I.J. Image, Graphics and Signal Processing, 2012, 12, 34-42
Published Online November 2012 in MECS (http://www.mecs-press.org/)
DOI: 10.5815/ijigsp.2012.12.05

A More Robust Mean Shift Tracker on Joint Color-CLTP Histogram

Pu Xiaorong, Zhou Zhihu
School of Computer Science and Engineering, University of Electronic Science and Technology of China
Chengdu, China
E-mail: puxiaor@uestc.edu.cn

Abstract— A more robust mean shift tracker using the joint of color and Completed Local Ternary Pattern (CLTP) histogram is proposed. CLTP is a generalization of Local Binary Pattern (LBP) which can be applied to obtain texture features that are more discriminant and less sensitive to noise. The joint of color and CLTP histogram based target representation can exploit the target structural information efficiently. To reduce the interference of background in target localization, a corrected background-weighted histogram and background update mechanism are adapted to decrease the weights of both prominent background color and texture features similar to the target object. Comparative experimental results on various challenging videos demonstrate that the proposed tracker performs favorably against several variants of state-of-the-art mean shift tracker when heavy occlusions and complex background changes exist.

Index Terms— MeanShift, Object Tracking, Completed Local Ternary Pattern, Joint Color-CLTP Histogram

I. INTRODUCTION

For object tracking within the field of computer vision， many efficient algorithms[1] have been proposed in recent years，mostly on the issues of changing appearance patterns of both the object and the scene, nonrigid object structures, object-to-object and object-to-scene occlusions. Among various object tracking methods, the mean shift tracker has recently attracted many researchers' attention [2] [3] [4] [5] [6] [7] [8][9]due to its simplicity and efficiency.

Mean shift tracking algorithm has been proved to be robust to scale, rotation, partial occlusion [2] [3] by using the color histogram to represent target object. However, it is inclined to fail when some of the target features present in the background. Comaniciu et al. [3] proposed a background-weighted histogram (BWH) to decrease background interference in target representation. However, Ning et al [4] demonstrated that the BWH-based mean shift tracker is equivalent to the conventional mean shift tracking method [2], and then a corrected background-weighted histogram (CBWH) is proposed to actually reduce the interference of background in target localization. Additionally, only utilizing color histogram to model the target object in the mean shift algorithm has the disadvantage that the spatial information of the target object is lost. Therefore, many other features, such as edge features [10], Local Binary Pattern (LBP) texture features [5], have been used in combination with color. Among these feature-combined mean shift trackers, the joint color-texture histogram proposed by Ning et al [5] achieves better tracking performance with fewer mean shift iterations and higher robustness.

Local Binary Pattern (LBP), firstly introduced by Ojala et al [11], is a simple yet efficient operator to describe local image pattern. It has great success in computer vision and pattern recognition, such as face recognition [12], texture classification [13] [14], unsupervised texture segmentation [15], dynamic texture recognition [16]. However, it still remains some issues to be further investigated. Tan and Triggs [17] proposed Local Ternary Patterns (LTP) that quantizes the difference between a pixel and its neighbours into three levels, which is less sensitive to noise in near-uniform image regions such as cheeks and foreheads. Guo et al [18] proposed Completed Local Binary Pattern (CLBP) where a local region is represented by its center pixel and a local difference sign-magnitude transform. CLBP is able to preserve an important kind of difference magnitude information for texture classification.

In this paper, a more robust mean shift tracker is discussed. We propose a novel texture descriptor, named Completed Local Ternary Pattern (CLTP), which is more discriminant and less sensitive to noise. The CLTP is firstly applied to represent the target texture features, and then combined with color to form a distinctive and effective target representation called joint color-CLTP histogram. In order to reduce the interference of background in target localization, corrected background-weighted histogram and background update mechanism [4] is adapted to decrease the weights of both prominent background color and texture features similar to the target. The proposed tracker is more robust especially in case of heavy occlusions and complex background changes than several state-of-art variants of mean shift tracker.

The rest of the paper is organized as follows. Section II briefly reviews conventional mean shift algorithm. Section III discusses Local Binary Pattern and

Completed Local Ternary Pattern. Section IV investigates mean shift tracker on joint color-CLTP texture histogram. Section V gives the comparative experimental results of the proposed tracker compared with several state-of-art mean shift trackers. Section VI concludes this paper.

II. CONVENTIONAL MEAN SHIFT ALGORITHM

A. *Target Representation*

In this section, target representation using the color histogram in the conventional mean shift tracker [2] is reviewed. Assume that the target object being tracked is defined by a rectangle or an ellipsoidal region in a frame of video.

Target model $\hat{q}$ of object being tracked can be obtained as

$$\begin{cases} \hat{q} = \{\hat{q}_u\}_{u=1\cdots m} \\ \hat{q}_u = C\sum_{i=1}^{n} k\left(\left\|x_i^*\right\|^2\right)\delta\left[b(x_i^*) - u\right] \end{cases} \tag{1}$$

where $\hat{q}_u$ represents the probabilities of feature u in target model $\hat{q}$, m is the number of feature spaces, δ is the Kronecker delta function, $\left\{x_i^*\right\}_{i=1\cdots n}$ is the normalized pixel positions in the target region centered at original position, $b(x_i^*)$ maps the pixel x_i^* to the histogram bin, $k(x)$ is an isotropic kernel profile and constant C is a normalization function defined by

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

Analogously, the target candidate model $\hat{p}_y$ of the candidate region can be computed as:

$$\begin{cases} \hat{p}(y) = \left\{\hat{p}_u(y)\right\}_{u=1\cdots m} \\ \hat{p}_u(y) = C_h \sum_{i=1}^{n_h} k\left(\left\|\frac{y - x_i}{h}\right\|^2\right)\delta\left[b(x_i) - u\right] \end{cases} \tag{3}$$

$$C_h = \frac{1}{\sum_{i=1}^{n} k\left(\left\|\frac{y - x_i}{h}\right\|^2\right)} \tag{4}$$

where $\hat{p}_u(y)$ represent the probabilities of feature u in target model $\hat{p}(y)$, $\left\{x_i^*\right\}_{i=1\cdots n_h}$ is the pixel positions in the target candidate region centered at y , h is the bandwidth and C_h constant is a normalization function.

Bhattacharyya coefficient is used to calculate the likelihood of the target model and the candidate model as:

$$\rho\left[\hat{p}(y), \hat{q}\right] = \sum_{u=1}^{m} \sqrt{\hat{p}(y)_u, \hat{q}_u} \tag{5}$$

The distance between the target model and the candidate model is defined as:

$$d\left[\hat{p}(y), \hat{q}\right] = \sqrt{1 - p\left[\hat{p}(y), \hat{q}\right]} \tag{6}$$

B. *Mean Shift Tracking*

To minimize the distance defined by Equation (6) is to maximize Equation (5), Taylor expansion is used to linearly approximate Bhattacharyya coefficient (5) as follows:

$$\rho\left[\hat{p}(y), \hat{q}\right] \approx \frac{1}{2}\sum_{u=1}^{m} \hat{p}_u(y_0)\hat{q}_u + \frac{1}{2}C_h \sum_{i=1}^{n_h} w_i k\left(\left\|\frac{y - x_i}{h}\right\|^2\right) \tag{7}$$

where y_0 is the target location in previous frame, and

$$\omega_i = \sum_{u=1}^{m} \sqrt{\frac{\hat{q}_u}{\hat{p}_u(y_0)}}\delta\left[b(x_i) - u\right] \tag{8}$$

It can be seen that the first term in (7) is independent of y , hence maximizing the second term in (7) can achieve the goal of minimizing the distance in (6). In the iterative optimization process, the new position moving from y to y_1 is defined as:

$$y_1 = \frac{\sum_{i=1}^{n_h} x_i \omega_i g\left(\left\|\frac{y - x_i}{h}\right\|^2\right)}{\sum_{i=1}^{n_h} \omega_i g\left(\left\|\frac{y - x_i}{h}\right\|^2\right)} \tag{9}$$

For simplicity, the Epanechnikov profile is chosen, then (9) is reduced to

$$y_1 = \frac{\sum_{i=1}^{n_h} x_i \omega_i}{\sum_{i=1}^{n_h} \omega_i} \tag{10}$$

from (10), iteratively obtained new position will converge to a fixed position that is the most similar region to the target.

III. LOCAL BIANRY PATTERN AND COMPLETED LOCAL TENARY PATTERN

A. Brief Review of Local Benary Pattern and Its Variants

The LBP [14] operator labels the pixel in an image by thresholding its neighborhood with the center value and considers the result as a binary number (binary pattern). The general version of the LBP operator is defined as:

$$LBP_{P,R}(x_c, y_c) = \sum_{p=0}^{P-1} s(g_p - g_c)2^p \tag{11}$$

$$s(x) = \begin{cases} 1 & x \geq 0 \\ 0 & x < 0 \end{cases} \tag{12}$$

where g_c is the gray value of the central pixel, g_p is the value of its neighbors, P is the total number of involved neighbors and R is the radius of the neighborhood.

The “uniform” LBP pattern, denoted by $LBP_{P,R}^{u2}$, is defined as the number of bitwise $0/1$ changes in that pattern:

$$U(LBP_{P,R}) = \left|s(g_{p-1} - g_c) - s(g_0 - g_c)\right| \\ + \sum_{p=1}^{P-1} \left|s(g_p - g_c) - s(g_{p-1} - g_c)\right| \tag{13}$$

It has $\mathrm{P}*(\mathrm{P}-1)+3$ distinct output values. Fig.1 shows all uniform patterns for P=8.

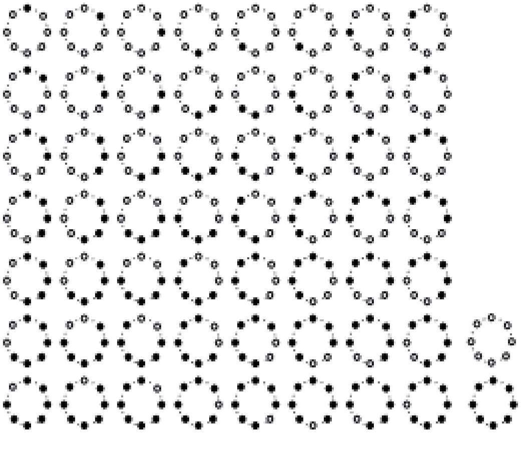

Figure 1.Uniform LBP patterns when P=8. The black and white dots represent the bit values of 1 and 0 in the 8-bit output of the LBP operator

Further, a locally rotation invariant“uniform” pattern is defined as:

$$LBP_{P,R}^{riu2} = \begin{cases} \sum_{p=0}^{P-1} s(g_p - g_c) & if U(LBP_{P,R}) \leq 2 \\ P+1 & otherwise \end{cases} \tag{14}$$

which has $\mathrm{P}+2$ distinct output values.

Recent years, many variants of LBP have emerged, such as Local Tenary Patterns (LTP) [17], Completed LBP [18], Dominant LBP [19], MONOGENIC-LBP [20], Multimodal LBP [21], Sobel-LBP [22]. Among numerous variants of LBP, CLBP and LTP are outstanding because of its significantly improved texture classification accuracy. CLBP has three parts: CLBP_Center (CLBP_C), CLBP-Sign (CLBP_S) and CLBP_Magnitude (CLBP_M). CLBP_C is formed by converting the center pixel into a binary code after global thresholding, CLBP_S is obtained by converting the local difference signs into an 8-bit binary code, and CLBP_M is gained by coding the difference magnitudes as an 8-bit binary code. Experimentally, the CLBP could achieve much better rotation invariant texture classification results than conventional LBP based schemes. More details of CLBP can be referred in [18]. LTP quantizes the difference between a pixel and its neighbours into three levels, and then each ternary pattern is split into its positive and negative halves. More details of LTP can be seen in [17].

B. Completed Local Tenary Patterns

Given a central pixel g_c, its P circularly and evenly spaced neighbors g_p ($p = [0, 1, \cdots, P-1]$), and its radius of the neighborhood R . The local difference between g_c and g_p is denoted by $d_p = g_p - g_c$, which can be viewed as two components: $d_p = s_p * m_p$, where $s_p = \begin{cases} 1, d_p \geq 0 \\ -1, d_p < 0 \end{cases}$ and $m_p = \left|d_p\right|$. To construct Completed Local Ternary Patterns, s_p is rewritten as:

$$s_p^{'} = \begin{cases} 1\ , d_p \geq t \\ 0\ , \left|d_p\right| < t \\ -1\ , d_p \leq -t \end{cases} \tag{15}$$

As ref. [17], a coding scheme that splits each ternary pattern into its positive and negative halves is applied. Thus, CLTP operator consists of three operators: CLTP_P (CLTP Positive operator), CLTP_N (CLTP Negative operator), and CLTP_M (CLTP Magnitude operator). They are defined as follows, respectively:

$$CLTP\_M_{P,R} = \sum_{p=0}^{P-1} s_1(m_p, \kappa)2^p, s_1(x, \kappa) = \begin{cases} 1, & x \geq \kappa \\ 0, otherwise \end{cases} \tag{16}$$

$$CLTP\_P_{P,R} = \sum_{p=0}^{P-1} s_2(d_p, t)2^p, s_2(x, t) = \begin{cases} 1, & x \geq t \\ 0, otherwise \end{cases} \tag{17}$$

$$CLTP\_N_{P,R} = \sum_{p=0}^{P-1} s_3(d_p, t)2^p, s_3(x, t) = \begin{cases} 1, & x \leq -t \\ 0, otherwise \end{cases} \tag{18}$$

where t is a threshold value, and κ is an adaptively determined threshold as the mean value of m_p.

Fig.2 shows an example of the CLTP. Fig.2(a) is the original 3×3 local sample block with central pixel being 25. The difference vector is $[4,38,24,-5,-13,22,-11,50]$, as shown in Fig.2(b). Absolute value of the difference vector is the local magnitude vector $[4,\ 38,24,5,13,22,11,50]$, as shown in Fig.2(c). By equation (15) with $t=5$, Ternary code vector $[0,1,1,0,-1,1,-1,1]$ (Fig.2(d)) is obtained, and then each ternary pattern is split into positive halves named CLTP Positive code of 01100101 (Fig.2(f)) and negative halves named CLTP Negative code of 00001010 (Fig.2(g)). Finally, CLTP Magnitude code of 01100101 (Fig.2(e)) can be obtained by equation(16).

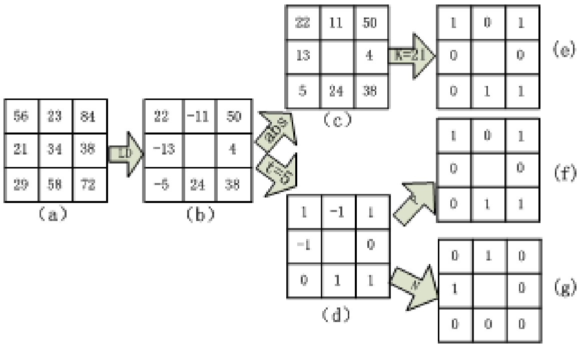

Figure 2.CLTP(P=8,R=1): (a) a 3×3 sample block ;(b) the local differences; (c) magnitude components;(d) ternary code; (e) CLTP Magnitude code;(f) CLTP Positive code (g) CLTP Negative code. LD means local difference operator, abs represents absolute operator, P and N means positive operator and negative operator respectively, threshold of CLTP Magnitude ($\kappa=21$) is the mean value of magnitude components m_p ($p=[0,1,\cdots,7]$) and set the threshold as $t=5$.

Similar to the definitions of $LBP_{P,R}^{riu2}$ and $CLTP\_M_{P,R}^{riu2}$, both $CLTP\_P_{P,R}^{riu2}$ and $CLTP\_N_{P,R}^{riu2}$ can also be defined to achieve rotation invariant classification.

IV. Mean Shift Tracking with Joint of Color and CLTP Texture Histogram

A. Target Representation Using Joint of Color and CLTP Texture Histogram

To model the target efficiently, rotation invariant CLTP patterns ($CLTP\_M_{P,R}^{riu2}$, $CLTP\_P_{P,R}^{riu2}$ and $CLTP\_N_{P,R}^{riu2}$) are firstly obtained. Then joint RGB-CLTP_M histogram, joint RGB-CLTP_P histogram and joint RGB-CLTP_N histogram are reaped by rewritten equation (1):

$$\begin{cases} \hat{q}_w = \left\{\hat{q}_{u_w}\right\}_{u_w=1\cdots m} \\ \hat{q}_{u_w} = C\sum_{i=1}^{n} k\left(\left\|x_i^*\right\|^2\right)\delta\left[b_w(x_i^*)-u_w\right] \end{cases} \tag{19}$$

where $w=1,2,3$. $\hat{q}_{u_w}$ represents the probabilities of feature u_w in target model $\hat{q}_w$ (i.e. $\hat{q}_1$, $\hat{q}_2$ and $\hat{q}_3$ corresponds to joint RGB-CLTP_M histogram, joint RGB-CLTP_P histogram, joint RGB-CLTP_N histogram respectively). $b_w(x_i^*)$ maps the pixel x_i^* to the corresponding histogram bin, $m=N_R\times N_G\times N_B\times N_p$ is the number of joint feature spaces , $N_R\times N_G\times N_B$ represents the quantized bins of color channels and N_p represents the bins of the CLTP texture features.

Similarly, the target candidate model $\hat{p}_w(y)$ corresponding to the candidate region is given by:

$$\begin{cases} \hat{p}_w(y) = \left\{\hat{p}_{u_w}(y)\right\}_{u_w=1\cdots m} \\ \hat{p}_{u_w}(y) = C_h\sum_{i=1}^{n_h} k\left(\left\|\frac{y-x_i}{h}\right\|^2\right)\delta\left[b_w(x_i)-u_w\right] \end{cases} \tag{20}$$

B. Corrected Background-Weighted Histogram and Background Model Updating Mechanism

The idea of corrected background-weighted histogram and background model updating mechanism in [4] is adapted to reduce the interference of background that has color and texture features similar to the tracked target. The background is represented as $\left\{\hat{o}_{u_w}\right\}_{u_w=1\cdots m}$ with ($\sum_{i=1}^{m}\hat{o}_{u_w}=1, w=1,2,3$) and it is calculated by the surrounding area of the target. Transformation between the representations of target model and target candidate model is defined by:

$$\left\{v_{u_w}=\min\left(\hat{o}_w^*/\hat{o}_{u_w},1\right)\right\} \tag{21}$$

where $\hat{o}_w^*$ is the minimal non-zero value in $\left\{\hat{o}_{u_w}\right\}_{u_w=1\cdots m}$. Hence the target model is defined as:

$$\begin{cases} \hat{q}_w^{'} = \left\{\hat{q}_{u_w}^{'}\right\}_{u_w=1\cdots m} \\ \hat{q}_{u_w}^{'} = C_w^{'}v_{u_w}\sum_{i=1}^{n} k\left(\left\|x_i^*\right\|^2\right)\delta\left[b_w(x_i^*)-u_w\right] \end{cases} \tag{22}$$

where $C_w^{'} = \dfrac{1}{\sum_{i=1}^{n} k\left(\left\|x_i^*\right\|^2\right) \sum_{u_w=1}^{m} v_{u_w} \delta\left[b_w(x_i^*) - u_w\right]}$

and then a new weight formula is computed as:

$$\omega_{w,i}^{'} = \sum_{u_w=1}^{m} \sqrt{\hat{q}_{u_w}^{'} / \hat{p}_{u_w}(y)}\,\delta\left[b_w(x_i^*) - u_w\right] \qquad (23)$$

In the tracking process, the background will often change due to the variations of illumination, viewpoint, occlusion and scene content etc. Therefore, it is necessary to dynamically update the background model for robust tracking. Here, a simple background model updating mechanism is proposed. Assume that we have acquired the background features $\left\{\hat{o}_{u_w}\right\}_{u_w=1\cdots m}$ and $\left\{v_{u_w}\right\}_{u_w=1\cdots m}$ in the previous frame. The background features $\left\{\hat{o}_{u_w}^{'}\right\}_{u_w=1\cdots m}$ and $\left\{v_{u_w}^{'}\right\}_{u_w=1\cdots m}$ are firstly obtained in the current frame. Then Bhattacharyya similarity between $\left\{\hat{o}_{u_w}^{'}\right\}_{u_w=1\cdots m}$ and the old background model $\left\{\hat{o}_{u_w}\right\}_{u_w=1\cdots m}$ is computed by :

$$\rho_w = \sum_{u_w=1}^{m} \sqrt{\hat{o}_{u_w} \hat{o}_{u_w}^{'}} \qquad (24)$$

$$\rho = (\lambda_1 \rho_1 + \lambda_2(\alpha_1 \rho_2 + \alpha_2 \rho_3)) \qquad (25)$$

If ρ is smaller than a threshold ξ_2 when there are considerable changes in the background, $\left\{\hat{o}_{u_w}^{'}\right\}_{u_w=1\cdots m}$ should be updated by $\left\{\hat{o}_{u_w}\right\}_{u_w=1\cdots m}$ and $\left\{v_{u_w}^{'}\right\}_{u_w=1\cdots m}$ be replaced by $\left\{v_{u_w}\right\}_{u_w=1\cdots m}$.

C. Tracking Using Joint of Color and CLTP Texture Histogram

In the iterative process of tracking, the estimated target moves from y to a new position y_{new} is defined as:

$$y_w = \frac{\sum_{i=1}^{n_h} x_i \omega_{w,i}^{'}}{\sum_{i=1}^{n_h} \omega_{w,i}^{'}} \qquad (26)$$

$$y_{new} = (\lambda_1 y_1 + \lambda_2(\alpha_1 y_2 + \alpha_2 y_3)) \qquad (27)$$

where the parameters λ_1, λ_2 , α_1 and α_2 in equation (27) are set by the same way as in equation(23).

By equation (27), the proposed approach can explore the target object in the new frame accurately and robustly. The scheme of the approach can be designed as:

1) Initialize the position y_0 of the target candidate region and set the iteration number $k = 0$, the maximum iteration number $N = 15$, the position error threshold ξ_1 (default value: 0.1) and the background model update threshold ξ_2 (default value: 0.5).
2) Calculate the target model by (19) and the background-weighted histogram $\left\{\hat{o}_{u_w}\right\}_{u_w=1\cdots m}$, then compute $\left\{v_{u_w}\right\}_{u_w=1\cdots m}$ by (21) and transformed target model $\hat{q}_w^{'}$ by (22) in the previous frame.
3) In the current frame, calculate the distribution of the target candidate model $\hat{p}_w(y_0)$.
4) Calculate the weights $\omega_{w,i}^{'}$ via (23).
5) Calculate the new position y_{new} of the target candidate region using (26) and (27).
6) Set $d \leftarrow \left\|y_{new} - y_0\right\|$, $y_0 \leftarrow y_{new}$, $k \leftarrow k+1$.

   If $d < \xi_1$ or $k \geq N$

   Calculate $\left\{\hat{o}_{u_w}^{'}\right\}_{u_w=1\cdots m}$ and $\left\{v_{u_w}^{'}\right\}_{u_w=1\cdots m}$ of tracked object in the current frame, and then obtain ρ based on (24) and (25). If ρ is smaller than ξ_2 , then update $\left\{\hat{o}_{u_w}\right\}_{u_w=1\cdots m} \leftarrow \left\{\hat{o}_{u_w}^{'}\right\}_{u_w=1\cdots m}$, $\left\{v_{u_w}\right\}_{u_w=1\cdots m} \leftarrow \left\{v_{u_w}^{'}\right\}_{u_w=1\cdots m}$, and update $\hat{q}_w^{'}$ by (22).

   Stop and go to Step 7.

   else

   go to step(3)
7) Set the next frame as the current frame with initial location y_0 and go to Step 1.

V. COMPARATIVE EXPERIMENTS OF STATE-OF-ART MEANSHIFT ALGORITHMS

To evaluate the effectiveness of the proposed method, we carried out a series of experiments on five challenging public video sequences, Table Tennis playing video sequences from [4], Camera sequences from PETS2001 database [23], Bird2 video sequences from [24], Walking Woman video sequences from [25], and Panda video sequences from [26]. We make comparison with three state-of-the-art mean shift trackers, mean shift tracker with background-weighted histogram (MS_BWH) [3], mean shift tracker with corrected background-weighted histogram (MS_CBWH) [4] and mean shift tracker using joint color-LBP texture histogram (MS_RGBLBP) [5]. For comparative purpose, the tracked object of MS_BWH, MS_CBWH, MS_RGBLBP and the proposed approach are represented by white, yellow, red and green rectangle box, respectively.

In the proposed approach and MS_RGBLBP, we set $N_R \times N_G \times N_B \times N_p = 8 \times 8 \times 8 \times 10$. In MS_BWH and MS_CBWH, we set $N_R \times N_G \times N_B = 8 \times 8 \times 8$. The incremental size of search region is 5 pixels and threshold t in CLTP is set manually to 5. The background region is three times the size of the target. Since Guo et al [13] mathematically demonstrated that the local difference reconstruction errors made by using s_p and m_p is $E_s = \chi^2$, and $E_m = 4\chi^2$, we set $\lambda_1 = 0.2$, $\lambda_2 = 0.8, \alpha_1 = \alpha_2 = 0.5$.

A. Comparative results on Table Tennis sequence (simple background, no occlusions).

The target to be tracked is the moving head of the player. Tracking results of the proposed approach are shown in Fig.3. All trackers can locate the target accurately, since the color differences between the target and the background are distinctive and the background has simple structural information.

The first row of Table 1 lists the total and average numbers of mean shift iterations among the four methods. It is clear that the proposed method achieves better result than the other trackers.

Fig.4 shows the comparative numbers of iterations in every frame. It can be seen that the proposed approach needs less iterations than MS_BWH and MS_RGBLBP in each frame at most case and performs as good as MS_CBWH.

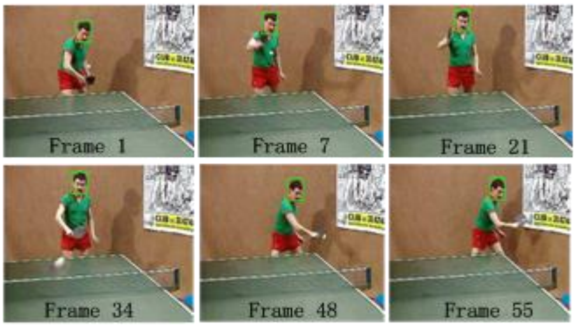

Figure 3. Tracking results of the proposed method on tennis playing video sequence. Frame1, 7, 21, 34, 48 and 55 are displayed.

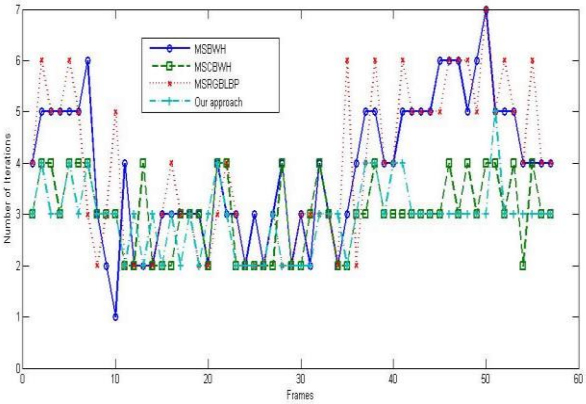

Figure 4. Number of iterations on the Table tennis sequence

B. Comparative results on Camera sequence (background change, no occlusions).

The target to be tracked is the walking man with the small blurred object of interest and background change. Tracking results of the four methods are shown in Fig.5. All the methods can accurately locate the tracking object among the first 150 frames. However, neither MS_BWH nor MS_RGBLBP can locate the target object while both MS_CBWH and the proposed approach can still track the target accurately, when the background changes from grass to road.

The second row of Table 1 lists the total and average numbers of mean shift iterations of the four methods, which indicates that the proposed approach needs less number of iterations than MS_BWH and MS_RGBLBP, and just only performs a little bit worse than MS_CBWH.

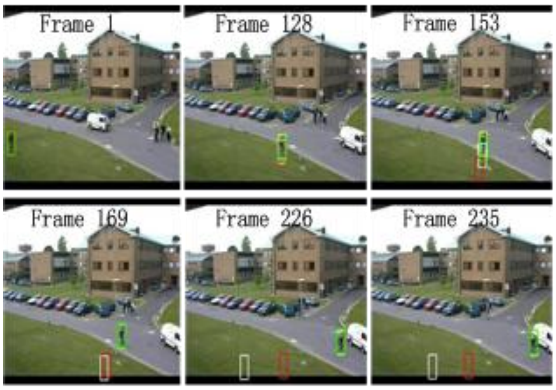

Figure 5. Tracking results of the four methods on Camera video sequence. Frame 1, 128, 153, 169, 226 and 235 are displayed.

C. Comparative results on Bird2 video sequence (*no background change, heavy occlusions*).

The target to be tracked is the flying bird with heavy occlusions. Fig.6 shows the tracking results of the four methods. It can be experimentally observed that MS_BWH and MS_CBWH drift away from the target into background regions but MS_CBWH and the proposed approach can still locate the target when heavy occlusion emerges. Moreover, the proposed approach is more stable and accurate than MS_CBWH when occlusion appears in frame 46 and 64.

The third row of Table 1 gives the numbers of iterations of the proposed approach and MS_CBWH, which indicates that the proposed approach performs better than MS_CBWH.

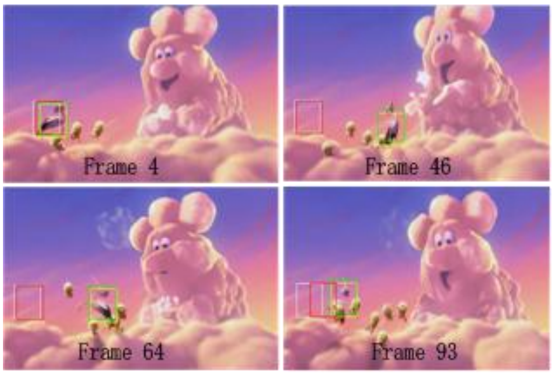

Figure 6. Tracking results of the four methods on Bird2 video sequence. Frame 4, 16, 46 and 93 are displayed.

D. *Comparative results on Walking Woman video sequences* (*complex background changes, heavy occlusions*)

The target to be tracked is the walking woman with heavy occlusions and complex background changes. Tracking results of the four methods are shown in Fig.7. Because the background is changing and heavy occlusion exists during tracking process, only the proposed approach can still track the target, while MS_BWH, MS_ CBWH and MS_RGBLBP all lose the target. This suggests that the joint color-CLTP histogram is more discriminant and can more efficiently exploit the target structural information than color histogram and joint color-LBP histogram.

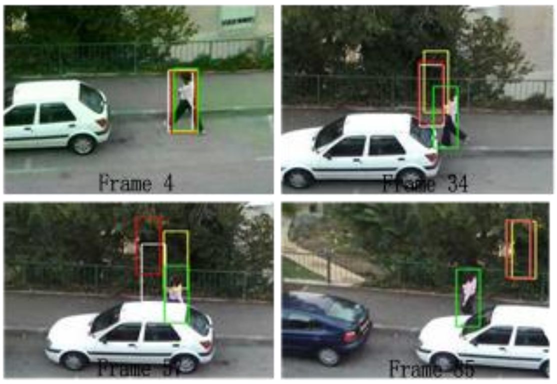

Figure 7.Tracking results of the four methods on Walking Woman video sequence. Frame 4, 34, 57 and 85 are displayed.

TABLE 1. THE NUMBERS OF ITERATION BY THE FOUR METHODS

Video Sequence	Frames	Method	Total Number of Iterations	Average Number of Iterations
Table Tennis	**58**	**MS_BWH**	220	3.86
		MS_CBWH	172	3.02
		MS_RGBLBP	222	3.89
		Our approach	167	**2.93**
Camera	**145**	**MS_BWH**	492	3.42
		MS_CBWH	428	**2.97**
		MS_RGBLBP	509	3.53
		Our approach	454	3.15
Bird2	**99**	**MS_BWH**	—	—
		MS_CBWH	357	3.64
		MS_RGBLBP	—	—
		Our approach	350	**3.57**

E. *Comparative results on Panda video sequences (background changes, heavy occlusions, large in-plane rotations and fast movement)*

The target to be tracked is the cartoon Panda with heavy occlusions, background changes, large in-plane rotations and fast movement. Tracking results of the four methods are shown in Fig.8. Because of the rotation invariant target model and fast convergence of meanshift, all the methods can handle large in-plane rotations and fast movement when there is no occlusion. However,as seen from frame 196,208 and 212, the other three trackers drift away after the target undergoes heavy occlusions and background changes at the same time whereas our proposed method performs well throughout this sequence. This also suggests that the joint color-CLTP histogram is more powerful than color histogram and joint color-LBP histogram.

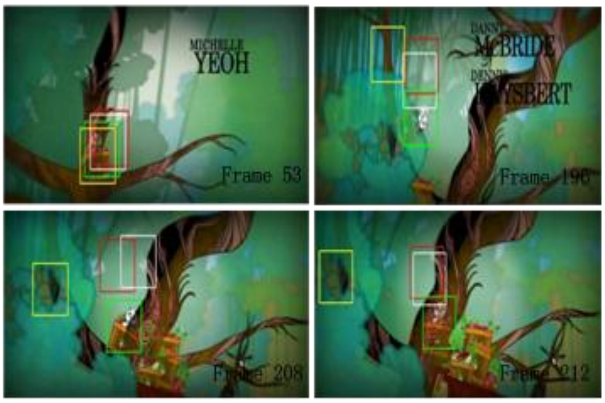

Figure 8.Tracking results of the four methods on Panda cartoon video sequence. Frame 53, 196, 208 and 212 are displayed.

VI. Conclusion

In this paper, a more robust mean shift tracker based on a more distinctive and effective target model is proposed. First, a novel texture descriptor, Completed Local Ternary Pattern (CLTP), is proposed to represent the target structural information, which is more discriminant and less sensitive to noise. Then, a new target representation, joint color-CLTP histogram, is presented to effectively distinguish the foreground target and its background. Finally, the corrected background-weighted histogram and background updating mechanism is adapted to reduce the interference of background that has color and texture features similar to the target object. Numerous experimental results and evaluations demonstrate the proposed tracker performs favorably against existing variants of state-of-the-art mean shift tracker.

ACKNOWLEDGMENT

The authors thanks for the support of National Science Foundation of China: "Vision Cognition and Intelligent Computation on Facial Expression" under Grant 60973070.

REFERENCES

[1] A. Yilmaz , O. Javed , and M. Shah, "Object Tracking: a Survey, " *ACM Computing Surveys*, 38, (4), Article 13, 2006.

[2] D. Comaniciu, V. Ramesh, and P. Meer, "Real-Time Tracking of Non-Rigid Objects Using Mean Shift, " *Proc. IEEE Conf. Computer Vision a nd Pattern Recognition*, pp. 142-149 , 2000.

[3] D. Comaniciu, V. Ramesh, and P. Meer, "Kernel-Based Object Tracking," *IEEE Trans. Pattern . Anal. Machine Intell.*, 25, (2), pp. 564-577, 2003.

[4] J. Ning, L. Zhang, D. Zhang and C. Wu, "Robust Mean Shift Tracking with Corrected Background-Weighted Histogram," *IET Computer Vision*, 2010.

[5] 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. 7 ,pp.1245–1263,2009.

[6] G. Bradski, "Compuer vision face tracking for use in a perceptual user interface, " *Intel Technology Journal*, 2(Q2) , 1998.

[7] J. Ning, L. Zhang, D. Zhang and C. Wu, "Scale and Orientation Adaptive Mean Shift Tracking," *IET Computer Vision*, 2011.

[8] Q. A. Nguyen, A. Robles-Kelly and C. Shen, "Enhanced kernel-based tracking formonochromatic and thermographic video," *Proc. IEEE Conf . Video and Signal Based Surveillance*, pp. 28–33,2006.

[9] C. Yang, D. Ramani and L. Davis, "Efficient mean-shift tracking via a new similiarity measure," *Proc. IEEE Conf . Computer Vision and Pattern Recognition*, pp. 176–183,2005.

[10] I. Haritaoglu and M. Flickner, "Detection and tracking of shopping groups in stores," *Proc. IEEE Conf . Computer Vision and Pattern Recognition* , Kauai, Hawaii, pp. 431–438, 2001.

[11] T. Ojala, M. Pietik äinen, and T. T. M äenp ää, "Multiresolution gray-scale and rotation invariant texture classification with Local Binary Pattern," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 24, no. 7, pp. 971-987, 2002.

[12] T. Ahonen, A. Hadid, and M. Pietikäinen, "Face recognition with Local Binary Patterns: application to face recognition," I*EEE Trans. on Pattern Analysis and Machine Intelligence,* vol. 28, no. 12, pp. 2037-2041, 2006.

[13] T.Ojala, M. Pietikainen, and D. Harwood, "A comparative study of texture measures with classification based on feature distributions," *Pattern Recognition*, vol. 29, no. 1, pp. 51–59,1996.

[14] T. Ojala, T. M äenp ää, M. Pietik äinen, J. Viertola, J. Kyllönen, and S. Huovinen, "Outex – new framework for empirical evaluation of texture analysis algorithm," *Proc. Inte'l. Conf. on Pattern Recognition*, pp. 701-706, 2002.

[15] T. Ojala and M. Pietikäinen, "Unsupervised texture segmentation using feature distributions," *Pattern Recognition*, 32, pp.477-486,1999.

[16] G. Zhao, and M. Pietikäinen, "Dynamic texture recognition using Local Binary Patterns with an application to facial expressions," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 27, no. 6, pp. 915-928, 2007.

[17] X. Tan and B. Triggs. "Enhanced Local Texture Feature Sets for Face Recognition Under Difficult Lighting Conditions," *IEEE Trans. on Image Processing*, 19(6): pp. 1635-1650, 2010.

[18] Z. Guo, L. Zhang and D. Zhang, "A Completed Modeling of Local Binary Pattern Operator for Texture Classification," *IEEE Trans. on Image Processing*, vol. 19, no. 6, pp. 1657-1663, June 2010.

[19] S. Liao, M. K. Law and A. S. Chung, "Dominant Local Binary Patterns for Texture Classification," *IEEE Trans. on Image Processing*, Vol. 18, No. 5, pages 1107 – 1118, May, 2009

[20] L. Zhang, L. Zhang, Z. Guo and D. Zhang, "Monogenic-LBP : A new approach for rotation invariant texture classification," *Inte'l. Conf. on Image Processing*, pp. 2677-2680, 2010

[21] R.M.N Sadat and S. W. Teng, "Texture Classification Using Multimodal Invariant Local Binary Pattern," *IEEE Workshop on Applications of Computer Vision*, pp 315-320, 2011.

[22] S. Zhao, Y. Gao, and B. Zhang, "Sobel-LBP," *15th IEEE International Conference on Image Processing*, pp. 2144–2147, 2008.

[23] "Pets2001: http://www.cvg.rdg.ac.uk/pets2001/," .

[24] S. Wang, H. Lu, F. Yang, M. Yang, "Superpixel Tracking," *13th International Conference on Computer Vision*, pp. 1323-1330, 2011

[25] A. Adam, E. Rivlin, and I. Shimshoni, "Robust fragments-based track-ing using the integral histogram," *IEEE Inte'l. Conf. on Computer Vision and Pattern Recogintion*, pp. 798–805, 2006.

[26] W. Zhong, H. Lu and M.H. Yang, "Robust Object Tracking via Sparsity-based Collaborative Model, " *IEEE Inte'l. Conf. on Computer Vision and Pattern Recogintion*, 2012.

AUTHORS' INTRODUCTION

Pu Xiao-Rong received the M.S. degree in artificial intelligence from Southwest Normal University, Chongqing, China, in 2002 and the Ph.D. degree in computational intelligence from the University of Electronic Science and Technology of China (UESTC), Chengdu, China, in 2007. She was a visiting scholar with the University of Manchester Institute of Science and Technology, Manchester, U.K., in 2004. Currently, she is an associate professor with the School of Computer Science and Engineering, UESTC. Her current research interests include neural networks, biometrics, and affective computing.

Zhou Zhihu is now a graduate student at School of Computer Science and Engineering in UESTC. His research interests include biometrics and affective computing.