Adaptive Scale Mean-Shift Tracking with Gradient Histogram

Changqing Xie, Wenjing Kang, and Gongliang Liu$^{(\boxtimes)}$

School of Information and Electrical Engineering, Harbin Institute of Technology (Weihai), Weihai, China
`l5650l3l973@yeah.net`, {`kwjqq,liugl`}`@hit.edu.cn`

Abstract. The mean-shift (MS) tracking is fast, is easy to implement, and performs well in many conditions especially for object with rotation and deformation. But the existing MS-like algorithms always have inferior performance for two reasons: the loss of pixel's neighborhood information and lack of template update and scale estimation. We present a new adaptive scale MS algorithm with gradient histogram to settle those problems. The gradient histogram is constructed by gradient features concatenated with color features which are quantized into the $16 \times 16 \times 16 \times 16$ bins. To deal with scale change, a scale robust algorithm is adopted which is called background ratio weighting (BRW) algorithm. In order to cope with appearance variation, when the Bhattacharyya coefficient is greater than a threshold the object template is updated and the threshold is set to avoid incorrect updates. The proposed tracker is compared with lots of tracking algorithms, and the experimental results show its effectiveness in both distance precision and overlap precision.

Keywords: Object tracking · Mean-shift · Scale estimation · Gradient

1 Introductions

Visual object tracking has always been a challenging work especially in the sequences with the deformation and rotation. The MS tracking has an outstanding performance and is easy to implement. It tracks by minimizing the Bhattacharyya distance between two probability density functions represented by a target and target candidate histograms. The histogram is a statistical feature that does not depend on the spatial structure within the search window. This makes it more robust than other algorithms. But it lacks the essential template update and pixel's neighborhood information leading to a worse accuracy.

The mean-shift algorithm is a nonparametric mode-seeking method for density functions proposed by Fukunaga and Hostetler [1]. Comaniciu et al. [2, 3] use it to track object. And Comaniciu et al. change the window size over multiple runs by a constant factor but produces little effect because the smaller windows usually have higher similarity. The image pyramids and an additional mean-shift algorithm for scale selection had been used after estimating the position to confirm the window size in Collins [4]. But its speed is lower than the conventional MS algorithm. A new histogram that exploits the object neighborhood has been proposed to help discriminate

the target which is called background ratio weighting (BRW) in Vojir et al. [5]. This approach is faster than others and has a superior effect in sequences with scale change but performs poorly for grayscale sequences.

Gradient information is crucial for appearance representation since it contains pixel's neighborhood information and is insensitive to illumination variation but it always be ignored. Based on this observation, we present a novel adaptive scale MS algorithm with gradient histogram. The gradient information is calculated by Canny edge detector which was developed by Canny [6].

Moreover, the BRW algorithm has been used to improve the performance of videos with scale change. The template of target will be updated by liner interpolation only if conditions are met to avoid addition of incorrect information. The template update also can cope with appearance variation of target. The proposed tracker is compared with lots of algorithms, and the experiment results show that it is more robust and accurate.

2   Canny Edge Detector

The Canny edge detector is an edge detection operator that uses a multi-stage algorithm to detect a wide range of edges in images. To remove the noise, a Gaussian filter is applied to the image; the Gaussian filter kernel of size $(2k + 1) \times (2k + 1)$ is given by:

$$H_{ij} = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{(i-(k+1))^2 + (j-(k+1))^2}{2\sigma^2}\right); \ 1 \le i,\ j \le (2k+1) \qquad (1)$$

Let I_o denote the original image and H is a classic Gaussian filter matrix; we get the image without noise $I = H * I_o$. Then we extract intensity gradient of the image with the Sobel operator proposed by Sobel [7]. The gradient information in horizontal and vertical directions is G_x and G_y. From this, the edge gradient can be determined by

$$G = \sqrt{G_x^2 + G_x^2} \qquad (2)$$

where

$$G_x = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 2 \end{bmatrix} * I \ \text{and} \ G_y = \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix} * I. \qquad (3)$$

The edge extracted from the gradient value is still quite blurred after processing the G with Gauss filter and Sobel operator. The non-maximum suppression and double-threshold joint should be applied to the processed image to improve effect of gradient. Then we can get the gradient image I_g to track the target.
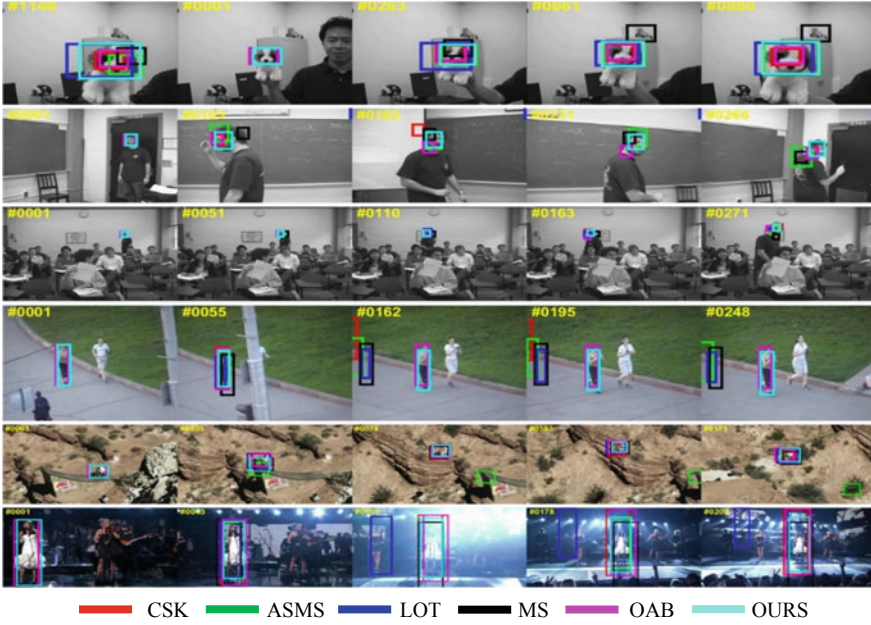
Fig. 1. Result of six trackers in six sequences. (The sequence names from top to bottom are dog1, freeman1, freeman3, jogging, mountainbike, and singer1)

3 The Tracking Algorithm

Different from the conventional MS tracking algorithm, we append the I_g to original image to get I_e. After the combination of the images, we extract the histogram $\hat{\mathbf{q}}$ from I_e. To cope with the problem which caused by the size of target changes, we use the BRW-MS instead of conventional MS. We can get $\hat{\mathbf{q}}$ from:

$$\hat{q}_u = C \sum_{i=1}^{N} k\left(\frac{\left(x_i^{*1}\right)^2}{a^2} + \frac{\left(x_i^{*2}\right)^2}{b^2}\right) \delta\big[b\big(x_i^*\big) - u\big] \tag{4}$$

and an ellipsoidal region is used to represent target $\frac{\left(x_i^{*1}\right)^2}{a^2} + \frac{\left(x_i^{*2}\right)^2}{b^2} < 1$ in current frame. The target candidate is given by

$$\hat{p}_u(\mathbf{y}, h) = C_h \sum_{i=1}^{N} k\left(\frac{\left(y^1 - x_i^1\right)^2}{a^2 h^2} + \frac{\left(y^2 - x_i^2\right)^2}{b^2 h^2}\right) \delta[b(\mathbf{x}_i - u)] \tag{5}$$

where the h is the scale factor. The location of the target is obtained by

$$\hat{y}_1^1 = \frac{1}{a^2} m_k^1(\hat{\mathbf{y}}_0, h_0) + \hat{y}_0^1, \ \hat{y}_1^2 = \frac{1}{b^2} m_k^2(\hat{\mathbf{y}}_0, h_0) + \hat{y}_0^2 \tag{6}$$

$$h_1 = \left[1 - \frac{A}{K}\right] h_0 + \frac{1}{h_0} \frac{B}{K} \tag{7}$$

where A is

$$\sum_{i=1}^{N} w_i k \left(\frac{(y_0^1 - x_i^1)^2}{a^2 h_0^2} + \frac{(y_0^2 - x_i^2)^2}{b^2 h_0^2} \right), \tag{8}$$

B is

$$\sum_{i=1}^{N} w_i \left(\frac{(y_0^1 - x_i^1)^2}{a^2} + \frac{(y_0^2 - x_i^2)^2}{b^2} \right) g \left(\frac{(y_0^1 - x_i^1)^2}{a^2 h_0^2} + \frac{(y_0^2 - x_i^2)^2}{b^2 h_0^2} \right). \tag{9}$$

And $g(x) = -k'(x)$ is the derivative of $k(x)$. w_i can be obtained by

$$w_i = \max(0, W) \tag{10}$$

where W is

$$W = \left(\sum_{u=1}^{m} \frac{1}{\rho[\hat{\mathbf{p}}(\hat{\mathbf{y}}_0, h_0), \hat{\mathbf{q}}]} \sqrt{\frac{\hat{q}_u}{\hat{p}_u(\hat{\mathbf{y}}_0, h_0)}} - \frac{1}{\rho\left[\hat{\mathbf{p}}(\hat{\mathbf{y}}_0, h_0), 3\widehat{\mathbf{bg}}\right]} \sqrt{\frac{\widehat{bg_u}}{\hat{p}_u(\hat{y}, h_0)}} \right) \delta[b(\mathbf{x}_i) - u] \tag{11}$$

The $\widehat{\mathbf{bg}}$ is the histogram of background computed over the neighborhood of the target in the first frame. Let us denote

$$K = \sum_{i=1}^{N} w_i g \left(\frac{(y_0^1 - x_i^1)^2}{a^2 h_0^2} + \frac{(y_0^2 - x_i^2)^2}{b^2 h_0^2} \right) \tag{12}$$

and

$$\mathbf{m}_k(\hat{\mathbf{y}}, h_0) = \frac{\sum_{i=1}^{N} x_i w_i g \left(\frac{(y_0^1 - x_i^1)^2}{a^2 h_0^2} + \frac{(y_0^2 - x_i^2)^2}{b^2 h_0^2} \right)}{K} - \hat{y}_0 \tag{13}$$

When the location was determined, the $\hat{q}$ is updated by

$$\hat{\mathbf{q}}_{\mathbf{new}} = \begin{cases} \hat{\mathbf{q}}_{\mathbf{old}} & \text{if } \rho[\hat{\mathbf{p}}(\mathbf{y}), \hat{\mathbf{q}}_{\mathbf{old}}] \le \alpha \\ (1 - \lambda)\hat{\mathbf{q}}_{\mathbf{old}} + \lambda \hat{p}_u(\mathbf{y}) & \text{if } \rho[\hat{\mathbf{p}}(\mathbf{y}), \hat{\mathbf{q}}_{\mathbf{old}}] > \alpha \end{cases} \tag{14}$$

4   Experiment

Experiments are conducted on sequences from Object Tracking Benchmark2013 (OTB2013) dataset [8]. The sequences in OTB2013 not only suffer the deformation but also have other change such as fast motion, background clutter, motion blur, and so on. So, we selected six sequences from OTB2013 dataset to show the results. We compared the proposed algorithm with conventional and state-of-the-art algorithms which are available as source code. They are conventional mean-shit algorithm [2], ASMS [5], OAB [9], LOT [10], and CSK [11]. The parameters for those algorithms are set default.

Figure 1 shows the result of six trackers in sequences. The score of distance precision (DP) rate, overlap success (OS) rate, and center location error (CLE) can be obtained from Table 1.

In general, the gradient histogram improves performance of MS algorithm for the gradient histogram. The data in Table 1 show that proposed algorithm has a higher score than others in DP, OS, and CLE which means our tracker is better than others. The sequence dog1 suffers the scale change, and results show standard MS failed to track target because it only uses the gray levels to calculate histogram and it lacks template update. But our tracker can deal well with this condition for the adopting of gradient histogram and RBW algorithm. The proposed tracker has a better performance than other trackers in sequence singer1 which suffers illumination variation since the addition of gradient histogram which does not dependent on the current pixel value but the difference of adjacent pixel. What is more we find that our tracker has a great improvement in gray image compared to the colorful image than the ASMS algorithm since the ASMS algorithm only can acquire information in the gray channel which is scanty.

Table 1. Scores of six trackers in six sequences

Sequence name		Dog1	Freeman1	Freeman3	Jogging1	Mountainbike	Singer1
Our tracker	**OS** rate (%)	*1*	*0.975*	*0.998*	*0.974*	*0.996*	*0.806*
	DP rate (%)	*1*	*0.837*	*0.304*	*0.958*	*0.965*	*0.299*
	CLE (pixel)	*3.18*	*6.23*	*6.57*	*8.1*	*8.36*	33.39
MS	**OS** rate (%)	0.224	0.423	0.598	0.15	0.75	0.14
	DP rate (%)	0.054	0.012	0.02	0.078	0.706	0.259
	CLE (pixel)	48.46	35.02	30.6	83.93	22.33	33.95
ASMS	**OS** rate (%)	0.842	0.696	0.696	0.228	0.132	0.792
	DP rate (%)	0.841	0.387	0.304	0.218	0.132	0.299
	CLE (pixel)	9.98	19.93	37.5	93.69	238.2	33.91
CSK	**OS** rate (%)	0.926	0.371	0.813	0.231	0.996	0.521
	DP rate (%)	0.59	0.224	0.165	0.208	0.925	0.296
	CLE (pixel)	8.91	182.64	33.94	112.91	8.49	17.45
LOT	**OS** rate (%)	0.507	0.11	0.593	0.192	0.715	0.165
	DP rate (%)	0.506	0.08	0.085	0.023	0.671	0.208
	CLE (pixel)	24.16	137.09	42.09	89.68	25.52	145.06
OAB	**OS** rate (%)	0.999	0.552	0.741	0.893	0.724	0.786
	DP rate (%)	0.583	0.236	0.122	0.827	0.684	0.248
	CLE (pixel)	7.36	18.09	51.7	13.9	13.85	*15.46*

(The best results are in italics)

5   Conclusion

In this paper, an adaptive scale mean-shift algorithm with gradient histogram has been proposed to improve the tracking performance of MS-like algorithms. The gradient histogram is constructed by color histogram and gradient feature calculated by Canny edge detector. To deal with the scale change, the RBW algorithm is adopted. Template update is used to cope with appearance variation when the Bhattacharyya coefficient between the current frame and the template is greater than the threshold. The setting of the threshold makes tracker more robust for incorrect information. The proposed tracker is compared with a lot of algorithms in OTB2013 dataset. The experiment results show the tracker's effectiveness in deformation, rotation, scale change, and illumination variation. Moreover, our tracker has a better preference in gray sequences than conventional MS-like algorithms.

References

 1. Fukunaga K, Hostetler L. The estimation of the gradient of a density function, with applications in pattern recognition. IEEE Trans Inf Theor. 1975;21(1):32–40.
 2. Comaniciu D, Ramesh V, Meer P. Real-time tracking of non-rigid objects using mean shift. In: Proceedings IEEE conference on computer vision and pattern recognition, 2000. IEEE; 2000, vol. 2. p. 142–9.
 3. Comaniciu D, Ramesh V, Meer P. Kernel-based object tracking. IEEE Trans Pattern Anal Mach Intell. 2003;25(5):564–77.
 4. Collins RT. Mean-shift blob tracking through scale space. In: Proceedings IEEE computer society conference on computer vision and pattern recognition, 2003. IEEE; 2003, vol. 2. p. II-234.
 5. Vojir T, Noskova J, Matas J. Robust scale-adaptive mean-shift for tracking. Pattern Recogn Lett. 2014;49:250–8.
 6. Canny J. A computational approach to edge detection. In: Readings in computer vision; 1987. p. 184–203.
 7. Sobel I. An isotropic 3×3 image gradient operator. In: Machine vision for three-dimensional scenes; 1990. p. 376–9.
 8. Wu Y, Lim J, Yang MH. Online object tracking: A benchmark. In: 2013 IEEE conference on computer vision and pattern recognition (CVPR). IEEE; 2013. p. 2411–8.
 9. Grabner H, Grabner M, Bischof H. Real-time tracking via on-line boosting. Bmvc; 2006, vol. 1, no. 5. p. 6.
10. Oron S, Bar-Hillel A, Levi D, et al. Locally orderless tracking. Int J Comput Vision. 2015;111(2):213–28.
11. Henriques JF, Caseiro R, Martins P, et al. Exploiting the circulant structure of tracking-by-detection with kernels. In: European conference on computer vision. Springer, Berlin, Heidelberg; 2012. p. 702–15.