Robust kernel-based tracking algorithm with background contrasting

Rongli Liu (刘荣利)* **and Zhongliang Jing (敬忠良)**

School of Aeronautics and Astronautics, Shanghai Jiao Tong University, Shanghai 200240, China

**Corresponding author: liurl@sjtu.edu.cn*

Received May 25, 2011; accepted July 8, 2011; posted online August 24, 2011

The mean-shift algorithm has achieved considerable success in object tracking due to its simplicity and efficiency. Color histogram is a common feature in the description of an object. However, the kernel-based color histogram may not have the ability to discriminate the object from clutter background. To boost the discriminating ability of the feature, based on background contrasting, this letter presents an improved Bhattacharyya similarity metric for mean-shift tracking. Experiments show that the proposed tracker is more robust in relation to background clutter.

OCIS codes: 100.4999, 330.7310.

doi: 10.3788/COL201210.021001.

The mean-shift algorithm is a kernel-based iterative procedure, which climbs the local mode along the gradient ascent direction in feature space[1]. The method does not require prior knowledge, efficiency, and invariance to object deformation. Thus, it is suitable for the analysis of unknown data set. Recently, it has been widely applied in pattern classification[2,3], image segmentation[4], and visual tracking[5]. However, the traditional mean-shift tracker[5]has one critical inherent drawback: it can be only used to find the local mode. Consequently, it is likely to lose the target when a similar color distribution exists in the vicinity of the target. A considerable number of works have been dedicated to overcome the defects in the generic mean-shift trackers. The focus of these works varies from model updating, which accommodates appearance variation[6−9], bandwidth selection, which decides the size of the tracking window[10,11], spatial information[12−15], to the more discriminative distance metrics[16,17].

In this letter, we utilize the background information to obtain a background contrasting map, which is subsequently introduced into the traditional Bhattacharyya similarity metric. Thus, an improved similarity metric is proposed to boost tracking accuracy.

The mean-shift procedure seeks the local modes along the gradient ascent direction, which maximizes the Bhattacharyya similarity iteratively from a given starting position.

Given a current location y_{old}, the new position y_{new} is computed as

$$y_{\mathrm{new}} = \frac{\sum_i K(x_i - y_{\mathrm{old}}) w(x_i)(x_i - y_{\mathrm{old}})}{\sum_i K(x_i - y_{\mathrm{old}}) w(x_i)}, \tag{1}$$

where x_i represents the pixels inside the candidate region. The weight $w(x_i)$ of the pixel location x_i from the candidate region is defined as

$$w(x_i) = \sqrt{q_u/p_u}\big|_{b(x_i)=u}, \tag{2}$$

where $b : R^2 \longrightarrow \{1, 2, \cdots, m\}$ is the mapping of the color at x_i into histogram bin u with $u \in \{1, 2, \cdots, m\}$. $q = \{q_u\}_{u=1}^m$ and $p = \{p_u\}_{u=1}^m$ are the discrete color density estimation smoothed by the kernel function $K(\cdot)$ of the template and candidate regions, respectively. Here, $K(\cdot)$ can be chosen as uniform, Epanechnikov, or Gaussian.

$w(x_i)$ in Eq. (1) is derived from the first-order Taylor expansion of the Bhattacharyya similarity metric used in[5]

$$\rho(y) \equiv \rho[p(y), q] = \sum_u \sqrt{p_u(y), q_u}. \tag{3}$$

Using Taylor expansion around a point y_0, Eq. (3) can be approximated as

$$\rho[p(y), q] \approx \frac{1}{2} \sum_u \sqrt{p_u(y_0) q_u} + \frac{C_{\mathrm{h}}}{2} \sum_i w(x_i) K(y - x_i), \tag{4}$$

where C_{h} is a normalizing constant.

By using Eq. (1), the location estimation of the object can be performed iteratively. The kernel profile $k(\cdot)$ assigns a smaller weight to the locations farther from the center of the target. Although this can increase the estimation robustness, the mean-shift algorithm still converges to local mode, and the tracker may drift away from the object when a similar color is present in the vicinity.

To solve the background clutter problem, we present a background contrasting procedure that will be incorporated into the traditional Bhattacharyya similarity metric. The improved metric focuses on features that are more likely to be in the target object rather than in the background.

Generally, color histogram is a common feature used in mean-shift tracking. Firstly, we obtain the probability of some color v being in the object model. If we denote $H_{\mathrm{o}}(v)$ and $H_{\mathrm{s}}(v)$ as the non-normalized histograms of the object patch and its surrounding, including the object patch, respectively, the probability of the color bin v $p_{\mathrm{f}}(\mathrm{Obj}|v)$ in the model can be computed as

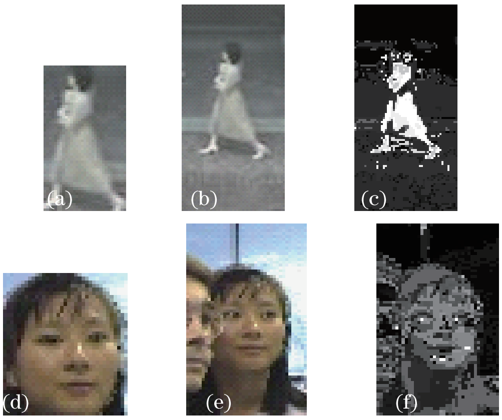

Fig. 1. Templates ((a),(d)), selected regions ((b),(e)), and corresponding probability maps ((c),(f)).

$$p_{\mathrm{f}}(\mathrm{Obj}|v) = \frac{H_{\mathrm{o}}(v)+1}{H_{\mathrm{s}}(v)+2}. \tag{5}$$

For convenience, we approximate $H_{\mathrm{o}}(v)$ by utilizing the non-normalized histogram of the template. Subsequently, $H_{\mathrm{s}}(v)$ is updated only during each mean-shift iteration. The probability maps computed according to Eq. (5) are shown in Fig. 1.

Introducing $p_{\mathrm{f}}(\mathrm{Obj}|v)$ into Eq. (3) yields

$$\widetilde{\rho}(y) = \widetilde{\rho}\left[p_c(y), p_m\right] \sum_y \sqrt{p_{\mathrm{f}}(\mathrm{Obj}|v)} \sum_u \sqrt{p_u(y) q_u}\,, \tag{6}$$

where $\widetilde{\rho}$ is a new Bhattacharyya similarity metric.

By denoting $\widetilde{w}(x_i) = \sqrt{p_{\mathrm{f}}[\mathrm{Obj}|I(x_i)]}\, w(x_i)$, Eq. (1) simply becomes

$$y_{\mathrm{new}} = \frac{\sum_i K(x_i - y_{\mathrm{old}}) \widetilde{w}(x_i)(x_i - y_{\mathrm{old}})}{\sum_i K(x_i - y_{\mathrm{old}}) \widetilde{w}(x_i)}. \tag{7}$$

Equation (6) emphasizes that the weights from colors are more likely to be in the object model than in the background. Thus, the tracker tends to follow features that are more discriminant from the background, that is, the tracker becomes more robust in relation to background clutter.

The tracking performance between the original mean-shift tracker and the proposed tracker is evaluated. For the background region, we use the area inside a rectangle which is slightly bigger than the target candidate window. In this letter, we consider objects with slight changes in scale, thus the size of the tracking window is set as a fixed value.

Figure 2 shows the tracking results of a woman wearing a white skirt and moving at the railway, where the color of the clothes is similar to the color of the crossing zebra. Figure 3 shows the tracking results of the face of woman in an indoor setting, where occlusion with a similar face happens. Figure 4 shows the tracking error versus the frame number. It demonstrates that the proposed approach succeeds during the tracking period. In Fig. 3, we observe that although some occlusions exist, the proposed tracker can always lock onto the object.

In conclusion, we propose an improved Bhattacharyya similarity metric based on background contrasting. The proposed method is more robust in relation to the background clutter problem. Experimental

Fig. 2. (Color online) Tracking results: (Top) proposed tracker. (Bottom) traditional mean-shift tracker. The tracking results are highlighted with a red box.

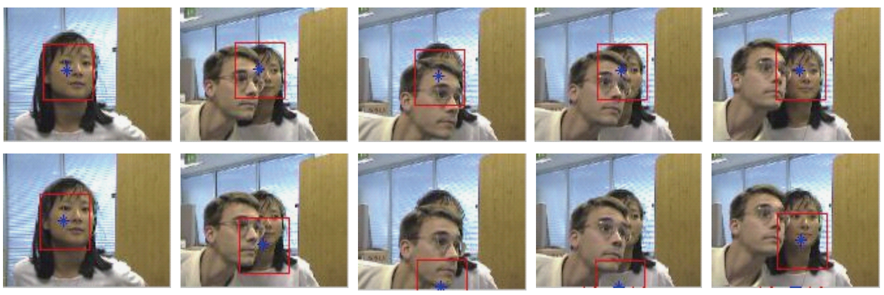

Fig. 3. (Color online) Tracking results: (Top) proposed tracker. (Bottom) traditional mean-shift tracker. The tracking results are highlighted with a red box.

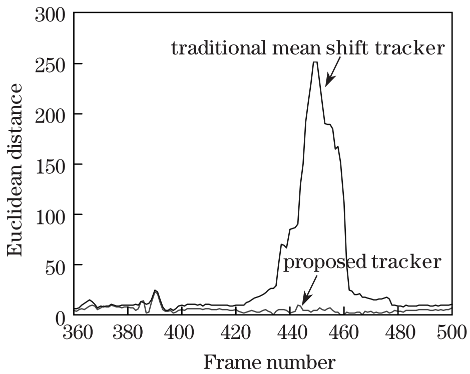

Fig. 4. Euclidean distance between the true and estimated positions.

results show that the proposed tracker has outperformed the traditional mean-shift tracker. These have highlighted the importance of scale estimation. As such, it will be the focus of our next work. We intend to obtain better results by incorporating background contrasting map into the kernel weighting.

This work was supported by the National Natural Science Foundation of China (No. 60775022) and the Ph.D. Programs Foundation of Ministry of Education of China (No. 20090073110045).

References

1. Y. Cheng, IEEE Trans. Pattern Anal. Mach. Intel. **17,** 790 (1995).
2. D. Comaniciu, IEEE Trans. Pattern Anal. Mach. Intel. **25,** 281 (2003).
3. B. Georgescu, I. Shimshoni, and P. Meer, in *Proceedings of Ninth IEEE International Conference on Computer Vision* **1,** 456 (2003).
4. D. Comaniciu and P. Meer, IEEE Trans. Pattern Anal. Mach. Intel. **24,** 603 (2002).
5. D. Comaniciu, V. Ramesh, and P. Meer, IEEE Trans. Pattern Anal. Mach. Intel. **25,** 564 (2003).
6. I. Leichter, M. Lindenbaum, and E. Rivlin, Computer Vis. Image Und. **114,** 400 (2010).
7. C. Shen, J. Kim, and H. Wang, IEEE Trans. Circ. Syst. Vid. **20,** 119 (2010).
8. P. Wang and H. Qiao, IEEE Trans. Circ. Syst. Vid. **21,** 156 (2011).
9. J. Tu, H. Tao, and T. Huang, in *Proceedings of International Conference on Computer Vision* **3851,** 694 (2006).
10. C. Shen, M. Brooks, and A. Van Den Hengel, IEEE Trans. Image Process **16,** 1457 (2007).
11. R. T. Collins, in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition* **2,** 234 (2003).
12. E. Maggio and A. Cavallaro, Computer Vis. Image Und. **113,** 544 (2009).
13. F. Wang, S. Yu, and J. Yang, Int. J. Electron. Comm. **64,** 614 (2010).
14. S. Birchfield and S. Rangarajan, in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition* **2,** 1158 (2005).
15. J. Jeyakar, R. V. Babu, and K. R. Ramakrishnan, Computer Vis. Image Und. **112,** 296 (2008).
16. C. Yang, R. Duraiswami, and L. Davis, in *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition* **1,** 176 (2005).
17. J. S. Hu, C. W. Juan, and J. J. Wang, Pattern Recogn. Lett. **29,** 2165 (2008).