate model from the target candidate region centred at position $\mathbf{y}$ is given by

$$\begin{aligned}\hat{p} &= \{\hat{p}_u(\mathbf{y})\}_{u=1,\ldots,m};\\ \hat{p}_u(\mathbf{y}) &= C_h\sum_{i=1}^{n_h} k\left(\left\|\frac{\mathbf{y}-\mathbf{x}_i}{h}\right\|^2\right)\delta[b(\mathbf{x}_i) - u]\end{aligned} \tag{2}$$

where $\hat{p}(\mathbf{y})$ is the target candidate model, $\hat{p}_u(\mathbf{y})$ is the probability of the u^{th} element of $\hat{p}(\mathbf{y})$, $\{\mathbf{x}_i\}_{i=1,\ldots,n_h}$ are pixels in the target candidate region centred at $\mathbf{y}$, h is the bandwidth and C_h is the normalised constant $C_h = 1/\Sigma_{i=1}^{n_h} k(||\mathbf{y} - \mathbf{x}_i/h||^2)$.

2.2 Mean-shift tracking algorithm

A key issue in the mean-shift tracking algorithm is the computation of an offset from the current location $\mathbf{y}$ to a new location $\mathbf{y}_1$ according to the mean-shift iteration equation

$$\mathbf{y}_1 = \frac{\Sigma_{i=1}^{n_h}\mathbf{x}_i w_i g(||(\mathbf{y}-\mathbf{x}_i)/h||^2)}{\Sigma_{i=1}^{n_h} w_i g(||(\mathbf{y}-\mathbf{x}_i/h)||^2)} \tag{3}$$

$$w_i = \sum_{u=1}^{m}\sqrt{\frac{\hat{q}_u}{\hat{p}_u(\mathbf{y})}}\delta[b(\mathbf{x}_i) - u] \tag{4}$$

where $g(\mathbf{x})$ is the shadow of the kernel profile $k(\mathbf{x})$: $g(\mathbf{x}) = -k'(\mathbf{x})$. For the convenience of expression, we denote by $g_i = g(\|\mathbf{y} - \mathbf{x}_i/h\|^2)$. Thus, (3) can be re-written as

$$\mathbf{y}_1 = \sum_{i=1}^{n_h}\mathbf{x}_i w_i g_i \Big/ \sum_{i=1}^{n_h} w_i g_i \tag{5}$$

With (5), the mean-shift tracking algorithm can find the most similar region to the target object in the new frame.

2.3 Background-weighted histogram (BWH)

In target tracking, often the background information is included in the detected target region. If the correlation between target and background is high, the localisation accuracy of the object will be decreased. To reduce the interference of salient background features in target localisation, a representation model of background features was proposed by Comaniciu *et al.* [3] to select discriminative features from the target region and the target candidate region.

In [3], the background is represented as $\{\hat{o}_u\}_{u=1,\ldots,m}$ (with $\Sigma_{i=1}^{m}\hat{o}_u = 1$) and it is calculated by the surrounding area of the target. The background region is three times the size of the target as suggested in [3]. Denote by $\hat{o}^*$ the minimal non-zero value in $\{\hat{o}_u\}_{u=1,\ldots,m}$. The coefficients

$$\{v_u = \min(\hat{o}^*/\hat{o}_u, 1)\}_{u=1,\ldots,m} \tag{6}$$

are used to define a transformation between the representations of target model and target candidate model. The transformation reduces the weights of those features with low v_u, that is, the salient features in the background. Then the new target model is

$$\hat{q}'_u = C'v_u\sum_{i=1}^{n} k(||\mathbf{x}_i^*||^2)\,\delta[b(\mathbf{x}_i^*) - u] \tag{7}$$

with the normalisation constant

$$C' = \frac{1}{\Sigma_{i=1}^{n} k(||\mathbf{x}_i^*||^2)\Sigma_{u=1}^{m} v_u\delta[b(\mathbf{x}_i^*) - u]}$$

The new target candidate model is

$$\hat{p}'_u(\mathbf{y}) = C'_h v_u\sum_{i=1}^{n_h} k\left(\left\|\frac{\mathbf{y}-\mathbf{x}_i}{h}\right\|^2\right)\delta[b(\mathbf{x}_i) - u] \tag{8}$$

where

$$C'_h = \frac{1}{\Sigma_{i=1}^{n_h} k(||(\mathbf{y}-\mathbf{x}_i)/h||^2)\Sigma_{u=1}^{m} v_u\delta[b(\mathbf{x}_i^*) - u]}$$

The above BWH transformation aims to reduce the effects of prominent background features in the target candidate region on the target localisation. In the next section, however, we will prove that BWH cannot achieve this goal because it is equivalent to the usual target representation under the mean-shift tracking framework.

3 CBWH scheme

3.1 Equivalence of BWH representation to usual representation

By the mean-shift iteration formula (5), in the target candidate region the weights of points (referring to (4)) determine the convergence of the tracking algorithm. Only when the weights of prominent features in the background are decreased, the relevance of background information for target localisation can be reduced.

Let us analyse the weight changes by using the BWH transform. Denote by w_i' the weight of point $\mathbf{x}_i$ computed by the BWH in the target candidate region. It can be derived by (4) that

$$w_i' = \sum_{u=1}^{m} \sqrt{\frac{\hat{q}_u'}{\hat{p}_u'(\mathbf{y})}} \delta[b(\mathbf{x}_i) - u] \tag{9}$$

Let u' be the bin index in the feature space which corresponds to point $\mathbf{x}_i$ in the candidate region. We have $\delta[b(\mathbf{x}_i) - u'] = 1$. So (9) can be simplified as

$$w_i' = \sqrt{\hat{q}_{u'}'/\hat{p}_{u'}'(\mathbf{y})} \tag{10}$$

Substitute (7) and (8) into (10), then there is

$$w_i' = \sqrt{\frac{C' v_{u'} \Sigma_{j=1}^{n} k(||\mathbf{x}_j^*||^2)\delta[b(\mathbf{x}_j^*) - u']}{C_h' v_{u'} k(||\mathbf{y} - \mathbf{x}_j/h||^2)\delta[b(\mathbf{x}_j) - u']}}$$

By removing the common factor $v_{u'}$ from the numerator and denominator and substituting the normalisation factors C and C_h into the above equation, we have

$$\begin{aligned} w_i' &= \sqrt{\frac{CC_h}{CC_h} \times \frac{C' \Sigma_{j=1}^{n} k(||\mathbf{x}_j^*||^2)\delta[b(\mathbf{x}_j^*) - u']}{C_h' \Sigma_{j=1}^{n_h} k(||(\mathbf{y} - \mathbf{x}_j)/h||^2)\delta[b(\mathbf{x}_j) - u']}} \\ &= \sqrt{\frac{C'C_h}{CC_h'}} \times \sqrt{\frac{\hat{q}_{u'}}{\hat{p}_{u'}}} = \sqrt{\frac{C'C_h}{CC_h'}} w_i \end{aligned} \tag{11}$$

where w_i calculated by (4) is the weight of point i in the target candidate region using the usual representation of target model and target candidate model.

Equation (11) suggests that w_i' is proportional to w_i. Moreover, by combining mean-shift iteration (5), we have

$$\mathbf{y}_1 = \frac{\Sigma_{j=1}^{n_h} \mathbf{x}_i g_i w_i'}{\Sigma_{j=1}^{n_h} g_i w_i'} = \frac{\Sigma_{j=1}^{n_h} \mathbf{x}_i g_i w_i \sqrt{C'C_h/CC_h'}}{\Sigma_{j=1}^{n_h} w_i g_i \sqrt{C'C_h/CC_h'}} = \frac{\Sigma_{j=1}^{n_h} \mathbf{x}_i g_i w_i}{\Sigma_{j=1}^{n_h} w_i g_i} \tag{12}$$

Equation (12) shows that the mean-shift iteration formula is invariant to the scale transformation of weights. Therefore BWH actually does not enhance mean-shift tracking by transforming the representation of target model and target candidate model. Its result is exactly the same as that without using BWH.

3.2 CBWH algorithm

Although the idea of BWH is good, we see in Section 3.1 that the BWH algorithm does not improve the target localisation. To truly achieve what the BWH wants to achieve, here we propose a new transformation method, namely the corrected BWH (CBWH) algorithm. In CBWH, (6) is employed to transform only the target model but not the target candidate model. That is to say, we reduce the prominent background features only in the target model but not in the target candidate model.

We define a new weight formula

$$w_i'' = \sqrt{\hat{q}_{u'}'/\hat{p}_{u'}(\mathbf{y})} \tag{13}$$

Note that the denominator in the above equation is different from that in (10). Similar to the previous derivation process in Section 3.1, we can easily obtain that

$$w_i'' = \sqrt{C'/C} \times \sqrt{v_{u'}} \times w_i \tag{14}$$

Since $\sqrt{C'/C}$ is a constant scaling factor, it has no influence on the mean-shift tracking process. We can omit it and simplify (14) as

$$w_i'' = \sqrt{v_{u'}} w_i \tag{15}$$

Equation (15) clearly reflects the relationship between the weight calculated by using the usual target representation (i.e. w_i) and the weight calculated by exploiting the background information (i.e. w_i''). If the colour at a point i in the background region is prominent, then the corresponding value of $v_{u'}$ is small. Hence in (15) this point's weight is decreased and its relevance for target localisation is reduced. This will then speed up mean shift's convergence towards the salient features of the target. Note that if we do not use the background information, $v_{u'}$ will be 1 and w_i'' will degrade to w_i with the usual target representation.

Fig. 1 plots the non-zero weights of the features in the first iteration of frame 2 of the benchmark ping-pang ball sequence (refer to Section 4). The weights w_i, w_i' and w_i'' are calculated, respectively, by using the three-target representation methods, that is, the original representation, BWH and CBWH. Fig. 1 clearly shows that w_i' is proportional to w_i with a constant rate ($w_i'/w_i = 0.5919$). Therefore the representation of target model and target candidate model using BWH is the same as the usual

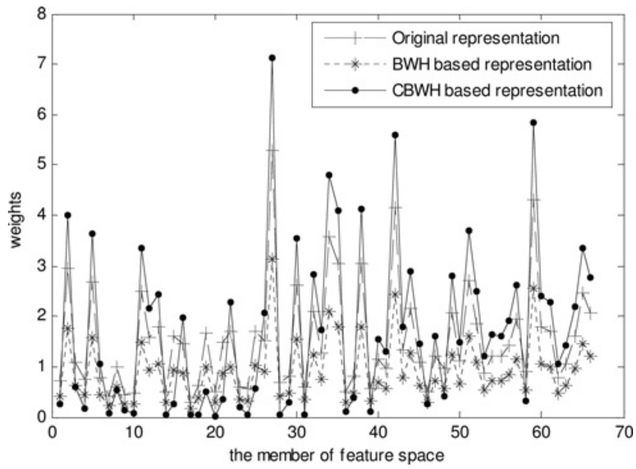

Fig. 1 *Weights of the features in the first mean-shift iteration of frame 2 (the ping-pang ball sequence) using the original representation, BWH and CBWH*

doi: 10.1049/iet-cvi.2009.0075

representation without using background features because the mean-shift iteration is invariant to scale transformation. Meanwhile, w_i'' is different from w_i and w_i'. Some w_i'', for example, of bins 27 and 42, are enhanced while some w_i'', for example, of bins 10 and 20, are weakened. In summary, BWH does not introduce any new information to mean-shift tracking, whereas CBWH exploits truly the background features and can introduce new information for tracking.

3.3 Background model updating in CBWH

In BWH and the proposed CBWH, a background colour model $\{\hat{o}_u\}_{u=1,\dots,m}$ is employed and initialised at the beginning of tracking. However, in the tracking process the background will often change due to the variations in illumination, viewpoint, occlusion and scene content, and so on. If the original background colour model is still used without updating, the tracking accuracy may be reduced because the current background may be very different from the previous background model. Therefore it is necessary to dynamically update the background model for a robust CBWH tracking performance.

Here we propose a simple background model updating method. First, the background features $\{\hat{o}_u'\}_{u=1,\dots,m}$ and $\{v_u'\}_{u=1,\dots,m}$ in the current frame are calculated. Then the Bhattacharyya similarity between $\{\hat{o}_u\}_{u=1,\dots,m}$ and the old background model $\{\hat{o}_u\}_{u=1,\dots,m}$ is computed by

$$\rho = \sum_{u=1}^{m} \sqrt{\hat{o}_u \hat{o}_u'} \tag{16}$$

If ρ is smaller than a threshold, this implies that there are considerable changes in the background, and then we update $\{\hat{o}_u\}_{u=1,\dots,m}$ by $\{\hat{o}_u'\}_{u=1,\dots,m}$ and update $\{v_u\}_{u=1,\dots,m}$ by $\{v_u'\}_{u=1,\dots,m}$. The transformed target model $\hat{q}_u'$ is then computed by (7) using $\{v_u'\}_{u=1,\dots,m}$. Otherwise, we do not update the background model. The proposed CBWH-based mean-shift tracking algorithm is summarised in Fig. 2.

4 Experimental results and discussions

Several representative video sequences are used to evaluate the proposed method in comparison with the original BWH-based mean-shift tracking, which is actually equivalent to the mean-shift tracking with usual target representation. The two algorithms were implemented under the programming environment of MATLAB 7.01. In all the

1) Calculate the target model $\hat{q}$ by Eq. (1) and the background-weighted histogram $\{\hat{o}_u\}_{u=1\cdots m}$, and then compute $\{v_u\}_{u=1\cdots m}$ by Eq. (6) and the transformed target model $\hat{q}'$ by Eq. (7). Initialise the position y_0 of the target candidate region in the previous frame.

2) Let $k \leftarrow 0$.

3) Calculate the target candidate model $\hat{p}(y_0)$ using Eq. (2) in the current frame.

4) Calculate the weights $\{w_i''\}_{i=1\cdots n_h}$ according to Eq. (13).

5) Calculate the new position y_1 of the target candidate region using Eq. (5).

6) Let $d \leftarrow \|y_1 - y_0\|$, $y_0 \leftarrow y_1$, $k \leftarrow k+1$. Set the error threshold ε_1 (default value: 0.1), the maximum iteration number N, and the background model update threshold ε_2 (default value: 0.5).

 If $d<\varepsilon_1$ or $k \ge N$

  Calculate $\{\hat{o}_u'\}_{u=1\cdots m}$ and $\{v_u'\}_{u=1\cdots m}$ based on the tracking result of the current frame. If ρ by Eq. (16) is smaller than ε_2, then $\{\hat{o}_u\}_{u=1\cdots m} \leftarrow \{\hat{o}_u'\}_{u=1\cdots m}$ and $\{v_u\}_{u=1\cdots m} \leftarrow \{v_u'\}_{u=1\cdots m}$, and $\{\hat{q}_u'\}_{u=1\cdots m}$ is updated by Eq. (7).

  Stop iteration and go to step 2 for next frame.

 Otherwise

  Go to step 3.

Fig. 2 *Proposed CBWH-based mean-shift tracking algorithm* experiments, the RGB colour model was used as the feature space and it was quantised into $16 \times 16 \times 16$ bins. Any eligible kernel function $k(\mathbf{x})$, such as the commonly used Epanechnikov kernel and Gaussian kernel, can be used. Our experiments have shown that the two kernels lead to almost the same tracking results. Here, we selected the Epanechnikov kernel as recommended in [3] so that $g(\mathbf{x}) = -k'(\mathbf{x}) = 1$.

To better illustrate the proposed method, in the experiments on the first three sequences we did not update the background feature model in CBWH because there are no obvious background changes, while for the last sequence we updated adaptively the background feature model because there are many background changes such as scene content, illumination and viewpoint variations. Tables 1 and 2 list, respectively, the average numbers of iterations and the target localisation accuracies by the two methods on the four video sequences. To calculate the target localisation accuracy, we manually labelled the target in each frame as ground-truth. The MATLAB codes and all the experimental results of this paper can be found in the web-link http://www.comp.polyu.edu.hk/~cslzhang/CBWH.htm.

Table 1 Average number of iterations by the two methods on the four sequences

Methods	Ping-pang ball sequence	Soccer sequence	Table tennis player sequence	Face sequence
BWH	8.14	3.57	4.25	4.16
CBWH	3.74	3.22	3.46	3.29

Table 2 Target localisation accuracies (mean error and standard deviation)

Sequence	BWH		CBWH	
	Mean error	Standard deviation	Mean error	Standard deviation
ping-pang ball	11.20	20.64	1.94	2.44
soccer	51.12	56.20	4.62	7.65
table tennis player	15.41	15.70	3.89	4.56
face	7.83	10.04	3.65	5.93

The first experiment is on the benchmark ping-pang ball sequence, which was used in [3] to evaluate BWH. This sequence has 52 frames of spatial resolution 352×240. The target is the ball that moves quickly (refer to Fig. 3). In frame 1, we initialised the target model with a region of size 27×31 (inner blue rectangle), which includes many background elements in it. The background model was then initialised to be a region of size 53×61 (external red rectangle excluding the target region), which approximately three times that of the target area. The tracking results in Fig. 3 and the statistics in Table 2 show that the proposed CBWH model (mean error: 1.94; standard deviation: 2.44) has a more accurate localisation accuracy than the original BWH model (mean error: 11.20; standard deviation: 20.64), because the former truly exploits the background information in target localisation. Fig. 4 illustrates the numbers of iterations by the two methods. The average number of iterations is 3.04 for CBWH and 8.14 for BWH. The CBWH method requires less computation. The salient features of target model are enhanced while the background features being suppressed in CBWH so that the mean-shift algorithm can more accurately locate the target.

The second video is a soccer sequence. In this sequence, the colour of sport shirt (green) of the target player is very similar to that of the lawn and thus some target features are presented in the background. Experimental results in Fig. 5 show that the BWH loses the object very quickly, whereas

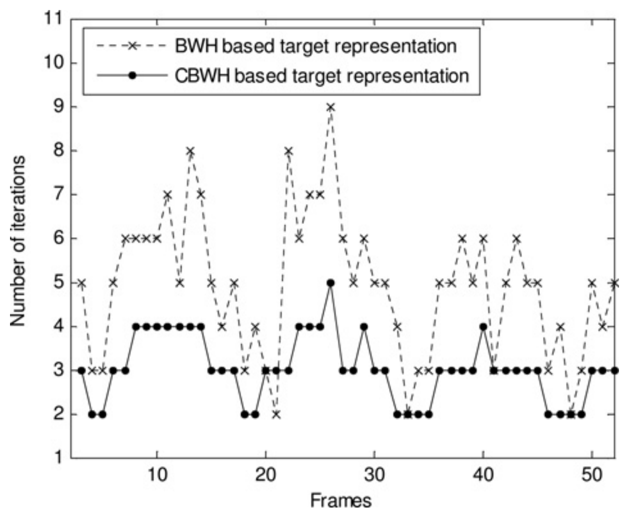

Fig. 4 *Number of iterations on the ping-pang ball sequence*

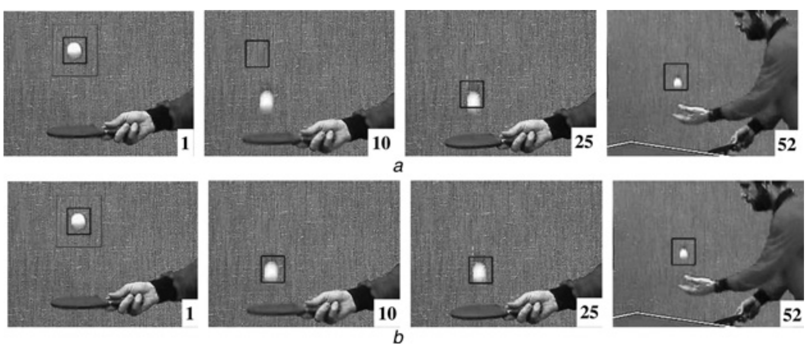

Fig. 3 *Mean-shift tracking results on the ping-pang ball sequence*

Frames 1, 10, 25 and 52 are displayed
a BWH-based mean-shift tracking
b Proposed CBWH-based mean-shift tracking

the proposed CBWH successfully tracks the player over the whole sequence.

The third experiment is on the benchmark sequence of table tennis player. The target to be tracked is the head of the player. We use this sequence to test the robustness of the proposed CBWH to inaccurate target initialisation. Refer to Fig. 6, in the first frame the initial target region (inner blue rectangle) was deliberately set so that it occupies only a small part of the player's head but occupies much background. The initial target model is severely inaccurate and it contains much background information. Fig. 7 compares the Bhattacharyya similarities between the tracking result and its surrounding background region by BWH and CBWH. We see that the Bhattacharyya similarity of CBWH is smaller than that of BWH, which implies that CBWH can better separate the target from background. Regarding the target localisation accuracy, the proposed CBWH-based method has a mean error of 3.89 and standard deviation of 4.56, which are much better than those of the BWH-based method whose mean error and standard deviation are 15.41 and 15.70, respectively.

Since CBWH reduces the impact of features shared by the target and background and enhances the prominent features in the target model, it decreases significantly the relevance of background for target localisation. The experiment in Fig. 6 suggests that the proposed CBWH method is a good candidate in many real-tracking systems, where the initial targets are often detected with about 60% background information inside them. In Fig. 8, we show the tracking results on this sequence by another inaccurate initialisation. The same conclusion can be drawn.

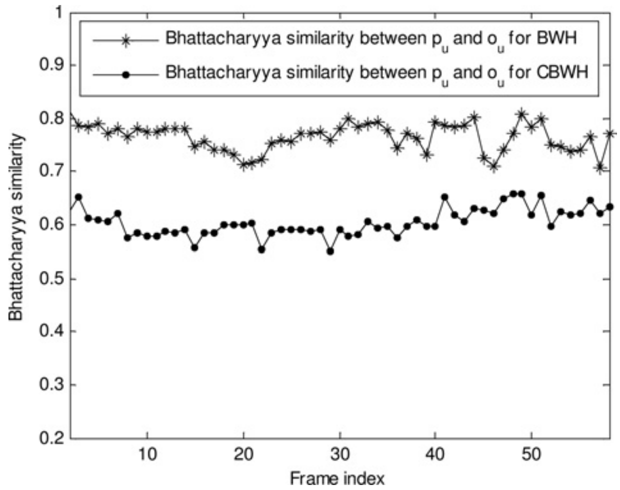

Fig. 7 *Bhattacharyya coefficients between the tracking result and its surrounding background region for the BWH and CBWH methods on the table tennis player sequence*

The last experiment is on a face sequence with obvious changes of background content, illumination and viewpoint. Usually, the background features $\{\hat{o}_u\}_{u=1,\ldots,m}$ are defined by the first frame. However, due to the evolution of video scenes, the background features will change and thus $\{\hat{o}_u\}_{u=1,\ldots,m}$ should be dynamically updated for better performance. Fig. 9 shows the tracking results, respectively, by BWH, CBWH without background update and CBWH with background update. Obviously, CBWH with

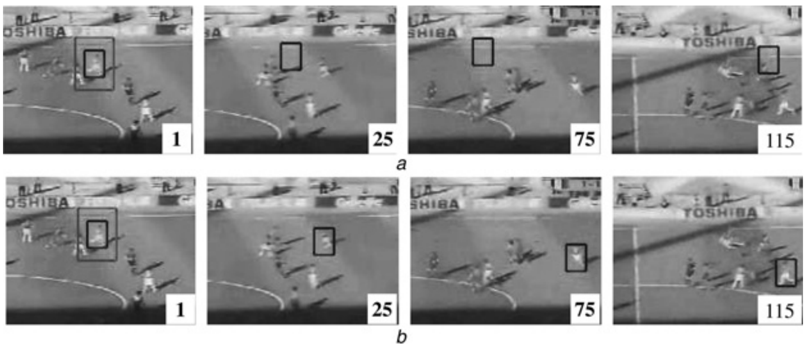

Fig. 5 *Mean-shift tracking results on the soccer sequence*

Frames 1, 25, 75 and 115 are displayed
a BWH-based mean-shift tracking
b Proposed CBWH-based mean-shift tracking

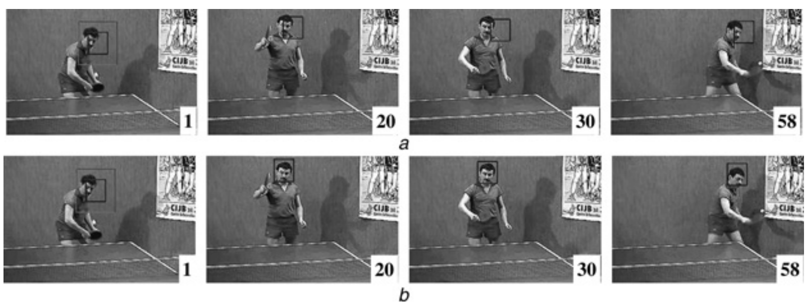

Fig. 6 *Mean-shift tracking results on the table tennis player sequence with inaccurate initialisation*

Frames 1, 20, 30 and 58 are displayed
a BWH-based mean-shift tracking
b Proposed CBWH-based mean-shift tracking

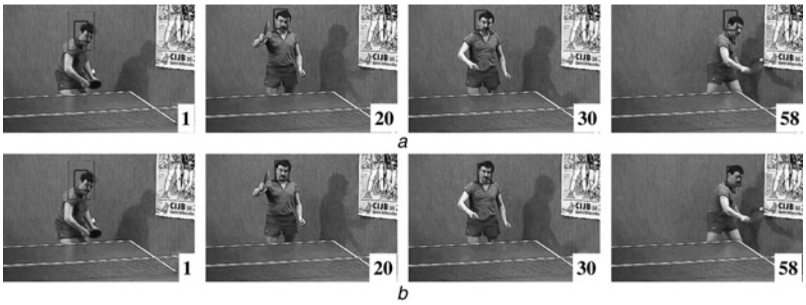

Fig. 8 *Mean-shift tracking results on the table tennis player sequence with another inaccurate initialisation*

Frames 1, 20, 30 and 58 are displayed
a BWH-based mean-shift tracking
b Proposed CBWH-based mean-shift tracking

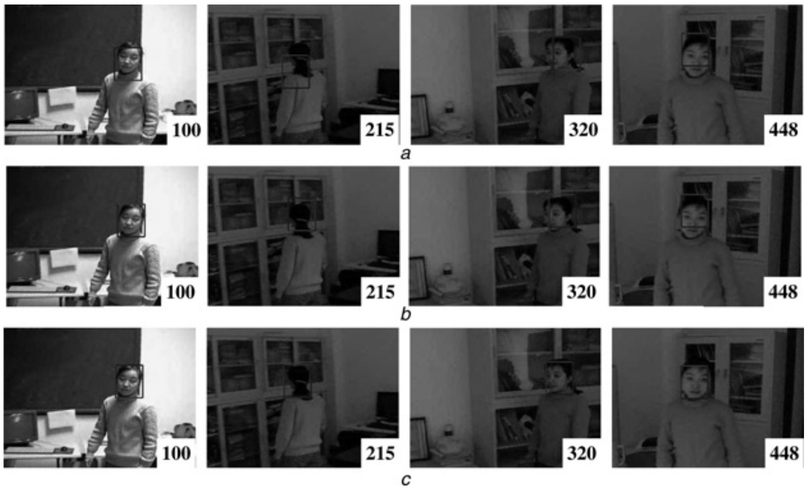

Fig. 9 *Mean-shift tracking results of the face sequence with the proposed CBWH target representation methods*

Frames 100, 215, 320 and 448 are displayed
a BWH-based mean-shift tracking
b Proposed CBWH-based mean-shift tracking without background update
c Proposed CBWH-based mean-shift tracking with background update

background update locates the target much more accurately than the other two methods, whereas BWH performs the worst.

The complexity of CBWH is basically the same as that of the original mean-shift tracking except for transforming the target model with background-weighted histogram. Since the proposed CBWH focuses on tracking the salient features which are different from background, the average number of iterations of it is much less than that of the original BWH. Meanwhile, Table 2 also shows that the proposed CBWH locates the target more reliably and more accurately than BWH. It achieves much smaller mean error and standard deviation than BWH.

5 Conclusions

In this paper, we proved that the BWH representation in [3] is equivalent to the usual target representation so that no new information can be introduced to improve the mean-shift tracking performance. We then proposed a CBWH method to reduce the relevance of background information and improve the target localisation. The proposed CBWH algorithm only transforms the histogram of target model and decreases the probability of target model features that are prominent in the background. The CBWH truly achieves what the BWH wants. The experimental results validated that CBWH can not only reduce the mean-shift iteration number but also improve the tracking accuracy. One of its important advantages is that it reduces the sensitivity of mean-shift tracking to the target initialisation so that CBWH can robustly track the target even if it is not well initialised.

6 Acknowledgment

This work is supported by the Hong Kong Polytechnic University Internal Research Grant (A-SA08), the National Natural Science Foundation of China under Grants 60775020 and 61003151, and the Fundamental Research Funds for the Central Universities under Grant QN2009091.

7 References

1 Yilmaz, A., Javed, O., Shah, M.: 'Object tracking: a survey', *ACM Comput. Surv.*, 2006, **38**, (4), Article 13
2 Comaniciu, D., Ramesh, V., Meer, P.: 'Real-time tracking of non-rigid objects using mean shift'. Proc. IEEE Conf. Computer Vision and Pattern Recognition, Hilton Head, SC, USA, June 2000, pp. 142–149
3 Comaniciu, D., Ramesh, V., Meer, P.: 'Kernel-based object tracking', *IEEE Trans. Pattern Anal. Mach. Intell.*, 2003, **25**, (2), pp. 564–577
4 Bradski, G.: 'Computer vision face tracking for use in a perceptual user interface', *Intel Technol. J.*, 1998, **2**, (Q2), pp. 12–21
5 Fukunaga, F., Hostetler, L.D.: 'The estimation of the gradient of a density function, with applications in pattern recognition', *IEEE Trans. Inf. Theory*, 1975, **21**, (1), pp. 32–40

doi: 10.1049/iet-cvi.2009.0075

6 Cheng, Y.: 'Mean shift, mode seeking, and clustering', *IEEE Trans. Pattern Anal. Mach. Intell.*, 1995, **17**, (8), pp. 790–799

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

8 Wang, J., Thiesson, B., Xu, Y., Cohen, M.F.: 'Image and video segmentation by anisotropic kernel mean shift'. Proc. European Conf. Computer Vision, Prague, Czech Republic, May 2004, vol. 3022, pp. 238–249

9 Paris, S., Durand, F.: 'A topological approach to hierarchical segmentation using mean shift'. Proc. IEEE Conf. Computer Vision and Pattern Recognition, Minnesota, USA, June 2007, pp. 1–8

10 Luo, Q., Khoshgoftaar, T.M.: 'Efficient image segmentation by mean shift clustering and MDL-guided region merging'. IEEE Proc. Int. Conf. Tools with Artificial Intelligence, Florida, USA, November 2004, pp. 337–343

11 Park, J., Lee, G., Park, S.: 'Color image segmentation using adaptive mean shift and statistical model-based methods', *Comput. Math. Appl.*, 2009, **57**, (6), pp. 970–980

12 Collins, R.: 'Mean-shift blob tracking through scale space'. Proc. IEEE Conf. Computer Vision and Pattern Recognition, Wisconsin, USA, June 2003, pp. 234–240

13 Zivkovic, Z., Kröse, B.: 'An EM-like algorithm for color-histogram-based object tracking'. Proc. IEEE Conf. Computer Vision and Pattern Recognition, Washington, DC, USA, July 2004, vol. I, pp. 798–803

14 Yang, C., Ramani, D., Davis, L.: 'Efficient mean-shift tracking via a new similarity measure'. Proc. IEEE Conf. Computer Vision and Pattern Recognition, San Diego, CA, June 2005, vol. I, pp. 176–183

15 Yilmaz, A.: 'Object tracking by asymmetric kernel mean shift with automatic scale and orientation selection'. Proc. IEEE Conf. Computer Vision and pattern Recognition, Minnesota, USA, June 2007, vol. I, pp. 1–6

16 Hu, J., Juan, C., Wang, J.: 'A spatial-color mean-shift object tracking algorithm with scale and orientation estimation', *Pattern Recognit. Lett.*, 2008, **29**, (16), pp. 2165–2173

17 Collins, R.T., Liu, Y., Leordeanu, M.: 'Online selection of discriminative tracking features', *IEEE Trans. Pattern Anal. Mach. Intell.*, 2005, **27**, (10), pp. 1631–1643

18 Tu, J., Tao, H., Huang, T.: 'Online updating appearance generative mixture model for mean shift tracking', *Mach. Vis. Appl.*, 2009, **20**, (3), pp. 163–173

19 Nummiaro, K., Koller-Meier, E., Gool, L.V.: 'An adaptive color-based particle filter', *Image Vis. Comput.*, 2003, **21**, (1), pp. 99–110

20 Jeyakar, J., Babu, R., Ramakrishnan, K.R.: 'Robust object tracking with background-weighted local kernels', *Comput. Vis. Image Underst.*, 2009, **112**, (3), pp. 296–309

21 Li, L., Feng, Z.: 'An efficient object tracking method based on adaptive nonparametric approach', *Opto-Electron. Rev.*, 2005, **13**, (4), pp. 325–330

22 Allen, J., Xu, R., Jin, J.: 'Mean shift object tracking for a SIMD computer'. Proc. Int. Conf. Information Technology and Applications, Sydney, Australia, July 2005, vol. I, pp. 692–697