Robust visual tracking based on multi-cue integration

Jiaqing Ma (马加庆)*[*] and Chongzhao Han (韩崇昭)

School of Electronics and Information Engineering, Xi'an Jiaotong University, Xi'an 710049, China

[*]*E-mail: jq_ma@126.com*

Received September 1, 2008

Traditional color-based mean shift tracking algorithm is unable to accurately track the object. To address this problem, we present an improved tracking algorithm. The improved tracker integrates the color and motion cues which characterize the appearance and motion information of the object, respectively. These two visual cues can complement each other and make for more precise target localization. Experiments show that the proposed tracking algorithm has better performance than the traditional mean shift tracker.

OCIS codes: 100.4999, 110.4155, 330.4150, 330.1710.

doi: 10.3788/COL20090705.0400.

Comaniciu presented a mean shift (MS) tracking algorithm that tracks the object by comparing the similarity between the reference color distribution of the target and the target's color distribution in the current frame[1]. Compared with the other tracking techniques, this algorithm was well-known for real-time computation and robustness against partial occlusion and view point changes due to its insensitivity to object appearance changes. In the past few years, several extensions to this algorithm have also been proposed to accommodate different tracking scenarios[2−6]. Despite its success in various scenarios, MS tracker does not perform well when objects have large variations in pose and appearance, or have a color similar to background. In these cases, the MS tracker often quickly drifts from target, and the tracking performance is severely affected. One way to alleviate the drift is multi-cue integration technology[7−10]. The idea behind this is very simple, i.e., when one cue does not work, another one may still work well, in other words, some of the visual cues can complement each other during the tracking process. Thus the multi-cue integration can improve the robustness of the tracker. Inspired by this idea, in this letter, we propose a robust mean shift (RMS) tracking algorithm based on the assumption that camera is static. In this tracker, motion cue based on histogramming successive frame difference is integrated into the traditional MS tracking framework, the color and motion cues can complement each other and make for more precise target localization during the tracking process. The contrast experiment results show that the proposed algorithm can significantly improve the robustness.

Denote the reference color histogram of the object as $\mathbf{q}^{\mathrm{c}} = \{q_u^{\mathrm{c}}\}_{u=1,\cdots,b_{\mathrm{c}}}$, and let $\{\mathbf{x}_i\}_{i=1,\cdots,n}$ be the pixel locations in the region with the object centered at $\mathbf{y}$. The color histogram[1] $\mathbf{p}^{\mathrm{c}}(\mathbf{y}) = \{p_u^{\mathrm{c}}(\mathbf{y})\}_{u=1,\cdots,b_{\mathrm{c}}}$ is then constructed as

$$p_u^{\mathrm{c}}(\mathbf{y}) = C_h \sum\nolimits_{i=1}^{n} k\left(\left\|\frac{\mathbf{y}-\mathbf{x}_i}{h}\right\|^2\right)\delta[b(\mathbf{x}_i)-u], \qquad (1)$$

where function $b(\cdot)$ maps the pixel location to the corresponding histogram bin, $k(\cdot)$ is the Epanechnikov kernel profile with radius h, δ is the Kronecker delta function, and C_h is a normalization constant.

Just as how the color cue can be used for tracking, one can use motion cues as well. Here we use the color histogram of absolute successive frame difference to describe the motion information. We call this histogram as motion histogram. Intuitively, if the region contains no movement, all the absolute frame difference measurements computed on successive pair of images will fall in the lower bins of histogram. When movement commences, the absolute frame difference measurements usually fall in all bins with no definitive pattern: uniform regions produce low absolute frame difference values, whereas higher values characterize the contours. The motion cue model is shown in Fig. 1. To accommodate these variations, we choose the reference motion histogram $\mathbf{q}^{\mathrm{m}} = \{q_u^{\mathrm{m}}\}_{u=1,\cdots,b_{\mathrm{m}}}$ to be uniform, namely, $q_u^{\mathrm{m}} = 1/b_{\mathrm{m}}$, $u = 1,\cdots,b_{\mathrm{m}}$. Let $\mathbf{p}^{\mathrm{m}}(\mathbf{y}) = \{p_u^{\mathrm{m}}(\mathbf{y})\}_{u=1,\cdots,b_{\mathrm{m}}}$ represent the motion histogram in region with object centered at $\mathbf{y}$, and $\mathbf{p}^{\mathrm{m}}(\mathbf{y})$ can be constructed like $\mathbf{p}^{\mathrm{c}}(\mathbf{y})$ in Eq. (1).

Now, with the given reference color histogram $\mathbf{q}^{\mathrm{c}}$ and initial location $\mathbf{y}_0$ of the object, the new location $\mathbf{y}_1$ can be obtained by maximizing the objective function

$$D(\mathbf{y}) = \lambda_{\mathrm{c}} B(\mathbf{q}^{\mathrm{c}}, \mathbf{p}^{\mathrm{c}}(\mathbf{y})) + \lambda_{\mathrm{m}} B(\mathbf{q}^{\mathrm{m}}, \mathbf{p}^{\mathrm{m}}(\mathbf{y})), \qquad (2)$$

where $B(\cdot)$ is the Bhattacharyya coefficient, λ_{c} and λ_{m} are weighted parameters ranging from 0 to 1, which

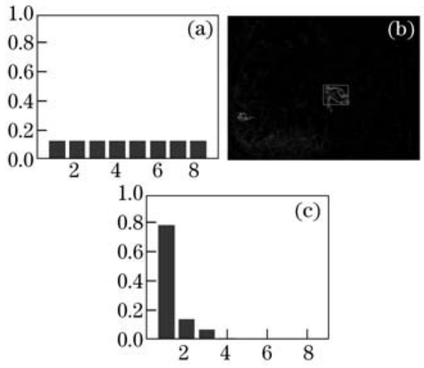

Fig. 1. Motion cue model. (a) Uniform reference motion histogram, (b) selected candidate motion region in frame difference image, and (c) motion histogram of the candidate motion region.

satisfy $\lambda_{\rm c} + \lambda_{\rm m} = 1$. By applying Taylor expansion at $\mathbf{p}^{\rm c}(\mathbf{y}_0)$ and $\mathbf{p}^{\rm m}(\mathbf{y}_0)$, we have

$$D(\mathbf{y}) \approx L(\mathbf{y}) = \lambda_{\rm c} L_{\rm c}(\mathbf{y}) + \lambda_{\rm m} L_{\rm m}(\mathbf{y}), \tag{3}$$

where

$$L_{\rm c}(\mathbf{y}) = \underbrace{\frac{1}{2}\sum\nolimits_{u=1}^{b_{\rm c}} \sqrt{p_u^{\rm c}(\mathbf{y}_0) q_u^{\rm c}}}_{c_1} + \frac{C_h}{2}\sum\nolimits_{i=1}^{n} \omega_i^{\rm c} k\left(\left\|\frac{\mathbf{y}-\mathbf{x}_i}{h}\right\|^2\right), \tag{4}$$

$$L_{\rm m}(\mathbf{y}) = \underbrace{\frac{1}{2}\sum\nolimits_{u=1}^{b_{\rm m}} \sqrt{p_u^{\rm m}(\mathbf{y}_0) q_u^{\rm m}}}_{c_2} + \frac{C_h}{2}\sum\nolimits_{i=1}^{n} \omega_i^{\rm m} k\left(\left\|\frac{\mathbf{y}-\mathbf{x}_i}{h}\right\|^2\right), \tag{5}$$

$$\omega_i^{\rm c} = \sum\nolimits_{u=1}^{b_{\rm c}} \sqrt{\frac{q_u^{\rm c}}{p_u^{\rm c}(\mathbf{y}_0)}}\delta[b(\mathbf{x}_i) - u],$$

$$\omega_i^{\rm m} = \sum\nolimits_{u=1}^{b_{\rm m}} \sqrt{\frac{q_u^{\rm m}}{p_u^{\rm m}(\mathbf{y}_0)}}\delta[b(\mathbf{x}_i) - u].$$

Introducing Eqs. (4) and (5) into Eq. (3), we can obtain

$$L(\mathbf{y}) = c_3 + \frac{C_h}{2}\sum\nolimits_{i=1}^{n} (\lambda_{\rm c}\omega_i^{\rm c} + \lambda_{\rm m}\omega_i^{\rm m}) \cdot k\left(\left\|\frac{\mathbf{y}-\mathbf{x}_i}{h}\right\|^2\right), \tag{6}$$

where $c_3 = \lambda_{\rm c} c_1 + \lambda_{\rm m} c_2$ is a constant.

Using the MS search procedure[1], $L(\mathbf{y})$ will be increased when the kernel is moved from the current location $\mathbf{y}_0$ to the new location $\mathbf{y}_1$ according to the relation

$$\mathbf{y}_1 = \frac{\sum\nolimits_{i=1}^{n} \mathbf{x}_i \omega_i g\left(\left\|\frac{\mathbf{y}_0-\mathbf{x}_i}{h}\right\|^2\right)}{\sum\nolimits_{i=1}^{n} \omega_i g\left(\left\|\frac{\mathbf{y}_0-\mathbf{x}_i}{h}\right\|^2\right)}, \tag{7}$$

where $g(x) = -k'(x)$, $\omega_i = \lambda_{\rm c}\omega_i^{\rm c} + \lambda_{\rm m}\omega_i^{\rm m}$. Set $\mathbf{y}_0 = \mathbf{y}_1$, and repeat this iterative procedure until $\|\mathbf{y}_1 - \mathbf{y}_0\| < \varepsilon$ or reach the maximal iterative number, where ε is a small predefined positive number. To handle scale variations, the radius h of $k(\cdot)$ is changed at ±10% in scale. Running the tracking algorithm to converge, the radius that makes $L(\mathbf{y})$ reach the maximum value is chosen.

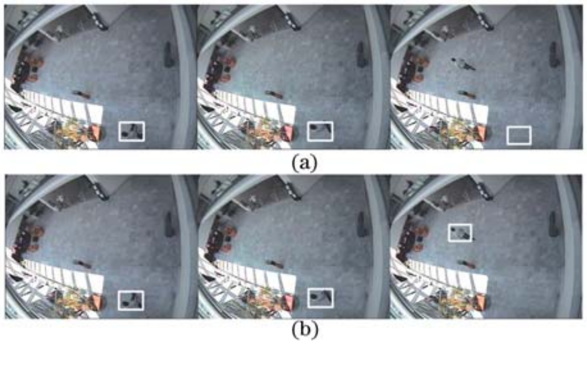

Fig. 2. Tracking results of the first sequence in frames 1, 8, and 78. Tracking results of (a) the traditional MS tracker and (b) the RMS tracker.

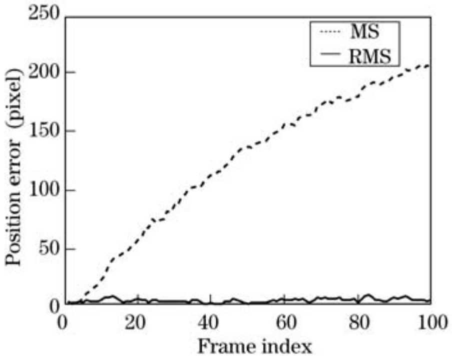

Fig. 3. Position error of the first sequence.

Two publicly available test sequences were used to evaluate the proposed tracking algorithm. For comparison, the traditional MS tracker was also utilized, and object in the first frame of each sequence was manually initialized. Parameters $\lambda_{\rm c}$ and $\lambda_{\rm m}$ were both set to be 0.5.

The first sequence contains one pedestrian walking from the bottom center to upper left, the challenge of this sequence is that the pedestrian has poorly distinguished color, meanwhile, the pedestrian cannot be modeled efficiently by a rectangle without considering rotation angle. As shown in Fig. 2, the traditional MS tracker begins to drift in fame 8, and finally loses the pedestrian. On the contrary, the proposed RMS tracker keeps tracking correctly throughout the entire sequence. Because in this case the color of pedestrian is similar to that of background, the color cue does not work well. However, the motion cue is very discriminant and can complement the deficiency of color cue. In order to further evaluate the RMS tracker, we define the position error as the Euclidean distance between the ground-truth of the object center location with the object center location obtained by the tracker. The quantitative comparison of position error for two trackers in this sequence is shown in Fig. 3.

The second sequence is one pedestrian walking from the bottom to the center, staying for a while

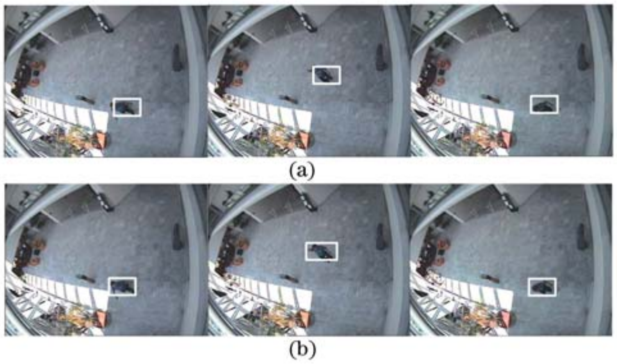

Fig. 4. Tracking results of the second sequence in frames 30, 92, and 117. Tracking results of (a) the traditional MS tracker and (b) the RMS tracker.

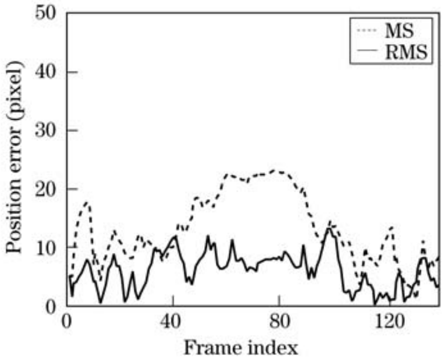

Fig. 5. Position error of the second sequence.

in the center location and then returning to the bottom. During this process, the pedestrian has large variations in pose. Some tracking results corresponding to MS and RMS trackers are shown in Fig. 4. It can be seen that the traditional MS tracker is unable to accurately track the pedestrian, while the RMS tracker keeps an accurate tracking. The reason is that the integration of two visual cues allows the object to be located with low ambiguity. The quantitative comparison of position error for two trackers in this sequence is shown in Fig. 5.

In conclusion, an improved MS tracking algorithm that integrates the color and motion cues is proposed. These two cues can complement each other during the tracking process. The contrast experiment results show that the proposed algorithm can significantly improve the performance of MS tracker.

This work was supported by the National " 973" Program of China under Grant No. 2007CB311006.

References

1. D. Comaniciu, V. Ramesh, and P. Meer, in *Proceedings of IEEE Conf. Computer Vision and Pattern Recognition 2000* **2,** 142 (2000).
2. R. T. Collins, in *Proceedings of IEEE Conf. Computer Vision and Pattern Recognition 2003* **2,** 234 (2003).
3. D. Comaniciu, V. Ramesh, and P. Meer, IEEE Trans. Pattern Anal. Mach. Intell. **25,** 564 (2003).
4. G. D. Hager, M. Dewan, and C. V. Stewart, in *Proceedings of IEEE Conf. Computer Vision and Pattern Recognition 2004* **1,** 790 (2004).
5. Z. Zivkovic and B. Kröse, in *Proceedings of IEEE Conf. Computer Vision and Pattern Recognition 2004* **1,** 798 (2004).
6. R. Han, Z. Jing, and Y. Li, Chin. Opt. Lett. **6,** 168 (2008).
7. E. Polat and M. Ozden, IEEE Trans. Multimedia **8,** 1156 (2006).
8. C. Shan, Y. Wei, T. Tan, and F. Ojardias, in *Proceedings of the Sixth International Conf. Automatic Face and Gesture Recognition* **1,** 669 (2004).
9. P. Pérez, J. Vermaak, and A. Blake. Proc. IEEE **92,** 495 (2004).
10. Z. Guan, Q. Chen, W. Qian, and Y. Hu. Acta Opt. Sin. (in Chinese) **28,** 860 (2008).