CrossMark

ORIGINAL PAPER

Improved object tracking via joint color-LPQ texture histogram based mean shift algorithm

Saadia Medouakh[1] · Mohamed Boumehraz[1] · Nadjiba Terki[2]

Abstract In this paper, a new robust mean shift tracker is proposed by utilizing the joint color and texture histogram. The contribution of our work is to take local phase quantization (LPQ) operator advantage of texture features representation, and to combine it with a color histogram mean shift tracking algorithm. The LPQ technique can be applied to obtain the texture features which represent the object. In texture classification, The LPQ operator is much robust to blur than the well-known local binary pattern operator (LBP). Compared with traditional color histogram mean shift algorithm which considers only color statistical information of the object, the joint color-LPQ texture histogram is more robust and overcome some difficulties of the traditional color histogram mean shift algorithm. Comparative experimental results on numerous challenging image sequences show that the proposed algorithm obtains considerably better performance than several state-of-the-art methods, especially traditional mean shift tracker. The algorithm is evaluated by numerical parameters: the center location and the average overlap, it proved the tracking robustness in presence of similar target appearance and background, motion blurring.

✉ Saadia Medouakh
    medsaadia@yahoo.fr

    Mohamed Boumehraz
    medboumehraz@mselab.org

    Nadjiba Terki
    t_nadjiba@yahoo.fr

[1] Energy Systems Modeling Laboratory, University of Biskra, Biskra, Algeria

[2] Department of Electrical Engineering, University of Biskra, Biskra, Algeria

1 Introduction

Object tracking is one of the most significant tasks in the field of computer vision [1–3]. It is widely applied in military and civil field such as surveillance, visual navigation [3] and human–computer interaction [1], etc. Object tracking purpose is to detect, extract, recognize, and track a moving object, get its state parameters, and understand the behavior of the object. Although many tracking algorithms [2] have been developed in recent years, it remains a very challenging problem due to factors such as illumination changes, occlusion, background clutters, scale appearance change and fast motions.

In the literature, several related surveys [1,4,5] of visual object tracking have been made to investigate the state-of-the-art tracking algorithms and their potential applications. Recently, tracking algorithms can be generally categorized into two main categories: generative tracking [2,6–9] and discriminative tracking [10–15].

Visual tracking using multiple types of visual features (intensity, color, spatio-temporal information, gradient and texture, etc) has been proved as a robust approach because features could complement each other [16]. In the literature, many tracking algorithms [4,17–22] attempt to combine multiple features to increase the representation accuracy against appearance variations and enhancing the discriminability between the target and its background.

In this paper, we propose a new hybrid color-texture feature for tracking based mean shift algorithm, which is a generative approach. Mean Shift is a good real-time performance, simple calculation and it is insensitive to edge

🖄 Springer

occlusion, rotation, deformation and background motion, as it adopts kernel histogram to represent its object model [23].

The mean shift algorithm [24] is a fast adaptive tracking procedure that finds the maximum of the Bhattacharyya coefficient given a target model and a starting region. This algorithm based on the color histogram used to depict the target region [25]. Color histogram estimates the point sample distribution to represent robustly the object appearance [17]. However, the use of color histograms to model the target object only makes the mean shift algorithm incapable to detect the spatial information which is lost and cannot distinguish between the target and the background, when the target has similar appearance [17,26].

In order to overcome the disadvantages, multiple features have been used in combination with color histogram [4, 27]. Recently, many researchers proposed various improved methods [17,20,26,28] used the joint color-texture histogram. It is more reliable than using only color histogram in tracking complex video scenes. The texture features introduce new information that the color histogram does not convey and reflect a stable spatial structure of the object, usually not subject to the impact of light and background color. The local binary pattern (LBP) method introduced by Ojala et al. [29] is a very effective technique which describes image texture features. In [17], Ning et al combined LBP technique and the color histogram mean shift tracking algorithm to construct a joint color-texture histogram algorithm. The results obtained outperformed the traditional color based method.

The LPQ operator presented in [30] uses a similar binary encoding scheme as the LBP method, where the descriptor is formed from a histogram of codewords computed for some image region. The results obtained have already shown that LPQ technique is insensitive to centrally symmetric blur that includes linear motion, out of focus, and atmospheric turbulence blur [31] and gave slightly better results for sharp texture images [30] than the LBP method. Therefore, adopting LPQ technique combined with color histogram mean shift algorithm for object tracking is a natural choice.

In this paper, we implement the LPQ to represent the target texture feature and then propose a joint color-texture histogram method for a more distinctive and effective target representation in mean shift tracker. The LPQ is firstly applied to represent the target texture features and then combined with color HSV to form a distinctive and effective target representation called joint color-LPQ histogram.

This paper is organized as follows. Section 2 proposes a short description of the mean shift tracking algorithm. The proposed joint color-LPQ texture histogram based mean shift tracker is described in Sect. 3. Section 4 gives the comparative experimental results of the proposed tracker against traditional mean shift tracker and several tracking methods. Finally, Sect. 5 gives the conclusion of this work.

2 Mean shift algorithm

Mean shift is a nonparametric estimator of gradient density. This algorithm calculates local maxima in any probability distribution [2,3].

According to [2], in the first frame, we select a rectangular region which contains the object. Let $\{x_i^*\}_{i=1...n}$ be the normalized pixel locations of the target model, $b(x_i^*)$ is the bin of x_i^* in the quantized feature space. The target probability model corresponds to each feature $u = 1, 2, \ldots, m$ can be expressed as:

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

where δ is the Kronecker delta and C is a normalization constant so that $\sum_{u=1}^{m} \hat{q}_u = 1$.

Let $\{x_i^*\}_{i=1...n_h}$ be the normalized pixel locations of the target candidate, centered at location y in the current frame. The target probability candidate corresponds to each feature, $u = 1, 2, \ldots, m$, can be expressed as:

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

To measure the similarity between the target model and the target candidate, mean shift uses the Bhattacharyya coefficient as the similarity function, it is defined as:

$$\rho(y) \equiv \rho\left[\hat{p}(y), \hat{q}\right] = \sum_{u=1}^{m} \sqrt{\hat{p}_u(y)\hat{q}_u} \tag{3}$$

The most probable location y of the target in the current frame is obtained by maximizing the Bhattacharyya coefficient $\rho(y)$. In the iterative process, the estimated target moves from y_0 to a new position y_1, which is defined as:

$$y_1 = \frac{\sum_{i=1}^{n_h} x_i w_i g\left(\left\|\frac{y_0 - x_i}{h}\right\|^2\right)}{\sum_{i=1}^{n_h} w_i g\left(\left\|\frac{y_0 - x_i}{h}\right\|^2\right)} \tag{4}$$

where

$$w_i = \sum_{u=1}^{m} \sqrt{\frac{\hat{q}_u}{\hat{p}_u(y_0)}} \delta\left[b(x_i) - u\right], \quad g(x) = -k'(x) \tag{5}$$

3 Proposed joint color-texture histogram for tracking based mean shift algorithm

In this work, we propose a novel target representation method by using the joint HSV color and LPQ texture features in the mean shift framework. Mean shift algorithm has been proved to be robust to partial occlusion, scale, rotation and non-rigid deformation of the target [2]. However, using only

color histogram in mean shift tracking has some problems [17]. First, the spatial information of the target is lost. Second, it caused object tracking inaccurately or fail easily when the object color was close to background color or the motion fast of the target or the camera (motion blur). To overcome the limitation of color histogram mean shift tracker, we use the HSV color space instead of the RGB color space, and the LPQ texture features have been used in combination with color histogram, which explores the blur invariant property and the spatial information. The details of the proposed method are explained in the following subsections.

3.1 LPQ texture features

The LPQ operator was proposed by Ojansivu et al. for texture description [30]. It was shown that the operator is robust to blur and outperform the Local Binary Pattern operator [29] in texture classification. LPQ utilizes the blur invariance property of the Fourier phase spectrum [30]. It is based on the quantized phase of the discrete Fourier transform computed locally for small image patches [31].

The spatial blurring is represented by a convolution between the image intensity and a point spread function (PSF). In the Fourier domain, this corresponds to

$$G(u) = F(u) \cdot H(u) \tag{6}$$

where $G(u)$, $F(u)$, and $H(u)$ are the discrete Fourier transforms (DFT) of the blurred image, the original image, and the PSF, respectively, and u is a vector of coordinates $[u, v]^T$ in the frequency domain. The magnitude and phase can be separated from:

$$|G(u)| = |F(u)| \cdot |H(u)| \text{ and } \angle G(u) = \angle F(u) \cdot \angle H(u) \tag{7}$$

The blur PSF is centrally symmetric, its Fourier transform H is always real-valued, and as a consequence its phase is only a two-valued function, given by:

$$\angle H(u) = \begin{cases} 0 & \text{if } H(u) \geq 0 \\ \pi & \text{if } H(u) < 0 \end{cases} \tag{8}$$

In the LPQ method, it is assumed that in the very-low frequency band, the value of $H(u)$ is positive with $\angle H(u) = 0$, so the phase information of $G(u)$ and $F(u)$ is the same and therefore a blur invariant representation can be obtained from the phase.

$$\angle G(u) = \angle F(u) \quad for \ all \ \angle H(u) \geq 0 \tag{9}$$

In LPQ the phase is examined in local M-by-M neighborhoods N_x (N_x is a window region of $M \times M$ pixels associated

with x position) at each pixel position x of the image $f(x)$. These local spectra are computed using a short term Fourier transform (STFT) defined by:

$$F(u, x) = \sum_{y \in N_x} f(x - y) e^{-j2\pi u^T y} \tag{10}$$

where $x = \{x_1, x_2, \ldots, x_N\}$ consist of simply 1-D convolution for the rows and columns successively. The local Fourier coefficients are computed at four frequency points $u_1 = [a, 0]^T$, $u_2 = [0, a]^T$, $u_3 = [a, a]^T$ and $u_4 = [a, -a]^T$, where a is a sufficiently small scalar to satisfy $H(u_i) > 0$. For each pixel position, this results in a vector:

$$F_x = [F(u_1, x), F(u_2, x), F(u_3, x), F(u_4, x)] \tag{11}$$

The phase information in the Fourier coefficients is recorded by observing the signs of the real and imaginary parts of each component in F_x. This is done by using a simple scalar quantizer:

$$q_j = \begin{cases} 1 & \text{if } g_j(x) \geq 0 \\ 0 & \text{if otherwise} \end{cases} \tag{12}$$

where $g_j(x)$ is the jth component of the vector $G_x = [Re\{F_x\}, Im\{F_x\}]$. The resulting eight binary coefficients $q_j(x)$ are represented as integer values between 0 and 255 using binary coding, given by:

$$f_{LPQ}(x) = \sum_{j=1}^{8} q_j(x) 2^{j-1} \tag{13}$$

As a result, we get the label image f_{LPQ} whose values are the blur invariant LPQ labels.

3.2 HSV color features

In the traditional mean shift tracking algorithm, the RGB color histogram has been used to target representation. However, the RGB space is not a perceptually uniform color space, that is, the difference between the colors in the RGB space do not correspond to the color differences perceived by humans. Additionally, the RGB dimensions are highly correlated [1]. To improve the performance of the mean shift tracking algorithm, we applied the HSV color histogram, which given robust tracking to lighting conditions, because HSV (hue, saturation and value) is an approximately uniform color space.

3.3 Tracking by mean shift with joint HSV color-LPQ texture histogram

The process of the proposed joint color-LPQ texture histogram based mean shift tracker is described in Fig. 1.
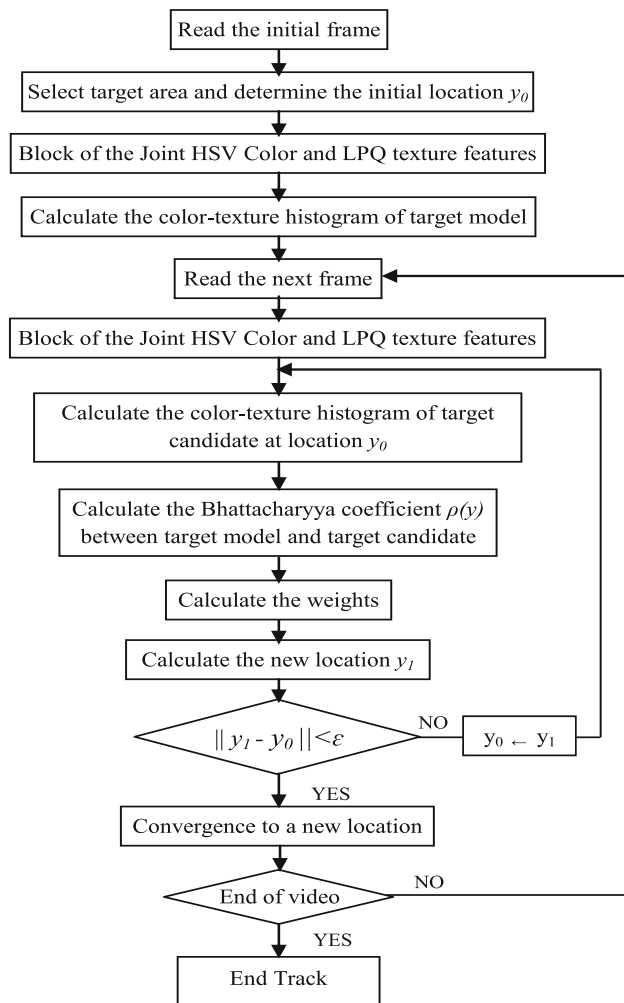
Read the initial frame

Select target area and determine the initial location y_0

Block of the Joint HSV Color and LPQ texture features

Calculate the color-texture histogram of target model

Read the next frame

Block of the Joint HSV Color and LPQ texture features

Calculate the color-texture histogram of target candidate at location y_0

Calculate the Bhattacharyya coefficient $\rho(y)$ between target model and target candidate

Calculate the weights

Calculate the new location y_1

$\| y_1 - y_0 \| < \varepsilon$ — NO → $y_0 \leftarrow y_1$

YES

Convergence to a new location

End of video — NO

YES

End Track

Fig. 1 Flow chart of the joint color-LPQ texture based mean shift tracking

Firstly, we calculate the LPQ feature of each pixel in the target region for obtaining LPQ target region, whose value is between 0 and 255. Then, we use the HSV channels and the LPQ patterns to jointly represent the target by the color and texture features, as shown in Fig. 2. To obtain the color and texture distribution of target region, we use (Eq. 1) to calculate the color and texture distribution of target model q, in which $m = N_H \times N_S \times N_v \times N_{LPQ}$ the number of joint feature spaces. The three dimensions (i.e., $16 \times 16 \times 16$) represent the quantized bins of color HSV channels and the fourth dimension (i.e., 16) is bins of the LPQ texture feature. The distribution color and texture of target model q has four dimensions 4D (i.e., $16 \times 16 \times 16 \times 16$). Similarly, the target candidate model $\hat{p}(y)$ is calculated with (Eq. 2). Figure 3 illustrates an example of the joint HSV color and LPQ texture distribution of target model q in the features $u_H = 3, u_S = 2, u_V = 4, u_{LPQ} = 1$.
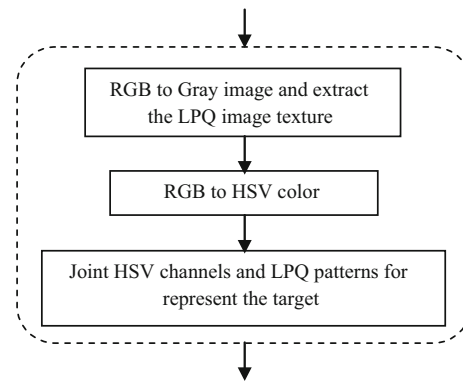
RGB to Gray image and extract the LPQ image texture

RGB to HSV color

Joint HSV channels and LPQ patterns for represent the target

Fig. 2 Block of the joint HSV color and LPQ texture features

4 Experimental results

In this section, we introduce the implementation details of the proposed joint color-LPQ texture based mean shift tracking algorithm MSLPQ and mean shift tracking algorithm MS and then report the experimental results by extensively evaluating the proposed tracking methods MSLPQ and MS on numerous video sequences including a comprehensive tracking benchmark [32]. To evaluate the effectiveness of the MSLPQ and MS, we implement it using HSV color histogram, and we use image gray for extracted texture feature with radius $R = 8$ for obtaining the LPQ image. The iteration numbers of mean shift tracking equal to 20 and the position error threshold ζ equal to 0.5. The proposed algorithms MSLPQ and MS are implemented by using the Matlab software. We validate the effectiveness of the proposed trackers by performing the experiments on several sequences from the OTB [32] and VOT 2015 datasets. These sequences include motion blur (MB), occlusion (OCC), background clutters (BC), deformation, fast motion (FM), rotation (IPR, OPR), and illumination (IV) and scale (SV) changes. Sequences *Boy, BlurFace, David3, Blur-Body, Human8, Lemming* and *BlurCar4* are from OTB, *Bolt2* is from VOT 2015. In the first frame, the target of interest is selected manually. These sequences are publicly available on the websites: http://www.visual-tracking.net, http://www.votchallenge.net

4.1 Quantitative comparison

In this section, we quantitatively compare MSLPQ and MS with four other popular algorithms: (1) the traditional mean shift tracking (KMS) [2]; (2) multiple instance learning (MIL) [13]; (3) Visual Tracking Decomposition (VTD) [9]; and (4) structured output tracking with kernels (Struck) [10]. We evaluate the performance of each tracker based on overlapping rate and center location error [32]. Center location error indicates the euclidean distance between the center of
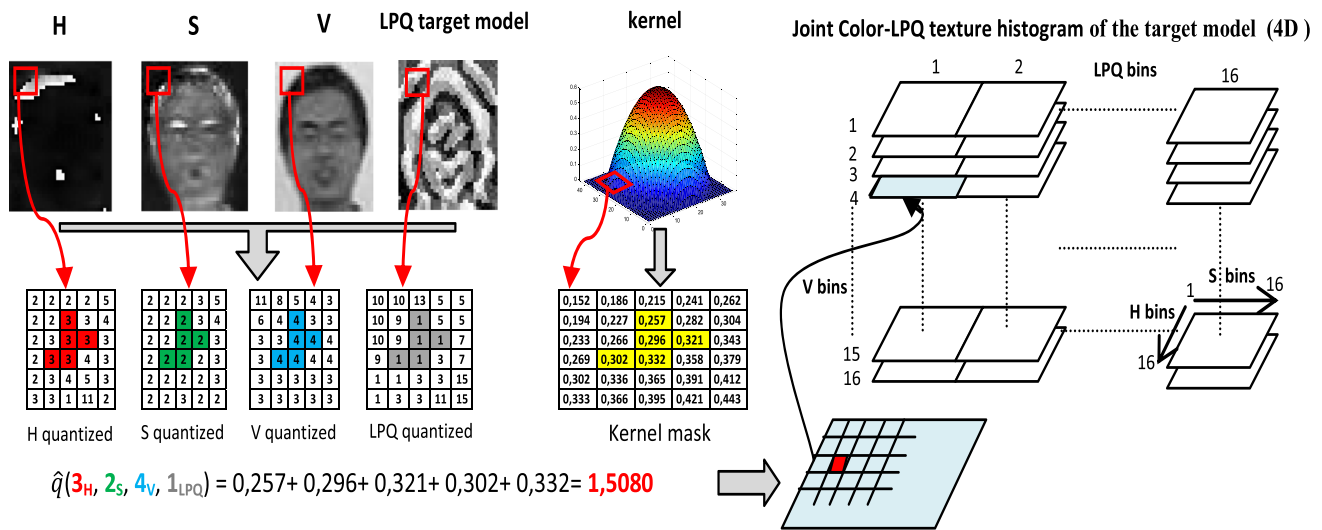
Fig. 3 Example of joint HSV color-LPQ texture histogram to target representation in mean shift algorithm

Table 1 Average overlap and center location error

Sequence	KMS	MIL	VTD	Struck	MS	MSLPQ
Boy	0.71/4.83	0.49/12.8	0.62/7.57	*0.75/3.84*	0.73/4.21	**0.76/3.66**
David3	0.65/*9.61*	0.53/29.6	0.40/66.7	0.29/106	*0.67*/11.8	**0.72/9.23**
BlurBody	0.62/27.3	0.03/206	0.23/146	**0.72**/*13.9*	0.64/22.8	*0.69*/**13.3**
BlurFace	0.70/14.5	0.27/71.9	0.41/40.6	0.46/42.3	**0.78/9.97**	*0.76*/11.2
Lemming	0.64/14.1	0.64/**12.0**	0.43/79.2	0.48/37.7	*0.66*/13.0	**0.66**/*12.2*
Human8	0.20/45.6	0.12/74.9	0.28/18.9	0.13/63.7	*0.48/5.68*	**0.50/3.36**
BlurCar4	0.70/21.8	0.06/197	0.07/185	**0.84/4.89**	0.67/25.0	*0.73/18.2*
Bolt2	0.40/39.9	0.68/*7.39*	0.50/17.1	0.22/86.4	**0.71/6.85**	*0.69*/7.49
Average	0.58/22.2	0.35/76.6	0.37/70.3	0.49/44.9	*0.67/12.4*	**0.69/9.84**

The bold number indicates the best performance, while italic indicates the second best

the bounding box around the tracked object and the ground truth. We use the precision to measure the overall tracking performance, which is defined as the percentage of frames whose estimated location within the given threshold distance of the ground truth. Usually, this threshold distance is set to 20 pixels [32,33].

Another evaluation is to compute the bounding box overlap S of r_t and r_g in each frame, where r_t is the bounding box outputted by a tracker and r_g is the ground truth bounding box. The bounding box overlap is defined as $S = |r_t \cap r_g| / |r_t \cup r_g|$, where $\cap$ and $\cup$ denote the intersection and union of two regions, respectively. In order to measure the overall performance of a given image sequence, we count the number of successful frames, whose overlap is larger than the given threshold 0.5. The success plot shows the ratios of successful frames at the thresholds varied from 0 to 1. We use the area under the curve (AUC) of each success plot to rank the comparison trackers.

The average overlap and the center location error of the four comparative trackers on the eight sequences are summarized in Table 1. Overall, the minimum center location

error and maximum overlap in all the tested sequences show that the proposed MSLPQ has the best average performances (0.6947) compared with the other methods. Moreover, MS appears to be more robust than original mean shift tracking KMS and achieves a better performance in this dataset especially in sequence Human8.

The overall performance of our tracker and other four compared trackers are shown as in Fig. 4. We use one pass evaluation (OPE) for the overall performance, and success rate and precision as an evaluation criterion, as shown in Fig. 4a, b, respectively. For comparison, we use 20 color images with different challenges. Obviously, the overall performance of our tracker outperforms the other 4 state-of-the-art trackers.

4.2 Qualitative comparison

In order to illustrate the qualitative comparison more clearly, Fig. 5 illustrates the tracking results of different trackers on several challenging sequences. The Boy and BlurFace sequences track human faces under MB, FM, IPR, SV and
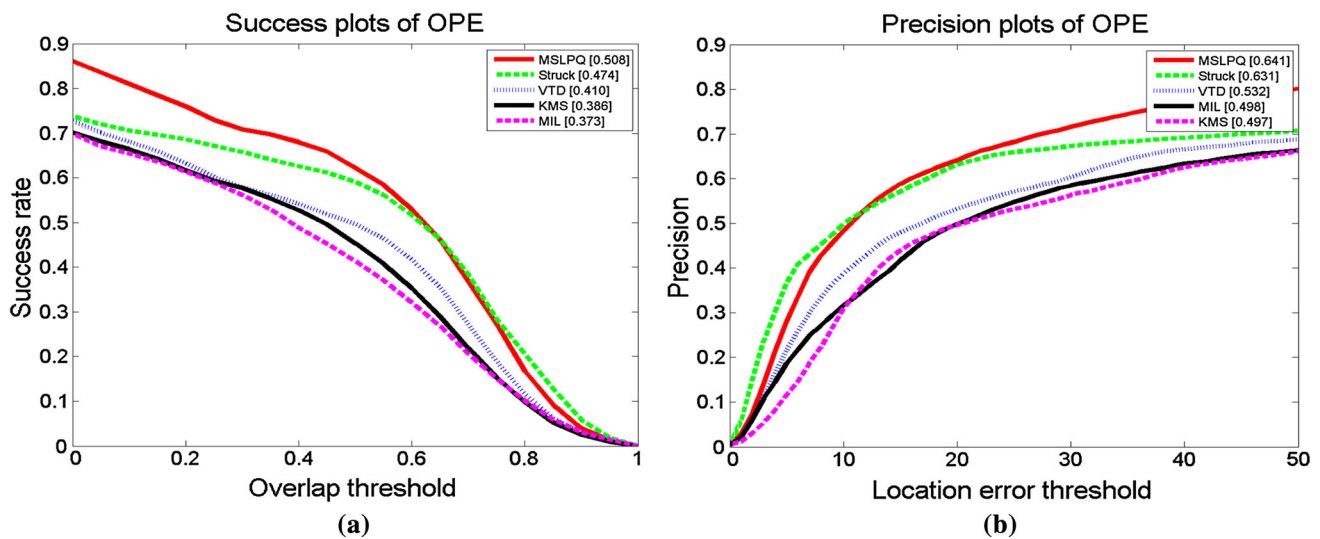
Fig. 4 Comparison with four state-of-the-art trackers on 20 color image sequences of the precision and success plots using one pass evaluation (OPE)

OPR. The experimental results show that MSLPQ obtains the best performance as shown in Fig. 5a, b. On the Boy sequence, the MIL fails to track the target from frame 577, where the frames include both fast motion and severe motion blur. In contrast, the MSLPQ can accurately follow the target at frames 206, 492 and 577 which are subject to motion blurring, fast motion and pose changes, when all the other trackers have failed. While, in BlurFace sequence, most trackers including MIL, KMS, VTD and Struck, fail to track the target from the early frames especially when severe blur motion occurs.

In Fig. 5c–f, we show some example frames for David3, BlurBody, Human8 and Bolt2 sequences, where the tasks are to track the human bodies in different settings. In the David3 and Bolt2 sequences, a moving Body is tracked, which presents many challenges such as partial occlusion and the background near the target has the similar color or texture as the target. In the David3 sequence, VTD, Struck and MIL are failed to track the target when a partial occlusion occurs at frame 130. In contrast, the proposed method is more robust when the background has the similar color as the target, this is the result of the employment of rich feature color-LPQ texture, while MS and KMS are less successful. In the Bolt2 sequence, the pose of target change rapidly and the appearance deform frequently, MSLPQ, MS and MIL succeed in tracking in the over the whole sequence, while VTD is only able to track a part of the target though does not lose it, but KMS loses the target when the background has the similar color as the target, after frame 185. However, Struck fails on the over the whole sequence. The BlurBody sequence includes Blur motion and DEF. MIL loses the target because of the blur motion, while VTD is able to track in the beginning, but it fails after frame 99. In contrast, the

MSLPQ and Struck obtain the best performance (frames 147, 273 and 321). Compared with MS and KMS, MSLPQ successfully tracks the target under different challenges due to the robustness of the additional LPQ texture features as in frame 273. On the Human8, most trackers MIL, VTD, Struck and KMS lose the target from the early frames onwards due to the target appearance is similar to the background. In contrast, the proposed methods MSLPQ and MS can accurately follow the target as in frames 88 and 122, despite the presence of many challenges such as background Clutters, scale and illumination changes.

In the last group of tested sequences, the tasks are varying from tracking car in the road in BlurCar4 sequence to tracking moving doll in Lemming sequence. Figure 5g–h shows some frames of these sequences. The Lemming sequence is much harder due to the significant appearance changes, OCC and out-of-plane rotations (OPRs). In this sequence, all trackers succeed in tracking the target in the beginning even frame 373. The VTD fails as in frame 739 due to heavy occlusion, while Struck gradually drifts from the target after the 948 frame and totally loses the target in the 963 frame (e.g., frame 1023) due to the pose changes. However, MSLPQ, MS, KMS and MIL are only able to track a part of the target though does not lose it (frame 1108). We can see that the MSLPQ and MS have the best performances than KMS and MIL, as a result of the use of color-LPQ texture feature for MSLPQ and 3D color histogram for MS. In the BlurCar4 sequence, the VTD and MIL lose the target from the start. Struck, MSLPQ, KMS and MS are able to track the moving car despite the dramatic severe blur motion and fast motion. In addition, these methods are prone to drift during the tracking due to the target appearance is somewhat similar to the background and in the presence another car has a similar color. However,

Fig. 5 Tracking results of different algorithms. Frame indexes are shown in the top left of each figure. **a** Boy, **b** BlurFace, **c** David3, **d** BlurBody, **e** Human8, **f** Bolt2, **g** Lemming, **h** BlurCar4

Struck and MSLPQ achieve the best performances than KMS and MS as shown in frame 331.

5 Conclusion

In this paper, we proposed an effective and robust algorithm of object tracking which uses color-LPQ texture histogram features and applying it to the mean shift tracking algorithm framework. The LPQ operator is used to represent the target structural information, which is more discriminant and insensitive to blurring. Then, the proposed target representation, joint color and LPQ texture, is used to effectively distinguish the foreground target and its background. The experimental results demonstrate that the proposed method is capable of taking advantage of color and texture features, compared with the original mean shift tracking algorithm. Comparative evaluations on challenging benchmark image sequences against several state-of-the-art trackers, such as KMS [2] and Struck [10], show that the proposed method obtains superior performance with respect to Background Clutters, blur motion, partial occlusions, and various appearance changes.

References

1. Yilmaz, A., Javed, O., Shah, M.: Object tracking: a survey. ACM Comput. Surv.: CSUR **38**(4), 1–45 (2006)
2. Comaniciu, D., Ramesh, V., Meer, P.: Kernel based object tracking. IEEE Trans. Pattern Anal. Mach. Intell. **25**(2), 564–577 (2003)
3. Zhao, Q., Yao, W., Xie, J.: Research on object tracking with occlusion. WIT Trans. Inf. Commun. Technol. **51**, 537–545 (2014)
4. Yang, H., Shao, L., Zheng, F., Wang, L., Song, Z.: Recent advances and trends in visual tracking: a review. Neurocomputing **74**, 3823–3831 (2011)
5. Li, X., Hu, W., Shen, C., Zhang, Z., Dick, A., Hengel, A.V.D.: A survey of appearance models in visual object tracking. ACM Trans. Intell. Syst. Technol. **4**(4), 1–58 (2013)
6. Ross, D.A., Lim, J., Lin, R.S., Yang, M.H.: Incremental learning for robust visual tracking. Int. J. Comput. Vis. **77**(1–3), 125–141 (2008)
7. Zhang, S., Yao, H., Zhou, H., Sun, X., Liu, S.: Robust visual tracking based on online learning sparse representation. Neurocomputing **100**, 31–40 (2013)
8. Mei, X., Ling, H.: Robust visual tracking using l1 minimization. In: IEEE 12th International Conference on Computer Vision, pp. 1436–1443 (2009)
9. Kwon, J., Lee, K.M.: Visual tracking decomposition. IEEE Conference on Computer Vision and Pattern Recognition (CVPR) **2010**, 1269–1276 (2010)
10. Hare, S., Saffari, A., Torr, P.H.: Struck: structured output tracking with kernels. In: IEEE International Conference on Computer Vision (ICCV), pp. 263–270 (2011)
11. Henriques, J.F., Caseiro, R., Martins, P., Batista, J.: High-speed tracking with kernelized correlation filters. IEEE Trans. Pattern Anal. Mach. Intell. **37**(3), 583–596 (2015)
12. Kalal, Z., Mikolajczyk, K., Matas, J.: Tracking-learning-detection. IEEE Trans. Pattern Anal. Mach. Intell. **34**(7), 1409–1422 (2012)
13. Babenko, B., Yang, M.H., Belongie, S.: Robust object tracking with online multiple instance learning. IEEE Trans. Pattern. Anal. Mach. Intell. **33**(8), 1619–1632 (2011)
14. Ma, C., Yang, X., Zhang, C., Yang, M-H.: Long-term correlation tracking. In: CVPR (2015)
15. Lan, X.Y., Yuen, P.C., Challappa, R.: Robust MIL-based feature template learning for object tracking. In: The Thirty First AAAI Conferences on Artificial Intelligence (AAAI), pp 4118–4125 (2017)
16. Lan, X.Y., Ma, A.J., Yuen, P.C., Challappa, R.: Joint sparse representation and robust feature-level fusion for multi-cue visual tracking. IEEE Trans. Image Process. **24**(12), 5826–5841 (2015)
17. Ning, J., Zhang, L., Zhang, D., Wu, C.: Robust object tracking using joint color-texture histogram. Int. J. Pattern Recognit. Artif. Intell. **23**(07), 1245–1263 (2009)
18. Wang, Q., Fang, J., Yuan, Y.: Multi-cue based tracking. Neurocomputing **131**, 227–236 (2014)
19. Lan, X.Y., Ma, A. J., Yuen, P.C.: Multi-cue visual tracking using robust feature-level fusion based on joint sparse representation. In: CVPR, pp 1194–1201 (2014)
20. Dou, J., Li, J.: Robust visual tracking based on joint multi-feature histogram by integrating particle filter and mean shift. Optik Int. J. Light Electron Opt. **126**, 1449–1456 (2015)
21. Lan, X.Y., Zhang, S., Yuen, P.C.: Robust joint discriminative feature learning for visual tracking. In: IJCAI, pp. 3403–3410 (2016)
22. Yuan, Y., Xiong, Z., Wang, Q.: An incremental framework for video-based traffic sign detection, tracking, and recognition. IEEE Trans. Intell. Transp. Syst. **18**(7), 1918–1929 (2017)
23. Chen, X., Zhang, M., Ruan, K., Xu, G., Sun, S., Gong, C., Min, J., Lei, B.: Improved mean shift target tracking based on self-organizing maps. SIViP **8**, S103–S112 (2014)
24. Wen, Z., Cai, Z.: A robust object tracking approach using mean shift. Proceedings of the Third International Conference on Natural Computation (ICNC) **2**, 170–174 (2007)
25. An, K.H., Chung, M.J.: Mean shift based object tracking with new feature representation. In: Proceedings of the Third International Conference on Ubiquitous Robots and Ambient Intelligence, pp 365–370 (2006)
26. Xiaorong, P., Zhihu, Z.: A more robust mean shift tracker on joint color-CLTP histogram. Int. J. Image Gr. Signal. Process. **4**(12), 34–42 (2012)
27. Phadke, G., Velmurugan, R.: Mean LBP and modified fuzzy C-means weighted hybrid feature for illumination invariant mean-shift tracking. SIViP **11**(4), 665–672 (2017)
28. Tavakoli, R.H., Moin, M.S., Heikkila, J.: Local similarity number and its application to object tracking. Int. J. Adv. Robot. Syst. **10**, 184 (2013)
29. Ojala, T., Pietikäinen, M., Mäenpää, T.: Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Trans. Pattern Anal. Mach. Intell. **24**(7), 971–987 (2002)
30. Ojansivu, V., Heikkila, J.: Blur insensitive texture classification using local phase quantization. In: Proceedings of Image and Signal Processing ICISP 2008, Cherbourg-Octeville, pp. 236–243 (2008)
31. Heikkila, J., Ojansivu, V.: Methods for local phase quantization in blur-insensitive image analysis. In: Proceedings of International Workshop Local and Non-Local Approximation in Image Processing, pp. 104–111 (2009)
32. Wu, Y., Lim, J., Yang, M. H.: Online object tracking: a benchmark. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2411–2418 (2013)
33. Qiao, L., Xiao, M., Weihua, O., Quan, Z.: Visual object tracking with online sample selection via lasso regularization. SIViP **11**(5), 881–888 (2017)