

Research on Dynamic Graph Target Tracking Method Fusing the Color Local Entropy

Junchang Zhang^{1,2*}, Chenyang Xia¹, Leili Hu², and Yanling Zhou¹

¹School of Electronics and Information, Northwestern Polytechnical University, Xi'an 710072, China

²Department of Luoyang Institute of electro-optical devices, Luoyang Henan 471009, China

Abstract. Focusing on the problems of target deformation, occlusion, background interference and rotation, a robust video tracking method is proposed in this paper, which is based on the superpixels and dynamic graph matching. Firstly, to make the superpixels edge fit better and structure tighter, the local gradient feature is fused into the simple linear iterative clustering (SLIC) method. Secondly, the candidate target superpixels set is generated by Graph Cuts and to obtain more accurate foreground superpixels set, the LASVM classification results are fused into the Graph Cuts energy function. Thirdly, in order to make the proposed tracker more robust, the color local entropy is fused into the diagonal elements of the affinity matrix. Experiment results show that the proposed algorithm has strong robustness and better tracking accuracy.

1 Introduction

Video tracking has become an important research direction in the field of computer vision. It is used in the field of human-computer interaction, video surveillance and human behavior analysis widely and has important application value in civil and military fields. According to the tracking principle, the video tracking algorithms can be considered as three aspects: template matching algorithm, state estimation algorithm and classification problem algorithm.

The template matching algorithm locates the target by finding the similarity between the target model and the candidate model according to the criterions. The classic template matching algorithm is the Mean Shift algorithm [1]. It has a better performance in a simple tracking scene. But the tracking performance becomes worse rapidly when target moves rapidly in complex scene. To reduce the background interference caused by the symmetric kernel, Chen [2] construct the asymmetric nuclear according to the segmentation of the target character. It can reduce the background interference and improve the tracking efficiency. To solve the problem of severe occlusion and do not reduce the real time, Xu [3] combined the weight of the over-segmentation. To reduce the interference of similar background, Li [4] introduced the target and the environment ratio coefficient to ensure the accurate update of the target model. The state estimation algorithm predicts the object state by using the state equation, and then extracts the real state information of the target. Particle filter is one of the typical algorithms, but the calculation is time-consuming, and is easy to occur the parti-

* Corresponding author: zhangjc@nwpu.edu.cn

cle degradation phenomenon. Then, Okuma⁴ proposed the Boosted particle filter algorithm. The classification tracking algorithm transforms the tracking problem into classification problem through training samples. It is common to estimate the target by the support vector machine. Then, Avidan proposed the support vector tracking and set tracking. Grabner⁵ fused the semi-supervised learning algorithm and boosting algorithm, and proposed the Semi-Boost tracking algorithm. It could solve the tracking drift problem. Babenko⁶ proposed multiple instance learning (MIL) algorithm, which is used to solve the problem that the training sample size is small or the sample is blur.

Only a few people have considered the occlusion and deformation problems at the same time. The performance of the mentioned algorithms would become poor when deal with the deformation and occlusion problem at the same time. However, Cai⁷ proposed the dynamic graph tracking (DGT) algorithm that has a good performance in dealing with deformation and occlusion problem. The classical SLIC method ⁸ is easy to generate the error segmentation at the edge pixels when the target is similar to the background in the DGT algorithm. Meanwhile, the color histogram is sensitive to similar background interference and illumination in the DGT algorithm. Therefore, the paper proposed an improved video tracking method to solve the above problems. It fuses the local gradient information into the SLIC method and combines the color local entropy ⁹ with the HSV color feature. The improved algorithm is more robust than the DGT algorithm.

2 DGT tracker

The DGT algorithm includes superpixels generation, image segmentation, graph model construction, graph matching and model update. The diagram is shown in Figure 1.

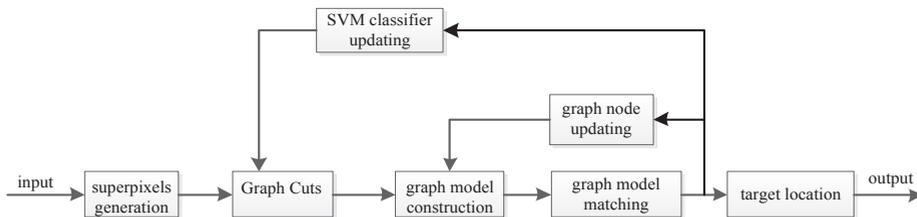


Fig. 1. The schematic diagram.

Firstly, the superpixels are generated by the SLIC algorithm which fuses the local gradient information. Then the superpixel is taken as a basic unit for the subsequent processing. Secondly, to obtain better segmentation results, the results of LASVM classification is fused into the Graph Cuts energy function. The DGT tracker takes the superpixels as foreground whose label value is 1 and the superpixels as background whose label value is 0 when the energy function is minimized. Thirdly, the target graph model and candidate graph model are constructed by ϵ -neighborhood way. Both structure information and appearance information can be exploited sufficiently in the affinity matrix. Fourthly, the matching result is achieved by spectral method. And then the target location is estimated according to the matching result. Finally, target graph model is updated according the matching result in real-time.

3 The improved SLIC algorithm

It is worth to notice that the SLIC method can generate superpixels according to the number of superpixels specified by users. And the pixel feature is expressed by the lab color space and the location information in the SLIC method. Then pixels will be classified as different

superpixels according to the similarity of features, so the feature accuracy is crucial to the superpixel segmentation. When the target and background are similar, they are easy to be considered as the same area. Therefore, in order to improve the accuracy of the segmentation, we fused the 3×3 neighborhood gradient feature into the distance formula. Adding the local gradient can help the superpixels move slowly in the boundary, and the boundary fitter. So we can obtain more accurate superpixels boundary. The proposed method is better than the conventional SLIC method. The local gradient formula can be expressed as (1).

$$gradient = \frac{1}{N} \sum_q \frac{(f(q) - f(p))^2}{dist(p, q)} \tag{1}$$

Where $f(\cdot)$ is the pixel value, $dist(p, q)$ is the distance between center point p and pixel point q . N is normalization constant.

$$N = \sum_q \frac{1}{dist(p, q)}$$

The distance can be expressed as (2):

$$D_s = \sqrt{d_{lab}^2 + \frac{m^2}{S^2} d_{xy}^2 + \lambda * gradient} \tag{2}$$

Where λ is a weight parameter.

The set $\{T_i\}$ is obtained by the SLIC algorithm. Then the candidate object superpixels set $\{T_i\}_{i=1}^{n_0}$ can be obtained by Graph Cuts algorithm. The energy function can be expressed as follows:

$$E(B) = \sum_{p \in S} D_p(b_p) + \sum_{p, q \in N} V_{p, q}(b_p, b_q) \tag{3}$$

Where $B = \{b_p | b_p \in \{0, 1\}, p \in S\}$ is the label of $\{T_p\}$. $\sum_{p \in S} D_p(b_p)$ is the regional energy function term. $V_{p, q}(b_p, b_q)$ is the boundary energy function term. The definition of $\sum_{p \in S} D_p(b_p)$ can be found in the DGT algorithm [11]. The definition of $V_{p, q}(b_p, b_q)$ can be redefined as follows:

$$V_{p, q}(b_p, b_q) = \exp(-2 * sqrt(abs(H_p - H_q))) \tag{4}$$

Where, p and q are neighboring superpixels. H is the value of the superpixel color local entropy.

The boundary energy function shows the segmentation boundary information in the physical sense. The color local entropy is more effective than the color feature in capturing the discontinuity between the neighborhood superpixels. When the value of color local entropy of p and q are similar, the value of $V_{p, q}(b_p, b_q)$ is large. Otherwise, the value is close to zero. Adding the binary energy term could reflect the penalty of error distribution in the Graph Cuts energy function effectively. The candidate target superpixels set can be obtained by the Graph Cuts by minimizing the energy function.

4 Color local entropy

The color local entropy is introduced in this paper. The color local entropy not only reflects the local color information, but also contains the location information and the entropy information. So the tracking system has a better performance in approaching the challenges of occlusion and deformation, and it is more robust in dealing with the challenges of rotation and similar background interference.

In this paper, we take a superpixel as a local area, and calculate the value of color local entropy in the superpixel. Color local entropy is a kind of the color feature expressed by the form of entropy. Meanwhile, taking the distance as weight coefficient is more effective in expressing the color local entropy. The color local entropy is expressed as follows:

$$H_c = - \sum_{u=1}^m p_u \log p_u \tag{5}$$

$$p_c = C \sum_{i=1}^n k \left(\left\| \frac{x_i - x_0}{d} \right\|^2 \right) \delta [b(x_i) - u] \tag{6}$$

Where x_i and x_0 are the pixel location and the center location in the superpixels area respectively. And d is the maximum distance from x_0 to any pixel point in the superpixel area. $\delta [b(x_i) - u]$ is a Dirac function, which is used to determine whether the pixel color value of the object area belongs to the u th bin. k is the location weight coefficient. C is the normalized factor. The formulation of k and C are as follows:

$$k(r) = \begin{cases} 1, & 0 < r \leq \alpha \\ 1 - \left(\frac{r - \alpha}{1 - \alpha} \right)^2, & \alpha < r \leq 1 \end{cases} \tag{7}$$

$$C = \left[\sum_{i=1}^n k \left(\left\| \frac{x_i - x_0}{d} \right\|^2 \right) \right]^{-1} \tag{8}$$

Compared with the traditional color feature, the color local entropy is better in describing and expressing the appearance feature. On the one hand, color local entropy retains the advantages of the color feature which is stable in the challenges of deformation and occlusion. On the other hand, it has the features of image local entropy that resisting geometric distortion and stability to illumination and similar background interference. At the same time, introducing the location information as weights coefficient can reduce the influence of the symmetry of entropy. The color local entropy feature can be used in the graph matching. Graph matching is the matching between the target graph $G(V, E)$ and the candidate graph $G'(V', E')$. $G(V, E)$ is obtained by updating the target graph nodes in every frame. $G'(V', E')$ is the set of the candidate target superpixels. It is constructed by ξ -neighborhood way. The matching between $G(V, E)$ and $G'(V', E')$ is a process of constructing and solving the affinity matrix. The node information and edge information are combined together in the affinity matrix. Therefore, the graph model is robust in the challenges of occlusion and deformation. In the affinity matrix, we fuse the color local entropy into the diagonal element $\Omega_1(c_i)$ instead of the simple color feature. The non-diagonal element $\Omega_2(c_i, c_j)$ is the edge corresponding relationship between the candidate graph and the

target graph. $\Omega_1(c_{ii'})$ and $\Omega_2(c_{ii'}, c_{jj'})$ are as follows:

$$\Omega_1(c_{ii'}) = \exp \left\{ -\frac{1}{\varepsilon_1^2} (D(f_i, f_{i'}))^2 \right\} \quad (9)$$

$$\Omega_1'(c_{ii'}) = \exp \left\{ -(H_i - H_{i'})^2 / (H_i + H_{i'}) \right\} \quad (10)$$

$$\Omega_1(c_{ii'}) = \min(\Omega_1(c_{ii'}), \Omega_1'(c_{ii'})) \quad (11)$$

Where $\varepsilon_1 = 1/\sqrt{2}$ is a constant, $\Omega_1(c_{ii'})$ is the χ^2 distance between superpixel i and i' . $\Omega_1'(c_{ii'})$ is the color local entropy distance between superpixel i and i' . The color local entropy is more holistic compared with the color feature. When the object is similar to the background, the color difference is small and the value of $\Omega_1(c_{ii'})$ is larger. It is easy to result in the error matching. While, the value of $\Omega_1'(c_{ii'})$ is smaller relatively because of the color local entropy including the location information and the entropy information. The possibility of matching the success is small. Therefore, the color local entropy can be used to reduce the error matching ratio, and the matching result is more robust.

$$\Omega_2(c_{ii'}, c_{jj'}) = \exp \left\{ -\frac{1}{\varepsilon_2^2} \left\| (l_i - l_j) - (l_{i'} - l_{j'}) \right\|_2^2 \right\} \quad (12)$$

Where, $\varepsilon_2 = r/\sqrt{2}$, $r = \sqrt{W \cdot H / N_s}$. $l_i - l_j$ is the distance where l_i and l_j are the node spatial location in the graph $G(V, E)$. Therefore, $\Omega_2(c_{ii'}, c_{jj'})$ reflects the change of edges when the object is deformed. Hence, it is insensitive to deformation relatively.

The details of the target location and updating part can be found in [10]. In the process of location, the weight coefficient consists of $\Omega_1(c_{ii'})$ and $\Omega_2(c_{ii'}, c_{jj'})$, as well as their neighborhood, thus it is more flexible. Concerning the value of the weight coefficient, there are two situations as follows: Firstly, if both the superpixels and its neighborhood are the target, the weight value is relatively large. Secondly, if a superpixel is matched successfully, while the neighborhood superpixels are not matched successfully, the weight value is relatively small. The process of updating includes updating retraining the LASVM and updating the target graph nodes.

5 Experiments

The proposed algorithm is implemented in C++ code on a PC with 2.20 GHz CPU and 2 GB memory. The selected video sequence is mainly characterized by deformation, occlusion, background interference, rotation and illumination. Meanwhile, the proposed algorithm is compared with other tracking algorithms (CT algorithm, Frag algorithm, TLD algorithm, DFT algorithm, DGT algorithm, LIAPG algorithm and SPT algorithm). The tracking performance is evaluated from the three aspect of qualitative analysis, quantitative analysis and the analysis of feature. We choose nine kinds of different video data (basketball, bolt, avatar, david3, football1, lemming, mountain-Bike, up, walking) to discuss the performance of some state-of-the-art trackers. We choose some data for the qualitative analysis. The tracking results can be seen in the Fig2.

5.1 Qualitative analysis

Deformation: Structure deformation is a great challenge for bounding box based trackers. For example in the sequences of basketball and bolt where the target changes structure quickly, the DGT algorithm and the improved algorithm have a better performance than other algorithm. But the bounding box of the improved algorithm is more accurate.

Occlusion: Sequences with target occlusion, such as basketball, bolt and lemming, brought difficulties to CT algorithm and TLD algorithm, because the lacking of the occlusion handling mechanisms. Meanwhile, the Frag algorithm and the SPT algorithm have an occlusion handling mechanisms, while they do not always have a good performance due to the interference of other challenges. The improved algorithm has a better performance in dealing with these challenges.

Illumination variations: Illumination variations have serious influence on the appearance features. The illumination variations sequence including avatar, basketball and lemming. It is obvious in the avatar sequence especially. In the avatar sequence, the other trackers had a drift. The improved algorithm has a better performance, because of fusing the color local entropy feature.

Rotation and Background interference: The most of sequence that we chose has the background interference and rotation on different degrees. It is typical in the bolt and lemming sequences. Many trackers have a poor performance in dealing with these challenges, including the DGT algorithm. While, the improved algorithm has a good performance because of the color local entropy feature. So, the improved algorithm is more robust than DGT algorithm.

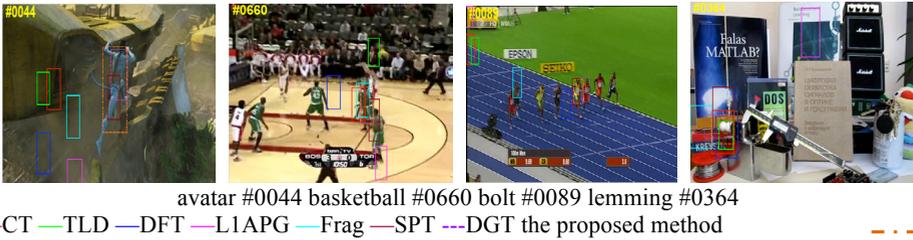


Fig. 2. Tracking results

5.2 The analysis of characteristics of performance

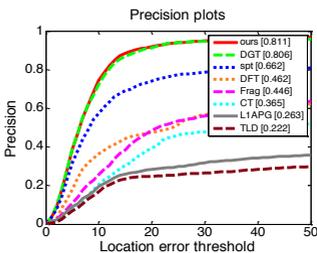


Fig. 3(a). The average ACEP.

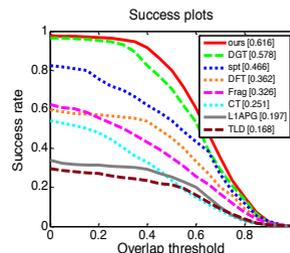


Fig. 3(b). The average SC.

The characteristics include deformation, occlusion, rotation, background interference and illumination. The Fig 3(a) shows the average ACEP value of the improved algorithm is more accurate than other trackers in these characteristics. Meanwhile, the Fig 3(b) shows the improved algorithm has a good performance than other trackers.

6 Conclusion

Due to only adopted the color feature to represent the object, the original DGT algorithm is sensitive to background interference and illumination, therefore, we take full use of the advantage of local gradient and color local entropy, and proposed an improved DGT algorithm. The proposed algorithm not only has a better performance in dealing with deformation and occlusion, but also is good in deal with background interference, rotation and illumination. The experiments results have shown that the improved algorithm is better than the DGT algorithm.

Reference

1. D. Comaniciu, et al, IEEE Trans. Pattern Anal. Mach. Intell, **5**, 564 (2003).
2. T. Liu, X P. Cheng, IEEE International Conference on Computer Engineering and Technology, 575(2010)
3. Y. Xu, Systems Process and Control, 224(2013)
4. N. Li, IEEE Conference on Electrical and Computer Engineering. Canadian, 3(2015)
5. K. Okuma, et al. Multitarget Detection and Tracking, **3021**, 28(2004)
6. H. Grabner, H. Bischof, IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 260(2006)
7. B. Babenko, M H. Yang, S. Belongie, IEEE Trans Pattern Anal Mach Intell, 983(2009)
8. Z. Cai , L. Wen, Z. Lei, et al, IEEE Trans. Image Processing, **12**, 5497(2014)
9. R. Achanta, A. Shaji, K. Smith, et al, SLIC superpixels. Epl, (2010)
10. W. Huan, W. Qinglin, W. Meng, et al, Third International Conference on Instrumentation, Measurement, Computer, Communication and Control, 114(2013)
11. A. Bordes, et al, Journal of Machine Learning Research, **6**, 1579(2012)