

Object tracking Based on Local Structure Orientation

Ruxi Xiang^{1,2,3,a}, Xifang Zhu^{1,2,3,b}, Feng wu^{1,2,3}, Qinquan Xu^{1,2,3}

¹ College of Optoelectronic Engineering, Changzhou Institute of Technology,213002

² Changzhou Institute of Modern Optical Technology,213002

³ Changzhou Key Laboratory of Optoelectronic Materials and Devices,213002

^axiangrx@czu.cn ^bzhuxf@czu.cn

Keywords: Particle filter; SPBRLSO; visual tracking; spatiogram

Abstract. In order to solve the poor robustness resulted from the pose variances, occlusion or clutter in some challenging videos, we propose a robust particle filter tracking method based on a new similarity measure SPBRLSO (spatio-bin-ratio based on the local structure orientation). First, the object tracked is represented by a histogram of the local structure orientation, and then the observation likelihood function is constructed using a spatial-bin-ratio similarity. Finally, the tracker is formed by the he observation likelihood function in the frame of the traditional particle filter tracking. Many experimental results show that the proposed method is able to robustly track the objects in visible videos and infrared videos from some challenging videos including pose variances, occlusion or clutter. Furthermore, our tracking performance is superior to other tracking methods including the VTD tracker, conventional particle filter tracker and the spatiogram tracker.

1. Introduction

Object tracking is an important research field in computer vision, and widely used in video surveillance, robot localization, human-computer interaction and more. Recently, many algorithms have been proposed for tracking the objects in real world videos , which are typically categorized into two major groups deterministic method and stochastic method^[1].The mean shift tracker^[2] introduced by Comaniciu is a famous method among various deterministic methods^[2-5], which is efficiently computed. However, it is sensitive to partly occlusions and quick moving objects, which has been solved by stochastic methods^[6-8]. Particle filter (PF)^[6], which is simple, robust and effective, can be effectively track the object under some simple conditions such as few occlusion and gently moving among these stochastic methods. Appearance of the object and the similarity measure are very important for particle filter tracker.

Histogram is an effective representation of the object, and widely used in visual tracking. However, it could not effectively represent the object in the complex scenes as it discards spatial information of the object. To improve the representation of the object, many researchers have made many efforts^[5,9]. For example, spatiogram^[5] is a histogram with spatial information by adding the spatial mean and covariance of the pixel positions to the similarity measure. A fragment-based method^[4] has solved the problem of occlusions, in which the object divided into multiple fragments is represented using histogram with spatial information. A dissimilarity measure BRD (Bin-ratio-dissimilarity)^[10] that considers the ratios between bin values of histograms is proposed , and successfully used for category and scene classification .

Motivated by spatiogram^[5]and BRD^[10], we propose a new similarity measure SPBRLSO that considers both spatial information and the ratios between bin values from the histogram of the local structure orientation which effectively represent the object . The new similarity SPBRLSO is different from the spatiogram and the BRD. Compared with the spatiogram, we add the ratios between bin values of histogram to the spatiograms to improve the discriminative of the object. Meanwhile, our method are also different from the BRD, we add spatial information to the BRD for improving the discriminative of the similarity measure BRD. Due to the fact that the discriminative power of the SPBRLSO is stronger than the spatiogram and the BRD, we construct the stronger discriminative likelihood function by means of the SPBRLSO. Compared with spatiogram and VTD^[8]

(Visual Tracking Decomposition) algorithm, a lot of experimental results show that our method are able to successfully track the objects some infrared videos including clutter and in visible videos from some challenging videos including the pose variances, occlusion .

2. SPBRLSO tracker

2.1. HLSO feature

HOG^[11](Histogram of oriented gradient) is an effective and significant feature for representing the object, which is widely used in computer vision and pattern recognition. However the information of the traditional HOG is smaller because of the gradient of the object is computed via some simple operators such as Sobel, Prewitt et.al .To further improve the discriminate ability of the object, we introduce a new gradient value that relies on the first order of the horizontal and vertical neighbor pixels .Let f_x and f_y denote the first order derivative of the x and the y orientation of the gray image f ,respectively. The local structure matrix A which is a semi-positive definite matrix at the central pixel p is write as :

$$A(p) = \begin{bmatrix} E & F \\ F & G \end{bmatrix} \quad (1)$$

where, E ,F and G are respectively given by

$$\begin{aligned} E &= \sum_{q \in w_p} (f_x^q(p))^2 & G &= \sum_{q \in w_p} (f_y^q(p))^2 \\ F &= \sum_{q \in w_p} f_x^q(p) f_y^q(p), \end{aligned} \quad (2)$$

in which w_p indicates the window at the central pixel p of the image f . After computing the local structure matrix, the eigenvalue $\gamma_{1,2}$ of the $A(p)$ is computed by

$$\gamma_{1,2} = \left[E + G \pm \sqrt{(E - G)^2 + 4F^2} \right] / 2, \quad (3)$$

and the corresponding eigenvector is given by

$$g_o(\cos(\theta),\sin(\theta)), \quad (4)$$

where $g_o(.)$ is a normalized value which is similar to normalized gradient value ,and θ is represented using the partial derivative values E,F and G as

$$\theta = \frac{1}{2} \arctan \frac{2F}{E-G} \quad (5)$$

The histogram of the local structure orientation of the object which are divided into mxn same size of blocks is represented by s different scales. In this paper, we select s different scale(s is 1.0 ,0.8,0.6 and 0.4),4x4 blocks with 9 histogram bins which is quantized with respect to the histogram of the local structure orientation, and then these histograms are cascaded as ch.

2.2. SPBRLSO tracker

Due to discarding all spatial information of the object, BRD could not efficiently represent the object from complex conditions. To further improve the effect of the representation, we introduce a new similarity measure SPBRLSO by adding the spatial mean and covariance from the pixels to the BRD.

Let q_r represent a second-order *spatiogram* of a reference image, which is defined as $q_r(i) = [h_i, m_i, c_i]$, where h_i is the ith bin value of the histogram h, and m_i , c_i is the mean vector and covariance matrices of the spatial positions of all pixels belonged to the ith bin, respectively. Likewise, p_c denotes a second-order spatiogram of a candidate image, which is also expressed as $p_c(i) = [h'_i, m'_i, c'_i]$, where h'_i, m'_i and c'_i are the same meaning as h_i, m_i and c_i . All histograms need to be normalized by L_2 norm. Let $d(q_r, p_c)$ denote the similarity measure between q_r and p_c , which is computed by the following formula

$$d(q_r, p_c) = \sum_{i=1}^U \lambda_i d(q_r^i, p_c^i) \quad (6)$$

where λ_i is the weight of the ith bin, which is defined as

$$\lambda_i = \alpha \exp(-\frac{1}{2}(m_i - m'_i)^T \Sigma_i^{-1} (m_i - m'_i))$$

where the factor α satisfies $0 \leq d(q_r, p_c) \leq 1$ and $d(q_r, q_r) = 1$ for any q_r ⁹, and Σ_i^{-1} also satisfies $\Sigma_i^{-1} = k(c_i^{-1} + c_i'^{-1})^1$. In the formula (4), $d(q_r^i, p_c^i)$ is a similarity measure between the i^{th} bin of the histograms h and h' , which is computed by the formula (7).

$$d(q_r^i, p_c^i) = \frac{\|h + h'\|_2^2 \sum_{u=1}^U \frac{h_u^b h_u'^b}{(h_u^b + h_u'^b)^2}}{N_e} \quad (7)$$

where U is the total number of bins, and N_e is the number of bins and each pair of bins has at least one non-zero value in the two histograms h and h' . If the spatiogram has only the color histogram, the SPBRLSO becomes the BRS.

The observation likelihood based on SPBRLSO is expressed as

$$\omega_t^k = p(Y_t | X_t^k) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1-d_k(q_r, p_c)}{2\sigma^2}}, \quad (8)$$

Where σ is a control parameter, and $d_k(q_r, p_c)$ is the similarity measure corresponding the k^{th} particle. As it can be seen from the formula (5), the weight of the k^{th} particle is directly related to the $d_k(q_r, p_c)$. Therefore, the bigger the similarity measure is, the more reliable sample is. According to the likelihood function, the current prediction state \hat{X}_t can be computed by

$$\begin{aligned} \hat{X}_t &= \sum_{k=1}^N X_t^k \omega_t^k \\ &= \sum_{k=1}^N X_t^k p(Y_t | X_t^k) \\ &= \sum_{k=1}^N X_t^k \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1-d_k(p,q)}{2\sigma^2}} \end{aligned} \quad (9)$$

From the formula (6), we can see that the discriminative ability of the similarity measure is very important because it can affect the accuracy of the state.

3. Experiments

We test our SPBRLSO tracker on several challenging video sequences including visible videos and infrared videos². In all experiments, a reference target is manually selected in the initial frame. The color space used in SPBRLSO tracker is **HSV** space. For the spatiogram tracker^[5], the color space used is also HSV space. The scale which may produce the maximum score is selected when scale changes are limited to $\pm 10\%$ in each video frame. The number of the particles is set to 50. VTD used the software of authors. All experiments are carried out by processor with a core 4 Duo 2.9 GHz and 2 GB RAM under Matlab2009a.

3.1. Comparing results in visible sequences

Fig.2 demonstrates the tracking results in basketball and skating sequences, where the objects tracked often make pose variation or are severely occluded by other object. While the object is severely occluded, or the appearance of the object is severely varied due to quick moving, the proposed method can robustly track the objects due to the fact that it adopts a stronger similarity measure SPBRLSO, as shown by red rectangles in Fig.3. The VTD tracker^[8]can also robustly track the objects because it constructs multiple observation models, as shown by green rectangles. However, the Spatiogram tracker^[5]fails to track the object with severe pose changes or severely occluded, as demonstrated by blue rectangles. Meanwhile, we can observe that our method and VTD produce smaller central mean errors from the table 1. However the central error of the spatiogram tracker is bigger. The speed of VTD takes 1~3 seconds per frame due to its high complexity, but our method takes only 0.04~0.1 seconds per frame, and the spatiogram tracker takes only 0.06~0.12 seconds per frame.

¹ In this paper, k is set to 4.

² <http://cv.snu.ac.kr/research/~vtd> , <http://www.cvg.rdg.ac.uk/PETS2006/data.html> and <http://www.cs.technion.ac.il/~idol/>



Fig2. Tracking result comparisons among our method, VTD tracker and the spatiogram tracker in basketball and skating sequences under severely pose variations and the occluded. The (a) shows the results in basketball sequence, and (b) shows the results in skating. Note that the object contained by red rectangle is the results of our method, green rectangle is the results of the VTD, and blue rectangle is the results of the spatiogram tracker.

Table1 Average Centre Position Errors in Fig.3

Video Clip	spatiogram tracker	VTD tracker	our tracker
basketball	36.87	7.23	10.37
skating	40.54	8.13	12.97

3.2. Comparing results in infrared sequences

The proposed method can track the object not only in visible sequences but also in infrared sequences. Here we show two experiments. One is the boat tracked from in sea and the other is the plane tracked from in air, both of which are affected by the clutter. In this subsection, we demonstrate the performance of the SPBRLSO tracker, the spatiogram tracker and the BPF tracker in infrared sequences. In Fig.3, we observe that three methods also track the objects, but our mean central error is the smallest ,as is shown in table 2.On the contrary, the mean central of the BPF tracker is biggest because of discarding the spatial information of the object tracked.

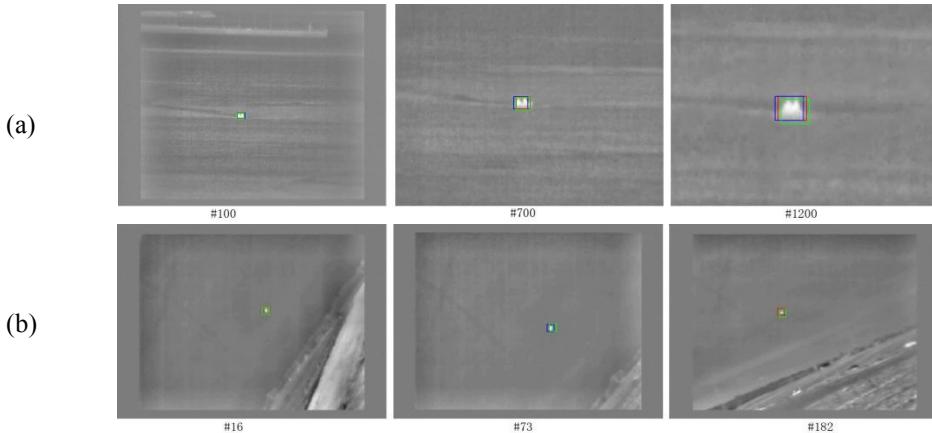


Fig.3 Tracking result comparisons among our method, BPF tracker and the spatiogram tracker in infrared sequences under heavy clutter. The (a) shows the results in boat sequence, and (b) shows the results in flying sequence. Note that the object contained by red rectangle is the results of our method, green rectangle is the results of the BPF, and blue rectangle is the results of the spatiogram tracker.

Table 2 Average Centre Position Errors in Fig.3

Video Clip	BPF	spatiogram tracker	HLSO tracker
boat	15.23	7.22	3.15
fly	10.97	6.12	4.54

4. Conclusions

We have proposed an effective tracking algorithm based on the new feature HLSO and a new similarity SPBRS which is more discriminative than the Bhattacharyya similarity measure, the BRS and the spatiogram. Compared with the conventional particle filter tracking, VTD and spatiogram tracker algorithm, our method outperforms the conventional particle filter and the spatiogram algorithm under challenging environments, and the tracking performance approximates VTD, but its speed is greatly faster.

Acknowledgment

This work was supported by the Research Foundation of Changzhou Institute of Technology (Grant No. YN1208), the Natural Science Foundation of China (61475027), Collaborative innovation fund of Jiangsu province (XYN1408) and the Research Foundation of Changzhou Institute of Modern Optical Technology (CZGY005)

References

- [1] H. Wang, D. Suter, K. Schindler and C. Shen, Adaptive object tracking based on an effective appearance Filter. *IEEE Transaction Pattern Analysis and Machine Intelligence*,2007, 29(9),1661-1667.
- [2] Comaniciu, D.; Ramesh, V.; Meer, P. Kernel-based object tracking. *IEEE Transaction Pattern Analysis and Machine Intelligence* ,2003, 25(5), 564-577.
- [3] Ruxi Xiang, Jianwei Li, Xuchu Wang, Fragments based tracking method with robust and adaptive mean shift, *ICIC-EL-An International Journal of Research and Surveys*,2011,5(12):4535-4541.
- [4] Wang Fanglin S J.Robust and efficient fragments-based tracking using mean shift.*AEU International Journal of Electronics and Communications*, 2010: 614-623.
- [5] Birchfield S T, Rangarajan S. Spatial histograms for region-based tracking. *ETRI Journal*, 2007,29(5), 697-699.
- [6] Isard, M.; Blake, A. Condensation—conditional density propagation for visual tracking. *International Journal of Computer Vision* 1998, 29, 5-28.
- [7] Fan Yang, chuan Lu andMing-Hsuan Yang. Robust superpixel tracking. *IEEE Transactions on Images Processing*,2014,23(4):1639-1651.
- [8] Junseok Kwon, Kyoung Mu Lee. Visual tracking decomposition. *Visual tracking decomposition. IEEE Conference on Computer Vision and Pattern Recognition* 2010, 1269-1276.
- [9] Shengfeng He, Qingxiong Yang, and Rynson W.H. Lau, J. Visual tracking via locality sensitive histograms, *Visual tracking decomposition IEEE Conference on Computer Vision and Pattern Recognition* 2013, 2427-2434 .
- [10] Xie N, Ling H, Hu W, et al. Use bin-ratio information for category and scene classification. *IEEE Conference on Computer Vision and Pattern Recognition* 2010, 2313 - 2319.
- [11] N.Dalal and B.Triggs. Histogram of oriented gradients for human detection. *IEEE Conference on Computer Vision and Pattern Recognition* 2005,886-893.