

# Object Tracking by Corrected Background-Weighted Histogram Mean Shift with Sum of Gradient Mode

Yu Yang, Yongxing Jia, Chuanzhen Rong, Lijuan Wang, Yuan Wang, Ying Zhu, Zhenjun Yue

PLA University of Science and Technology  
College of Communication Engineering  
Nanjing, China  
e-mail: idsp@qq.com

**Abstract**— With a focus on complex environments, the present paper describes a new algorithm in scale changed object tracking through color feature. Mean shift (MS) iterative procedure is the best color-based algorithm to find the location of an object. The algorithm performance is not acceptable once tracking scale changed objects in complex environments. In this paper, the main aim is to improve the MS method, using corrected background-weighted histogram (CBWH) algorithm to reduce the interference of background in target localization. To fit the object scale change, the sum of gradient mode (SGM) is employed. The experimental results show that the proposed method is superior to the traditional mean shift tracking in the following aspects: 1) it provides consistent object tracking throughout the video; 2) it is not influenced by the tracked objects scale changes; 3) it is less prone to the background clutter.

**Keywords**—object tracking; mean shift; background information; sum of gradient mode

## I. INTRODUCTION

In computer vision, tracking refers to the task of generating the trajectories of the moving objects by computing its motion in a sequence of images. Object motion recorded in the form of a trajectory commonly contains translational motion [1]. Numerous approaches have been dedicated to computing the translation of an object in consecutive frames, among which the mean shift (MS) method is one of the most common methods. The MS algorithm was originally proposed by Fukunaga in 1975 [2], and the theory was developed by Cheng 20 years later, with the concept of kernel function, which became the practical foundation of MS [3]. Later, Comaniciu imported the MS to visual object tracking field satisfactorily [4,5]. The popularity of the MS method is due to its ease of implementation, real time response and robust tracking performance.

However, the kernel bandwidth in traditional MS algorithm is fixed, which cannot be adapt to the change of object size. Though valuable research results were achieved in adaptive kernel bandwidth, they mostly use Bhattacharyya coefficient as the similarity measurement to adjust the original kernel bandwidth [5,6]. But these methods perform satisfactorily only when object becomes smaller. Bhattacharyya coefficient may not be bigger with large window than with small window, so these methods do not work well under this circumstance when object size

becomes larger. Since Bhattacharyya coefficient based similarity measurement could not reflect the change of object size accurately, a novel kernel bandwidth adaptation algorithm is proposed in [7], in which Gaussian scale-space is introduced to object tracking field. And it is proved that, when the scale parameter of Gaussian scale-space increases, the sum of gradient mode (SGM) of all pixels in tracking window is non-increasing. So SGM can be employed as a measurement which can reflect the variation trend of object size.

The second limitation of the traditional MS method is prone to local minima when some of the target features present in the background. Therefore, in [5], Comaniciu *et al.* further proposed the background-weighted histogram (BWH) to decrease background interference in target representation. Such an idea is reasonable and intuitive, and some works have been proposed to follow this idea [8-10]. Unfortunately, all of them do not notice that the BWH transformation formula proposed in [5] is actually incorrect and this is proved in [11]. And the corrected background-weighted histogram (CBWH) is proposed, which can truly achieve what the original BWH method wants: reduce the interference of background in target localization.

This paper is organized as follows. Section II provides briefly an overview of the classical MS algorithm and introduces the CBWH scheme in detail. Section III analyzes updating kernel bandwidth according to the ratio of SGM in consecutive frame and describes the proposed object tracking algorithm using the CBWH algorithm and SGM. Finally, the experimental results and concluding remark are given in Section IV and V, respectively.

## II. THE CBWH MEAN SHIFT ALGORITHM

### A. Mean Shift Tracking: An Overview

MS is a semi-automatic tracking method. The desired object is manually selected in the initial frame, and it is shown as a rectangle. The inside pixel location of the rectangles is presented as  $\{x_i^*\}_{i=1,2,\dots,n}$ . The selected area is considered as the object model [12,13]. The color histogram of the object model is then calculated by:

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

where  $u$  represents the feature value of the target model,

$b(x_i^*)$  associates the pixel  $x_i^*$  to the histogram bin,  $k(x)$  is an isotropic kernel function and constant  $C$  is  $C = 1/\sum_{i=1}^n k\|x_i^*\|^2$ .

Similarly, the probability of the feature  $u = 1, 2, \dots, m$  in the target candidate model from the target candidate region centered at position  $y$  is given by:

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

Based on the Bhattacharyya coefficient, a metric which aims to estimate the similarity between the target model and the candidate model is defined as:

$$\rho(y) = \sum_{n=1}^m \sqrt{\hat{p}_u(y) q_u} \quad (3)$$

More similar between the two density distribution,  $\rho$  is greater. Find the maximum of the density estimation to obtain the new location of  $y$  by using MS vector.

$$y_1 = \frac{\sum_{i=1}^{n_k} x_i \omega_i g\left(\left\|\frac{y_0 - x_i}{h}\right\|^2\right)}{\sum_{i=1}^{n_k} \omega_i g\left(\left\|\frac{y_0 - x_i}{h}\right\|^2\right)} \quad (4)$$

where

$$\omega_i = \sum_{u=1}^m \delta[b(x_i) - u] \sqrt{\frac{q_u}{p_u(y_0)}}, \quad g(x) = -k'(x) \quad (5)$$

### B. The Corrected Background-Weighted Histogram Scheme

In target tracking, the background information is often included in the detected target region. If the correlation between target and background is high, the localization accuracy of the object will be decreased. To reduce the interference of salient background features in target localization, a representation model of background features was proposed by Comaniciu [5] as background-weighted histogram (BWH) algorithm, which selects discriminative features from the target region and the target candidate region.

In [5], the new target model is:

$$\hat{q}'_u = C' v_u \sum_{i=1}^n k\left(\left\|x_i^*\right\|^2\right) \delta[b(x_i^*) - u] \quad (6)$$

And the new target candidate model is:

$$\hat{p}'_u(y) = C'_h v_u \sum_{i=1}^{n_h} k\left(\left\|\frac{y - x_i}{h}\right\|^2\right) \delta[b(x_i) - u] \quad (7)$$

where

$$\left\{v_u = \min\left(\frac{\hat{\sigma}^*}{\hat{\sigma}_u}, 1\right)\right\}_{u=1,2,\dots,m} \quad (8)$$

and  $\{\hat{\sigma}_u\}_{u=1,2,\dots,m}$  (with  $\sum_{i=1}^m \hat{\sigma}_u = 1$ ) is the background, which is three times the size of the target as suggested in [4].  $\hat{\sigma}^*$  is the minimal non-zero value in  $\{\hat{\sigma}_u\}_{u=1,2,\dots,m}$ .

The weight of point  $x_i$  computed by BWH in the target candidate region can be derived by (5) that

$$\omega'_i = \sum_{u=1}^m \sqrt{\frac{\hat{q}'_u}{\hat{p}'_u(y)}} \delta[b(x_i) - u] \quad (9)$$

The above BWH transformation aims to reduce the effects of prominent background features in the target candidate region on the target localization. However, in [11] prove that BWH cannot achieve this goal, because it is equivalent to the usual target representation under the MS tracking framework. Then the corrected BWH (CBWH) algorithm was proposed.

In CBWH, (8) is employed to transform only the target model but not the target candidate model. A new weight formula is defined as:

$$\omega''_i = \sqrt{\frac{\hat{q}'_u}{\hat{p}_u(y)}} \quad (10)$$

And simplify (10) as:

$$\omega''_i = \sqrt{v_u} \omega_i \quad (11)$$

Equation (11) clearly reflects the relationship between the weight calculated by using the usual target representation (i.e.  $\omega_i$ ) and the weight calculated by exploiting the background information (i.e.  $\omega''_i$ ). If the color of point  $i$  in the background region is prominent, the corresponding value of  $v_u$  is small. Hence in (11) this point's weight is decreased and its relevance for target localization 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  $\omega''_i$  will degrade to  $\omega_i$  with the usual target representation.

## III. OBJECT TRACKING

### A. The Sum of Gradient Mode

Bhattacharyya coefficient can not correctly reflect variation trend of object scale, when object becomes large and extends tracking window [7]. This will make tracking window deviating from object center.

**Theorem 1:** Within Gaussian scale-space, gradient mode of pixel will be non-increasing when scale parameter increases.

When Theorem 1 is applied in object tracking, it is supposed that scale parameter increases when object shrinks, and decreases when object becomes large. So when object

size shrinks, SGM of pixels in tracking window decreases monotonously. It is clear that SGM increases when object becomes larger and decreases when object becomes smaller.

In two dimensional spaces, gradient mode of pixel  $x_i$  and SGM for image are defined respectively as:

$$M(x_i) = \sqrt{m_x^2(x_i) + m_y^2(x_i)} \quad (12)$$

$$SGM = \sum_{i=1}^N M(x_i) \quad (13)$$

where  $m_x(x_i)$  and  $m_y(x_i)$  are horizontal and vertical gradient and  $N$  is the amount of pixels in image.

#### B. The Tracking Algorithm with CBWH and SGM

The SGM of tracking window in two adjacent frames  $n$  and  $n+1$  is denoted as  $SGM_n$  and  $SGM_{n+1}$ , and then the kernel bandwidth adaptive algorithm is formulated by

$$W_{n+1} = W_n \sqrt{\frac{SGM_{n+1}}{SGM_n}} \quad (14)$$

$$H_{n+1} = H_n \sqrt{\frac{SGM_{n+1}}{SGM_n}} \quad (15)$$

where  $W_n$ ,  $H_n$  and  $W_{n+1}$ ,  $H_{n+1}$  represent width and height of kernel bandwidth in frame  $n$  and  $n+1$ .

Based on the above analyses, the scale of the target can be estimated and then a scale adaptive MS tracking algorithm, i.e. the SGM-CBWH-MS algorithm, can be developed. The implementation of the whole algorithm is summarized as follows:

- 1) Calculate the target model  $\hat{q}$  by (1) and the BWH  $\{\hat{o}_u\}_{u=1, \dots, m}$ , and then compute  $\{v_u\}_{u=1, \dots, m}$  by (8) and the transformed target model  $\hat{q}'$  by (6). Initialize 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 (2) in the current frame.
- 4) Calculate the weights  $\{\omega_i^n\}_{i=1, \dots, n_h}$  according to (11).
- 5) Calculate the new position  $y_1$  of the target candidate region using (4).
- 6) Let  $d \leftarrow \|y_1 - y_0\|$ ,  $y_0 \leftarrow y_1$ ,  $k \leftarrow k+1$ . Set the error threshold  $\varepsilon_1$  (default value: 0.1), the maximum iteration number  $N$ .
- If  $d < \varepsilon_1$  or  $k > N$  Stop and go to step 7;
- Otherwise  $k \leftarrow k+1$  and go to step 3.
- 7) Calculate SGM of new tracking window and adapt the kernel bandwidth using (14) and (15). Then jump to step 3.

## IV. EXPERIMENTS

This section evaluates the proposed SGM-CBWH-MS algorithm in comparison with the MS algorithm [4], the SGM-MS algorithm [7] and the CBWH-MS algorithm [11]. Experiments are carried out on two sequences clipped from PETS dataset with software environment MATLAB 7.01. The resolution of video is 720×576 and frame rate is 25fps. In the experiments, the RGB color space was used. Each color band was equally divided into 16 bins (16×16×16).

The first clipped video (141 frames) is a car sequence where the object scale (a white car) increasing gradually as shown in Fig. 1. Neither the MS algorithm nor the CBWH-MS algorithm achieves good tracking results. On the other hand, the car is well tracked by both SGM-MS and SGM-CBWH-MS in the sequence. However, towards the latter half of the sequence, such as in frames 87 and 120, the SGM-MS and SGM-CBWH-MS algorithm estimate the target scale more perfectly fit on the car than MS and CBWH-MS, which due to the car scale getting much larger than the initial tracking window.

The second experiment is on the same object (the white car) from another camera angle with a complex background (186 frames) in where the object still exhibits scale zooming-in. Since the complicated background, the tracking results are much more different. As can be seen in Fig. 2, both CBWH-MS algorithm and SGM-CBWH-MS algorithm can track the target over the whole sequence, but neither MS algorithm nor SGM-MS algorithm. However, the SGM-CBWH-MS scheme works much better in tracking the scale zooming-in target, especially in the frames with background getting similar to the object or complicated.

## V. CONCLUSIONS

In this paper, a novel kernel bandwidth adaptive algorithm for object tracking has been proposed with a combination of CBWH-MS and SGM. The proposed algorithm is employed to provide an optimum solution to object tracking problems. The traditional MS algorithm use fixed kernel bandwidth, which may end up tracking failure because kernel bandwidth can't be adapted to size change of the object, especially in the complex environment. The CBWH could reduce the relevance of background information and improve the target localization. Gaussian scale-space is imported to object tracking, and the SGM of all pixels in tracking window is non-increasing when scale parameter increases, based on which to achieve the bandwidth adaptive. The experimental results show SGM-CBWH-MS algorithm can reliably track the object with scale changes and complicated background in video sequences.

## ACKNOWLEDGMENT

This work is supported by College of Sciences under the youth research fund (KYLZY001232).

## REFERENCES

- [1] Alper Yilmaz, "Object tracking by asymmetric kernel mean shift with automatic scale and orientation selection," Proc. IEEE Conf. Computer Vision and Pattern Recognition, 17(2), June 2007 pp.1– 6.

- [2] K. Fukunaga, L. D. Hostetler, "The estimation of the gradient of a density function, with application in pattern recognition," *IEEE Trans. Information Theory*, 1975, 21 (1), pp. 32–40.
- [3] Y. Cheng, "Mean-shift mode seeking and clustering," *IEEE Trans. Pattern Anal. Machine Intell.* 1995, 17 (8), pp. 790–799.
- [4] D. Comaniciu, V. "Ramesh, P. Meer. Real-time tracking of non-rigid objects using Mean-Shift," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2000, pp. 142–149.
- [5] D. Comaniciu, P. Meer, "Kernel-based object tracking," *IEEE Trans. Pattern Anal. Machine Intell.* 2003, 25, (2), pp. 564–577.
- [6] R. T. Collins, "Mean-Shift blob tracking through scale space," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2003, pp. 234–240.
- [7] Yundong Li, Weigang Zhao, "Adaptive updating of kernel band-width for mean-shift tracking," *Journal of Computational Information Systems*, 2012, 8(22), pp. 9569–9579.
- [8] Jeyakar J., Babu R., and Ramakrishnan K. R, "Robust object tracking with background-weighted local kernels," *Computer Vision and Image Understanding*, 2009, 112, (3), pp. 296–309.
- [9] Li L., and Feng Z., "An efficient object tracking method based on adaptive nonparametric approach," *Opto-Electronics Review*, 2005, 13, (4), pp. 325–330.
- [10] Allen J., Xu R., and Jin J., "Mean Shift Object Tracking for a SIMD Computer," *Proc. International Conference on Information Technology and Applications*. Sydney, Australia, vol. I, July 2005, pp. 692–697.
- [11] Jifeng Ning, Lei Zhang, David Zhang and Chengke Wu, "Robust mean shift Tracking with corrected background-weighted histogram," *IET Computer Vision*. vol. 6, January 2012 pp. 62–69.
- [12] Motavalli S., "A part image reconstruction system for reverse engineering of design modifications," *J Manuf Syst*, 1991, 10(5), pp.383–395.
- [13] Cheng Jin D, Da Wen, "Recent developments in the applications of image processing techniques for food quality evaluation," *Trends Food Sci. Tech.* 2004, 15(5), pp. 230–249.

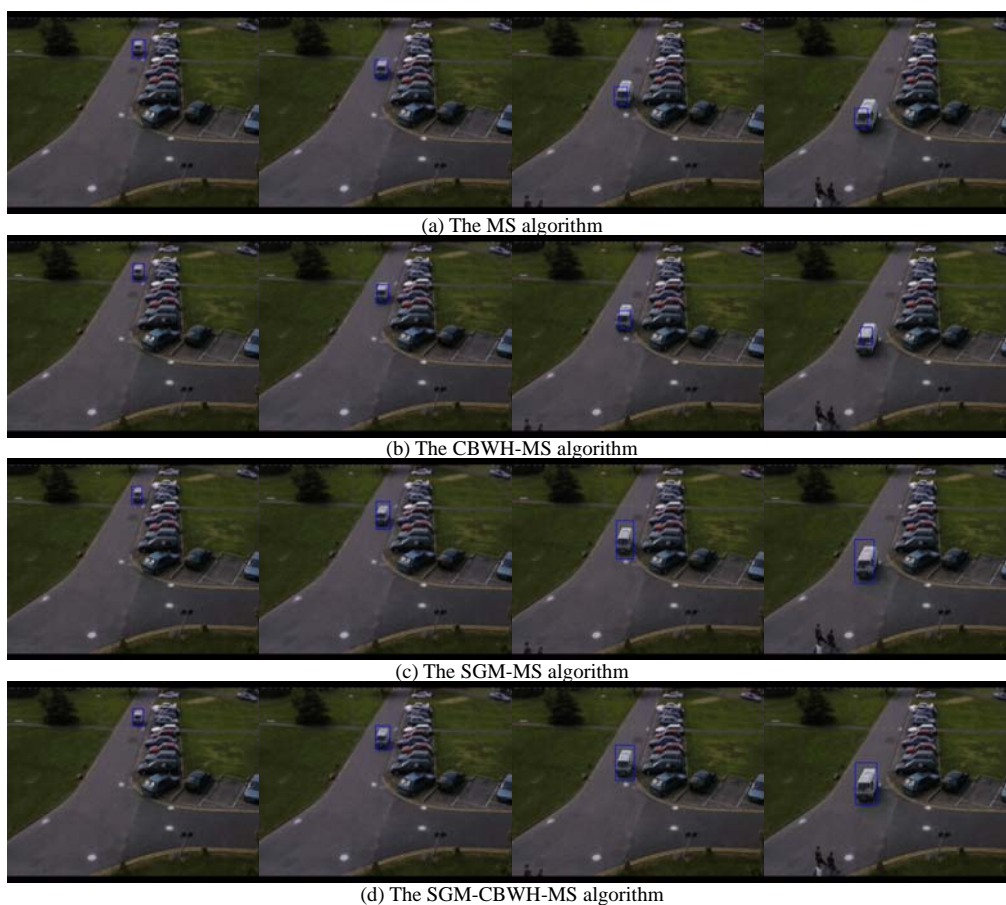


Figure 1. Tracking results of the car sequence 1 by different tracking algorithms. The frames 3, 47, 87, and 120 are displayed.



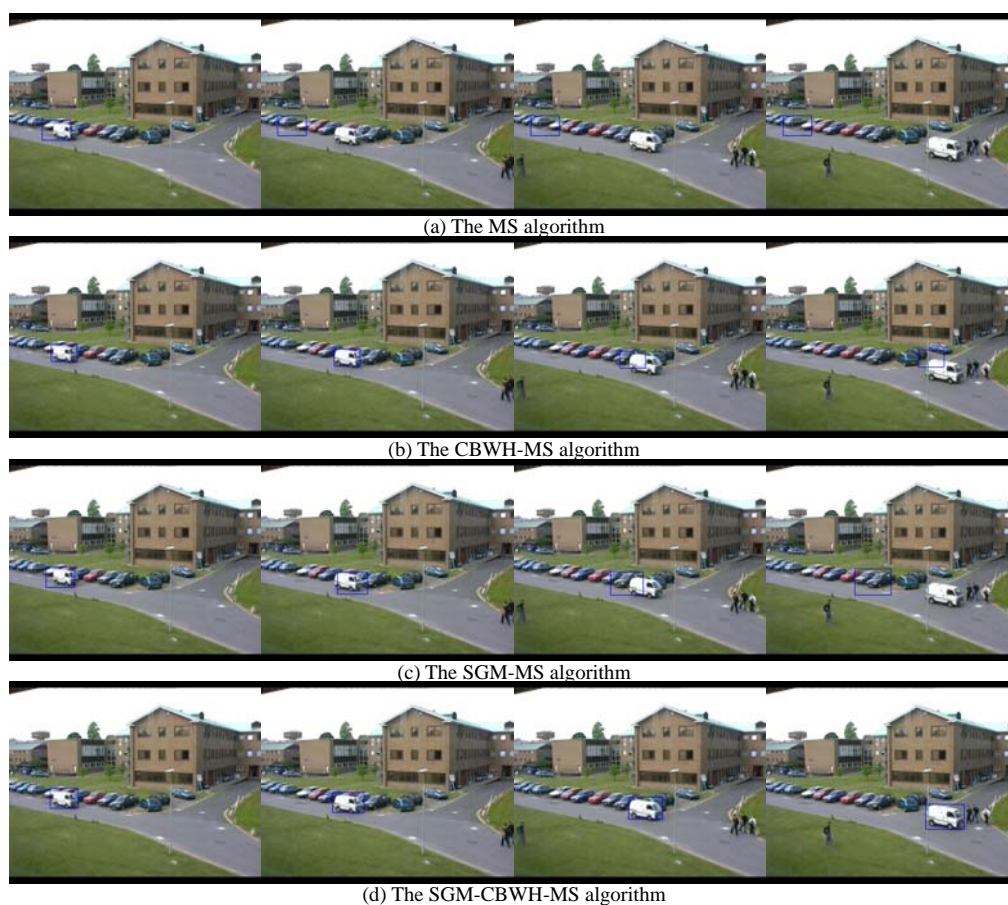


Figure 2. Tracking results of the car sequence 2 by different tracking algorithms. The frames 30, 60, 108, and 163 are displayed.