A Novel object Tracking Algorithm Based on mean shift algorithm and SURF

Xiaoqin Ma
School of big data and artificial intelligence
Chizhou University
Chizhou, China
czumxq@163.com

Lulu Li
School of Computer Science & Engineering
Yulin Normal University
Yulin, China
ylsyli66@126.com

***Abstract*—Mean shift algorithm (MSA) is a powerful object tracking technique due to its simplicity and robustness. However, it causes easily the inaccurate tracking when the scene is complex, for example, the object is seriously occluded, the color of the object is similar to that of the background, the background is dynamic, and the camera shakes or moves. Aiming at the above problems, a novel object tracking algorithm based on MSA and SURF (Speeded Up Robust Features) is proposed. Firstly, the object area in the current frame is determined by MSA, and secondly, SURF algorithm is used to match the feature points in the object area of initial frame with those in the object area of current frame, and finally, the coordinates of the center point in the object area of current frame are adjusted according to the matching results. The experimental results on three videos in complex scenes show that the proposed algorithm can track the object more accurately than MSA, and realize the real-time and accurate tracking of video objects in complex scenes.**

Keywords-object tracking; MSA; SURF

I. INTRODUCTION

A. Research background

The traditional MSA has been used in the field of object tracking for a long time. It has gradually attracted people's attention and become a research focus because of its advantages of low computational burden, easy implementation and good real-time performance [1]. MSA was first proposed by Fukunaga and Hostetler [2]. Cheng et al. extended it in two aspects and gave the possible application fields [3]. Comaniciu et al. applied MSA to object tracking for the first time [4].

MSA only uses the object color histogram to represent the object features, so it has many limitations. For example, it can not be applied to the situation where the background is similar to the color of object, the object is seriously occluded, the camera moves or shakes and the background is dynamic.

In view of the above-mentioned facts, many researchers have proposed a variety of algorithms. Jeyakar et al. split up the object into blocks and used the linear weighting method of color histogram of each block to represent the object features [5]. Leichter et al. extracted the color histogram of the object from multiple angles to represent the object features in all directions [6]. Li et al. put forward the pyramid algorithm of MSA to actualize the real-time global tracking of the object [7]. Li et al. added Kalman filter to MSA as a prediction mechanism, which can predict the position of linear moving object, but can't deal with the nonlinear situation [8]. Leichter proposed to use cross-bin color histogram to represent the object features, which can ameliorate the robustness of MSA [9]. Bousetouanef et al. combined texture features with color features to ameliorate the robustness of MSA [10]. Chen et al. combined MSA with frame-difference method to ameliorate the reliability of object tracking [11]. Du et al. proposed a new MSA based on Haar local binary pattern texture features. It improves the accuracy of object tracking in complex background [12]. Duan et al. proposed a new object tracking algorithm based on MSA and fast templet matching with good robustness and real-time performance [13]. Liu et al. proposed an adaptive smoothing method of infrared image with ship object based on MSA and block-based method [14]. The method effectively suppressed noise and highlighted ship object regions. Wang et al. proposed a real-time object tracking algorithm by using MSA and RGB-D feature. It performs well in accuracy and robustness when the object is disturbed by similar background [15]. Prajna et al. proposed an efficient visual tracking framework that is formulated by the exponential quantum particle filter and the traditional MSA for object tracking in complex videos [16].

B. Motivation and contribution

Various improved MSA does not completely solve the problems faced by MSA. SURF is a robust local feature point detection and description algorithm. SURF not only keeps the excellent performance of SIFT (Scale Invariant Feature Transform), but also overcomes its shortcomings, such as high computational complexity and time-consuming, and provides the possibility for the application in the real-time computer vision system. Various noises in the video cause inaccurate object tracking of MSA. Our idea is to use the powerful image matching ability of SURF to adjust the inaccurate video frames in real-time.

The innovations and contributions of this paper are as follows:

1)A novel object tracking algorithm is proposed by combining SURF algorithm with MSA. The proposed algorithm solves the problem of inaccurate object tracking of MSA when the video is occluded, the background and foreground colors are similar, and the background is dynamic and the camera moves.

2)Due to the accurate matching and low complexity of the SURF algorithm, the proposed algorithm realizes the real-time and accurate tracking of video objects in complex scenes by introducing the SURF algorithm.

2473-3636/22/$31.00 ©2022 IEEE
DOI 10.1109/DCABES57229.2022.00011

II. MATERIALS AND METHODS

A. MSA

MSA realized semi-automatic object tracking in video. In the start frame, the tracking object is selected by manually determining the search window.

The object model of MSA is a normalized color histogram $A = \{a_u\}_{u=1}^{m}$,where m is the number of bins. The normalized color distribution of a candidate object is $P(B) = \{p_u(B)\}_{u=1}^{m}$, where

$$p_u(B) = C_h \sum_{i=1}^{n_h} k(\left\| \frac{B - X_i}{h} \right\|^2) \delta[b(X_i) - u] \quad (1)$$

, $\{X_i\}_{i=1}^{n_h}$ denote n_h pixel locations in the object area, δ denotes Kronecker delta function, $b(X_i)$ indicates which histogram bin X_i is in, $k(x)$ denotes a kernel function, and C_h denotes the normalization constant. The color distribution of A can also be obtained from the above equation.

Bhattacharyya coefficient can evaluate the similarity between object and candidate, and its calculation formula is as follows:

$$\rho(B) = \rho[P(B), A] = \sum_{u=1}^{m} [p_u(B) a_u]^{\frac{1}{2}} \quad (2)$$

The position of the object in the current frame corresponds to the maximum value of Bhattacharyya coefficient that can be solved by running the mean-shift iterations. Suppose that the search for a new object location in the current frame starts at location B_0 . At every step of the iterative process, the estimated object moves from B_0 to the new location B_1 that can be calculated according to the following formula:

$$B_1 = \frac{\sum_{i=1}^{n_h} X_i w_i g(\|(B_0 - X_i)/h\|^2)}{\sum_{i=1}^{n_h} w_i g(\|(B_0 - X_i)/h\|^2)} \quad (3)$$

where $w_i = \sum_{u=1}^{m} [\frac{a_u}{p_u(B_0)}]^{\frac{1}{2}} \delta[b(X_i) - u]$ and $g(x) = -k'(x)$.

More information on the mean-shift object tracking algorithm can be found in the fourth reference.

B. SURF algorithm

Herbert Bay et al. proposed SURF algorithm in 2006 [17]. It is a image interest point detector and descriptor scheme with good robustness. SURF descriptor can be regarded as the gradient information extracted from SIFT (Scale-invariant feature transform) and its variants [18]. SURF and SIFT have similar performance, but SURF runs faster. The reason for the fast running speed is to use the integral images that can drastically reduce the number of simple box convolution operations without dependence on the chosen scale.

1) Interest point localization

SURF detector based on Hessian matrix has good performance in accuracy. The scale selection is determined by the determinant of Hessian matrix. Hessian matrix $H(P, \theta)$ at scale θ is defined as follows:

$$H(P,\theta) = \begin{pmatrix} C_{xx}(P,\theta) & C_{xy}(P,\theta) \\ C_{xy}(P,\theta) & C_{yy}(P,\theta) \end{pmatrix} , \quad (4)$$

where $P = (x, y)$ is any point in image I , $C_{xx}(P,\theta)$, $C_{xy}(P,\theta)$ and $C_{yy}(P,\theta)$ are convolutions of Gaussian second order partial derivative at point P . SIFT approximates Laplacian of Gaussian by Difference of Gaussian, while SURF approximates the second order partial derivatives of Gaussian by box filters, and therefore SURF runs much faster than SIFT.

2) Interest point matching

Matching of descriptors is achieved by the method of nearest neighbour distance ratio. Firstly, the distance is calculated between the descriptors of the matched feature points, and then if the ratio of first two minimum distances is less than a threshold ε , the descriptors are taken as a match.

C. A novel object tracking algorithm based on MSA and SURF

Aiming at the drawbacks of MSA, a novel object tracking algorithm based on MSA and SURF is proposed. The steps of the proposed algorithm are as follows：

- Manually determine the search window with the moving object in the start frame, as shown in Figure 1.

Figure 1. Manually selected tracking object in the start frame

- In the next frame, MSA is utilized for determining the search window with the moving object, as shown in Figure 2.

Figure 2. Tracking results of other frame using MSA

- Using SURF algorithm to match interest points of two object boxes, as shown in Figure 3.

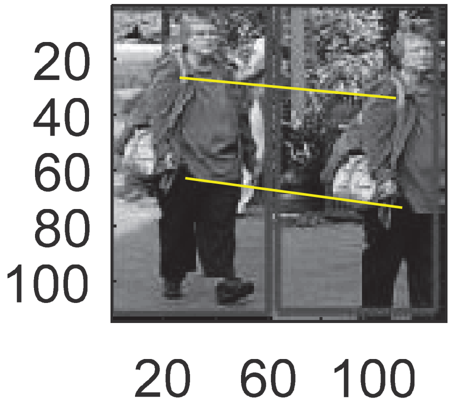

Figure 3. Matching results of two object boxes

- Calculate the coordinates of the matched interest points in each object box.
- Adjust the tracking object position according to the average difference of coordinates in two object boxes.The adjusted tracking results are shown in Figure 4.


Figure 4. Adjusted tracking results using our algorithm

III. EXPERIMENTAL RESULTS AND DISCUSSION

A. *Experimental Setting*

All the experiments are conducted on a 64 bit Windows 10 system with Core(TM) i7-4790 CPU and 8 GB memory. Both our algorithm and MSA provide MATLAB program.

B. *Data Set Preparation*

In this study, the chosen data set is a complex video that is taken from PETS 2009 Benchmark Data . The test video features similar color of object and background, camera’s jitter and motion, dynamic background, and poor illumination.

C. *Performance comparison*

To test the performance of our algorithm, we compare it with MSA.

1) Speed

The experimental results are shown in table 1. The speed is the average processing time of each frame when processing 30 images. The unit of speed is seconds. The size of each frame is 320×240.

TABLE I. SPEED COMPARISON BETWEEN MSA AND OUR ALGORITHM

Algorithm	Video
MSA	0.07s
Our algorithm	1.09

Table 1 show that MSA is much faster than our algorithm in terms of speed. However, tracking objects in the video with 300 frames, our algorithm takes about 5 minutes, which is acceptable.

2) Accuracy

The experimental results are shown in table 2 and figure 5. This paper only shows the tracking results of 30 frames in the test video.

Table 2 shows the coordinates of the center points of the object boxes obtained by MSA and our algorithm respectively in the test video.

As shown in table 2 , the abscissa gradually decreases because the woman in the video walks from right to left. The abscissa of the center of some frames does not change much because of camera movement.

Figure 5 shows the distance between the actual center point and the center point of the object frame tracked by our algorithm and MSA.

As shown in Figure 6, the MSA does not track the object accurately because of camera movement and similar color between background and object. Figure 7 show that our algorithm overcomes the interference of camera movement and similar color between background and object and tracks the object accurately.

In summary, the tracking results of our algorithm are better than that of MSA, and the position of the object frame found by our algorithm is accurate on the whole, and the tracking error of the proposed algorithm is relatively stable, while the tracking error of MSA increases with the increase of the number of tracked frames, which fully verifies that the proposed algorithm effectively solves the problems of MSA for object tracking in complex scenes, as shown in Table 2 and Figures 5-7.

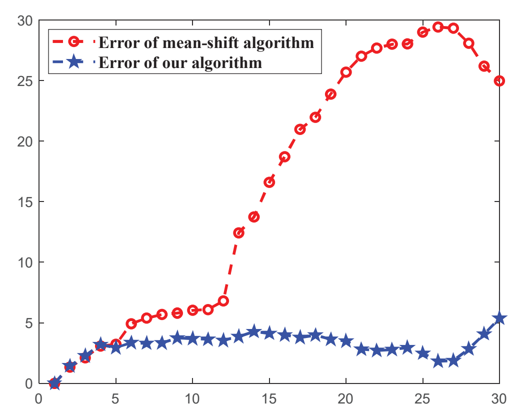

Figure 5. The distance between the actual center point and the center point of the tracked object by MSA and our algorithm

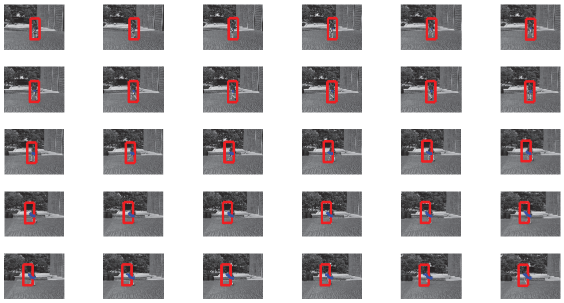

Figure 6. Tracking results of the test video by MSA

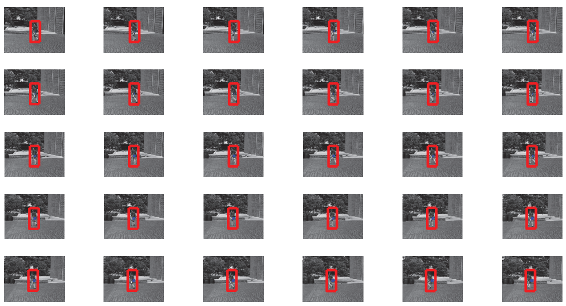

Figure 11. Tracking results of the second video by our algorithm

TABLE II. THE COORDINATES OF THE CENTER POINTS OF THE OBJECT BOXES OBTAINED BY MSA AND OUR ALGORITHM IN THE TEST VIDEO

Frame number	The actual coordinates of the center points of the object boxes	The coordinates of the center points of the object boxes obtained by MSA	Error of MSA	The coordinates of the center points of the object boxes obtained by our algorithm	Error of our algorithm
1	(164.50,130.00)	(164.50,130.00)	0	(164.50,130.00)	0
2	(163.00,129.00)	(164.19,129.63)	1.35	(164.36,129.49)	1.45
3	(162.00,128.00)	(163.85,129.01)	2.11	(164.02,129.05)	2.28
4	(162.00,126.00)	(163.44,128.73)	3.09	(163.75,128.68)	3.20
5	(162.00,126.00)	(162.92,129.09)	3.22	(163.37,128.61)	2.95
6	(161.00,126.00)	(161.60,130.88)	4.92	(162.97,128.72)	3.36
7	(161.00,126.00)	(161.11,131.39)	5.39	(162.58,128.90)	3.30
8	(161.00,126.00)	(160.59,131.68)	5.69	(162.22,129.10)	3.33
9	(160.00,126.00)	(160.05,131.80)	5.80	(161.93,129.21)	3.75
10	(160.00,126.00)	(159.49,132.01)	6.03	(161.70,129.28)	3.69
11	(160.00,126.00)	(158.83,131.98)	6.09	(161.45,129.33)	3.63
12	(160.00,126.00)	(157.42,132.30)	6.81	(161.18,129.34)	3.54
13	(159.00,126.00)	(146.58,126.60)	12.43	(160.89,129.37)	3.86
14	(158.00,126.00)	(144.26,126.05)	13.74	(160.59,129.40)	4.27
15	(158.00,126.00)	(141.67,122.97)	16.61	(160.27,129.44)	4.12
16	(158.00,126.00)	(140.11,120.46)	18.72	(159.88,129.52)	3.99
17	(158.00,126.00)	(138.73,117.69)	20.99	(159.46,129.50)	3.79
18	(157.00,126.00)	(137.26,116.33)	21.98	(158.99,129.45)	3.98
19	(157.00,126.00)	(135.87,114.84)	23.90	(158.47,129.32)	3.63
20	(156.00,126.00)	(133.97,112.74)	25.71	(157.83,128.98)	3.50
21	(156.00,126.00)	(132.64,112.39)	27.04	(157.14,128.57)	2.81
22	(155.00,126.00)	(131.10,111.99)	27.70	(156.44,128.30)	2.71
23	(154.00,126.00)	(129.77,111.94)	28.01	(155.71,128.20)	2.79
24	(153.00,126.00)	(128.20,112.92)	28.04	(154.93,128.23)	2.95
25	(153.00,126.00)	(127.06,113.00)	29.02	(154.11,128.22)	2.48
26	(152.00,127.00)	(126.01,113.18)	29.44	(153.31,128.27)	1.82
27	(151.00,127.00)	(125.11,113.18)	29.35	(152.34,128.32)	1.88
28	(149.00,127.00)	(124.26,113.69)	28.09	(151.45,128.46)	2.85
29	(147.00,127.00)	(123.59,115.23)	26.20	(150.74,128.61)	4.07
30	(145.00,127.00)	(122.79,115.56)	24.98	(150.05,128.87)	5.39

IV. CONCLUSIONS

In this paper, a novel object tracking algorithm based on MSA and SURF is proposed. Firstly, the moving object is roughly located using MSA, and then the location of the tracked object is precisely adjusted using SURF algorithm. Our algorithm still maintains high accuracy, stability and continuity of object tracking in complex scenes while solving the problems of MSA. In addition, unlike the deep learning model with a large number of labeled training data, our algorithm needs only the bounding box of the first frame.

ACKNOWLEDGMENT

Thanks to the support of 2020 Domestic Visiting Program for Outstanding Young Backbone Talents in Colleges and Universities in Anhui Province. (grant NO. gxgnfx2020113).

REFERENCES

[1] He L., Xu Y., Chen Y., Wen J., "Recent advance on mean shift tracking: a survey," International Journal of Image and Graphics, vol. 13, July. 2013, pp. 1-29, doi:10.1142/s0219467813500125.

[2] Fukunaga K., Hostetler L., "The estimation of the gradient of a density function with applications in pattern recognition," IEEE Transactions on Information Theory, vol. 21, Nov. 2006, pp. 32-40, doi:10.1109/TIT.1975.1055330.

[3] Cheng Y., "Mean shift, mode seeking and clustering," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.17, Sep. 1995, pp. 790-799, doi:10.1109/34.400568.

[4] Comaniciu D., Ramesh V., Meer P., "Kernel-based object tracking," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.25, Dec. 2003, pp. 564-577, doi:10.1109/ TPAMI. 2003. 1195991.

[5] Jeyakar J., Babu R., Ramakrishnan K., "Robust object tracking with background-weighted local kernels," Computer Vision and Image Understanding, vol.112, Dec. 2008, pp. 296-309, doi: 10.1016/ j.cviu.2008.05.005.

[6] Leichter I., Lindenbaum M., Rivlin E., " Mean shift tracking with multiple reference color histograms," Computer Vision and Image Understanding, vol.114, Mar. 2010, pp. 400-408, doi:10.1016/ j.cviu.2009.12.006.

[7] Li S X., Chang H., " Adaptive pyramid mean shift for global real-time visual tracking," Image and Vision Computing, vol.28, Mar. 2010, pp. 424-437, doi: 10.1016/j.imavis.2009.06.012.

[8] Li X., Zhang T., Shen X., et al. Object tracking using an adaptive Kalman filter combined with mean shift. Optical Engineering 2010, 49,020503-1-020503-3.

[9] Leichter I. Mean shift trackers with cross-bin metrics. IEEE Transactions on Pattern Analysis and Machine Intelligence 2012, 34, 695-706.

[10] [10]Bousetouanef F., Dib L., Snoussi H. Improved mean-shift integrating texture and color features for robust real time object tracking. Visual computer 2013, 29, 155-170.

[11] [11] Chen S., Wang S. Moving object tracking based on five frame difference and improved mean-shift algorithm. Computer science 2016, 43, 203-206.

[12] [12] Du J., Huang S., Yang S. Mean-shift target tracking algorithm of adaptive HLBP texture feature. Computer science 2017, 44, 217-220.

[13] [13] Duan L., Song P., Chen Z., Zhao P. A target tracking algorithm based on mean-shift and fast template matching. Journal of Northwestern polytechnical university 2018, 36, 792-799.

[14] [14] Liu Z., Bai X., Sun C., Zhou F., Li Y. Multi-modal ship target image smoothing based on adaptive mean shift. IEEE Access 2018, 6, 12573-12586.

[15] [15] Wang G., Tian J., Zhu W., Fang D. Real-time object tracking combined RGB-D with Mean-shift. Journal of Central South University 2019, 50, 2163-2170.

[16] [16] Dash P., Patra D. An efficient hybrid framework for visual tracking using Exponential Quantum Particle Filter and Mean Shift optimization. Multimedia Tools and Applications 2020, 79, 1-25.

[17] [17] Bay H., Tuytelaars T., Gool L. SURF: Speeded Up Robust Features. Computer Vision and Image Understanding 2008, 110, 346-359.

[18] [18] Lowe D. Distinctive image features from scale-invariant feature points. International Journal of Computer Vision 2004, 60, 91-110.