

Tracking Deformable Target via Multi-cues Active Contours

Peng Lv^(✉) and Qingjie Zhao

Beijing Key Lab of Intelligent Information Technology, School of Computer Science,
Beijing Institute of Technology, Beijing 100081, People's Republic of China
{p1v,zhaoqj}@bit.edu.cn

Abstract. In this study, we present a novel multi-cues active contours based method for tracking target contours using edge, region, and shape information. To locate the target position, a contour based meanshift tracker is designed which combines both color and texture information. In order to reduce the adverse impact of sophisticated background and accelerate the curve motion, we extract rough target region from the coming frame by the proposed target appearance model. What's more, both discriminative pre-learning based global layer and voting based local layer are integrated into our appearance model. For obtaining the detailed target boundaries, we embed edge, region, and shape information into the level sets based multi-cues active contour model (MCAC). Experiments on seven video sequences demonstrate that the proposed method performs better than other competitive contour tracking methods under various tracking environment.

Keywords: Object contour tracking · Active contours · Level sets · Segmentation

1 Introduction

Tracking boundaries of moving objects is an important and challenging task in computer vision. Unlike conventional tracking methods limited to use the rectangular bounding-box or other stationary shapes [1] to represent the target, contour tracking aims to extract the detailed target boundary information. In order to meet this goal, a variety of non-rigid target contour tracking methods have been proposed during last two decades.

In [2], Paragios et al. firstly use the geodesic active contour model [3] to drive the curve to target boundaries during the evolution. Zhang et al. [4] introduce a background mismatching based method for tracking the region of moving target. Niethammer et al. [5] also propose a region based method for target contour tracking. However, due to that only simple edge or region information is used,

P. Lv—This work is supported by the National Natural Science Foundation of China (No. 61175096 and No. 61273273) and Specialized Fund for Joint Building Program of Beijing Municipal Education Commission.

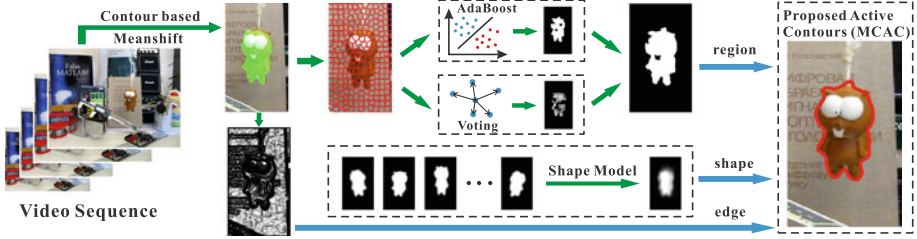


Fig. 1. Framework of the proposed contour tracking method.

these methods could not cope with the complex tracking environment, such as pose variation, occlusion, and sophisticated background. Recently, Bibbly et al. [6] propose a pixel-based contour tracking method which use generative model, however, without the edge information, the tracker may lose precise target boundary information. What's more, in [7], Vaswani et al. use a particle filter based mode tracker to track the target boundaries during the tracking. Combing both region and edge information, Cai et al. [8] propose a contour tracking framework. In [9], Fan et al. introduce an image matting based method to track the target region on a scribble trimap. However, these methods could not capture the various appearance in time, which may result in false segmentation.

In this paper, we propose a novel level sets based framework for tracking non-rigid target boundaries using edge, region and shape information. Our work mainly has the following three-fold contributions: (1) We propose a contour based meanshift target locating algorithm which integrates joint color and texture cues. (2) We propose a novel superpixel based dynamic appearance model using both global and local layers to extract the discriminative rough target region. In our method, a AdaBoost based pre-learning model and a voting algorithm are embedded into the global and local layers, respectively. Besides, for capturing changes of the target under sophisticated background, we update the appearance model dynamically during the tracking. (3) We also propose a multi-cues active contour model (MCAC) which combines edge, discriminative region and shape information for segmenting the target. To avoid the reinitialization procedure and reduce the computing time during the curve evolution, a distance regularization term is added in to the active contour model. With the discriminative region and shape information, the target could be segmented under various environment. The framework of our method is shown in Fig. 1.

The rest of the paper is organized as follows: Sect. 2 describe our target contour tracking framework. We show the qualitative and quantitative results in Sect. 3. In Sect. 4, we summarize the paper.

2 Proposed Method

2.1 Contour Based Meanshift Target Locating

To reduce the impact of complex background and extract the target more effectively, firstly we locate the target region before extracting its boundaries.

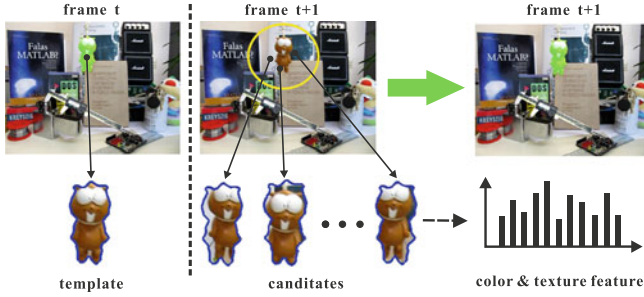


Fig. 2. Illustration of our contour based meanshift tracker.

A natural approach to track and locate the target position would be simply use meanshift, which use the rigid or elliptical region to represent the target. However, this approach has two drawbacks: (1) important target contour information may be lost during the tracking; and (2) the tracker may be interfered by the background in target boundingbox. To cope with these two problems, we use non-rigid region to represent the target in our meanshift tracker.

Considering that the target shape changes continuously during the tracking, which means that the shapes of targets between two continuous frame are highly similar. As shown in Fig. 2, at frame $t + 1$, we use the non-rigid target region $I_C(t)$ in frame $I(t)$ as the target template, which provides precise target information. To enable our tracker achieve more robust tracking under various environment, we extract both color and texture information from the target region. For color information, a histogram is extracted in RGB&HSV color space, while for texture information, LBP feature is used:

$$\begin{cases} \mathbf{f}_R = \{ \mathbf{f}_{color}, \mathbf{f}_{texture} \} \\ \mathbf{f}_{color} = f_{RGB\&HSV} \\ \mathbf{f}_{texture} = LBP(I_C(t)) \end{cases} \quad (1)$$

To measure the similarity between template region and candidates, we use the following distance:

$$d(\mathbf{f}_R, \mathbf{f}) = \sqrt{1 - \rho[\mathbf{f}_R, \mathbf{f}]} \quad (2)$$

where $\rho[\cdot]$ is the Bhattacharyya distance between two discrete distributions, which defined as:

$$\rho[\mathbf{f}_R, \mathbf{f}] = \sum_i^N \sqrt{\mathbf{f}_{i,R} \cdot \mathbf{f}_i} \quad (3)$$

Then we use meanshift algorithm to find the target position \mathbf{y}'_R in frame $t + 1$ as follows:

$$\mathbf{y}'_R = \frac{\sum_{i=1}^n \mathbf{y}_{i,R} w_i g(\cdot)}{\sum_{i=1}^n w_i g(\cdot)} \quad (4)$$

where w_i is the candidates weights, and $g(\cdot)$ is the kernel, respectively. After several iterations, a new non-rigid target position could be obtained in frame

$I(t+1)$, as shown in Fig. 2. In our method, this non-rigid target region provides important information for our appearance model, which will be described more detailedly in Sect. 2.2.

2.2 Appearance Model Combing Global and Local Layers

Considering that sophisticated background may affect the curve motion during the segmentation procedure, so we propose an appearance model to extract rough target region from the coming frame $I(t+1)$. Some prior works tend to use pixel-based or sparse-based models to represent the target, however, those models usually lost the detailed target boundaries information. In our method, we build a target appearance model based on superpixels, which enables the model to retain both target region and boundary information simultaneously during the tracking. Besides, rather than the single-layer target model in traditional methods, we combine both global and local layers for obtaining the rough region more robustly.

Discriminative Pre-learning Based Global Layer: In global layer, we use discriminative method to extract the global rough target region. For every superpixel sp , a histogram based feature descriptor \mathbf{s} is extracted in RGB and HSV color space. Those feature descriptors are labeled by $+1/-1$ according to the following criteria:

$$\mathbf{s} = \begin{cases} \mathbf{s}^+, & \text{if } (\mathbf{s} \cap \text{Target})/\mathbf{s} \geq \eta \\ \mathbf{s}^-, & \text{if } (\mathbf{s} \cap \text{Target})/\mathbf{s} < \eta \end{cases} \quad (5)$$

To extract the rough target region during the tracking, an online AdaBoost classifier is trained and updated dynamically based on the labeled samples.

However, the target appearance may change during the tracking, which may lead to false classification. To avoid this problem, we pre-learning the target appearance from the coming frame $I(t+1)$ before classifying superpixels. Recall that in Sect. 2.1, the meanshift tracker locates the non-rigid target region in frame $I(t+1)$, which enables us to use this information to update the AdaBoost classifier. In pre-learning procedure, we select some unlabeled superpixels randomly from the region $I_C(t+1)$ in next frame as the positive examples to update the classifier. What's more, the internal superpixels have higher probability to be selected than ones closed to the periphery. After pre-learning the target appearance, our model could capture changes of the target. Therefore, in the global layer, a rough target region R_{t+1}^{global} finally obtained, as shown in Fig. 3(e).

Voting Based Local Layer: However, under various tracking conditions, the global layer may miss some local regions, which would result in false segmentation. To reduce the adverse impact of noises caused by global layer, we propose a local voting algorithm to extract the target region. In our model, in order to retain the local features, every unlabeled superpixel in coming frame $I(t+1)$ is voted by the surrounded labeled superpixels in prior frame.

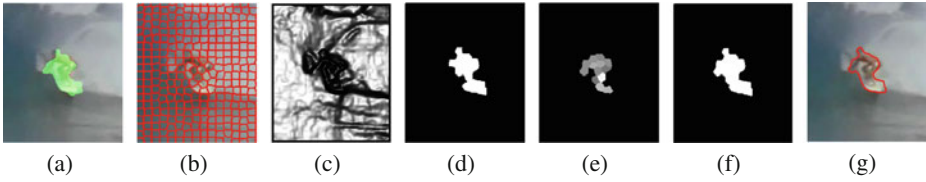


Fig. 3. Illustration of our global and local based appearance model: (a) the located position by our meanshift tracker; (b) superpixel segmentation; (c) the target edge information; (d) discriminative region of the global layer; (e) result of voting in local layer; (f) the final rough target region; (g) final segmentation result on target region.

In local layer, we use the following distance to measure the similarity between two superpixels:

$$d_{sp}(\mathbf{s}_i, \mathbf{s}_j) = \exp\left(-\frac{\rho^2[\mathbf{s}_i, \mathbf{s}_j]}{\sigma}\right) \tag{6}$$

where $\rho(\cdot)$ is the Bhattacharyya distance given in Eq. 3. For every superpixel in $I(t+1)$, the score which voted by the surrounded superpixels in $I(t)$ is computed by the following formula:

$$\text{score}(sp_{i,t+1}) = \frac{\sum_j^{sp_{j,t} \in \Omega_r} \chi(d_{sp}(\mathbf{s}_i, \mathbf{s}_j))}{\|\chi(d_{sp}(\mathbf{s}_i, \mathbf{s}_j))\|_0} \tag{7}$$

where Ω_r is the region of radius r surrounding the superpixel $sp_{i,t+1}$ in frame $I(t)$. Besides, the kernel function $\chi(\cdot)$ is given by:

$$\chi(d_{sp}(\mathbf{s}_i, \mathbf{s}_j)) = \begin{cases} d_{sp}(\mathbf{s}_i, \mathbf{s}_j), & \text{if } d_{sp}(\mathbf{s}_i, \mathbf{s}_j) \geq \zeta \\ 0, & \text{if } d_{sp}(\mathbf{s}_i, \mathbf{s}_j) < \zeta \end{cases} \tag{8}$$

After the voting procedure, the local target region R_{t+1}^{local} is obtained, as shown in Fig. 3 (f).

For obtaining more stable target region, we combine the global and local layers as follows: $R_{t+1} = R_{t+1}^{global} \cup R_{t+1}^{local}$. This rough target region provides important region information for our active contour model, which will be discussed more detailedly in Sect. 2.4. Moreover, in order to reduce the noise cause by those two layers, open operator is used to the expanded rough target region:

$$R'_t = (R_t \ominus B_1) \oplus B_2 \tag{9}$$

where B_1 and B_2 denote the erosion and dilation structuring element, respectively. As shown in Fig. 3 (g) and (h), after integrating both global and local information into the appearance model, target could be extracted accurately.

2.3 Dynamic Shape Model

During the segmentation, various noise such as illumination and target appearance changes may affect the curve evolution, which would results in the false

segmentation. What's more, some false negative regions generated by our appearance model may also cause over-segmentation. In order to handle these problems, we build a dynamic shape model to guide curve motion during the evolution.

For representing target shape S_t at time t , a gaussian kernel is applied to target region: $S_t = G(C_t)$, where C_t is the target region mask which is labeled by 1s and 0s. During the tracking, our shape model is updated dynamically as follows:

$$\begin{aligned} S_t &= \sum_{k=1}^t p^{t-k} G(C_k) = G(C_t) + \sum_{k=1}^{t-1} p^{t-k} G(C_k) \\ &= G(C_t) + S_{t-1}. \end{aligned} \quad (10)$$

2.4 Multi-cues Active Contours and Curve Evolution

In this section, we will introduce our multi-cues active contour model for segmenting target which combines edge, region, and shape information. Because conventional active contour models [3, 10, 11] only consider edge or region information, therefore, the curve is vulnerable to be interfered by the complicated background or obvious boundaries and may stop at the false position after evolution. To handle these limitations, in our method, we embed our dynamic appearance model and shape model into active contours.

Edge Information: As many works refer, an edge-detector is defined for extracting the image boundaries: $g(|\nabla I|) = 1/(1 + |\nabla \hat{I}|^2)$. Note that the rough expanded target region R'_t , which is obtained in our appearance model as described in Sect. 2.2, could reduce the negative effect of the background. Therefore, to accelerate the curve evolution, we just let the curve move on the extended rough target region $I'_R(t)$, where $I'_R(t) = R'_t \cdot I(t)$. Then the edge information of the rough target region could be represented as follows:

$$g_{edge} = \frac{1}{1 + |\nabla R'_t \cdot I(t)|^2} = R'_t \cdot g(|\nabla I(t)|) - R'_t + 1. \quad (11)$$

According to the edge information g_{edge} , we define an edge term in our active contour model:

$$\mathcal{F}_1 \triangleq \int_{\Omega} g_{edge} \cdot \delta(\varphi) |\nabla \varphi| dx. \quad (12)$$

Region Information: In many situations, it is hard to extract target boundaries due to the blurred edge or sophisticated background, which would affect the curve motion during the evolution. In order to enable the curve to stop at the target boundaries correctly, target region information is embedded into our active contour model.

Recall that in Sect. 2.2, the rough target region R_t provides important information of target region for the active contour model. To embed the region information into our model, we transform the region R_t into homologous edge information beforehand:

$$\begin{aligned}
 g_{region} &= g(|\nabla R_t \cdot I(t)|) + g(|\nabla R_t|) - 1 \\
 &= R_t \cdot g(|\nabla I(t)|) + g(|\nabla R_t|) - R_t.
 \end{aligned}
 \tag{13}$$

Therefore, we define the following region term in our active contour model:

$$\mathcal{F}_2 \triangleq \int_{\Omega} g_{region} \cdot \delta(\varphi) |\nabla \varphi| dx.
 \tag{14}$$

Shape Information: During the tracking, our appearance model may generate some false negative regions. Caused by the false negative regions information, the curve may move across the target boundaries, and stop at the wrong position. To cope with this problem, we add the target shape information to the active contour model:

$$\begin{cases} g'_{edge} = S_t \cdot g_{edge} \\ g'_{region} = S_t \cdot g_{region} \end{cases}
 \tag{15}$$

where S_t is the target shape model. Then we use Eq. 15 to update Eqs. 12 and 14, respectively. After integrating with shape information, our active contour model could produce more stable results.

Energy Functional and Curve Evolution: Combining the edge, region, and shape information, we propose a multi-cues active contour model (MCAC):

$$\mathcal{E}(\varphi) = \alpha \mathcal{F}_1(\varphi) + \beta \mathcal{F}_2(\varphi) + \mu \mathcal{R}(\varphi) + \tau \mathcal{A}(\varphi)
 \tag{16}$$

where $\mathcal{A}(\varphi)$ and $\mathcal{R}(\varphi)$ is area accelerate term and non-reinitialization term to speed up the curve evolution procedure, respectively. These two terms are given by:

$$\mathcal{A}(\varphi) \triangleq \int_{\Omega} g(|\nabla I(t)|) H(-\varphi) dx
 \tag{17}$$

$$\mathcal{R}(\varphi) \triangleq \int_{\Omega} p(|\nabla \varphi|) dx
 \tag{18}$$

where $H(\cdot)$ is the Heaviside function and $p(\cdot)$ is a potential function define in [11]. By using the finite difference calculation framework, the following gradient flow is obtained to optimize the energy functional $\mathcal{E}(\varphi)$:

$$\begin{aligned}
 \frac{\partial \varphi}{\partial t} &= \delta_{\epsilon}(\varphi) [\alpha \operatorname{div}(S_t \cdot g_{edge} \cdot \mathbf{F}) + \beta \operatorname{div}(S_t \cdot g_{region} \cdot \mathbf{F})] \\
 &\quad + \mu \operatorname{div}(d_p(|\nabla \varphi|) \nabla \varphi) + \tau g(|\nabla I|) \delta_{\epsilon}(\varphi),
 \end{aligned}
 \tag{19}$$

where $\mathbf{F} = \nabla \varphi / |\nabla \varphi|$.

3 Experimental Results

3.1 Experimental Setup

The proposed method is implemented in MATLAB R2010b under Red Hat Enterprise Linux platform on a Intel(R) Core(TM)i7 3.4 GHz processor with 3 GB memory.

Parameters Setting: In Sect. 2.2, the radius r of voting region Ω_r is set to 20, and $\zeta = 0.3$. In Eq. 9, the erosion and dilation structuring element are 5×5 and 12×12 , respectively. The updating parameter p in our dynamic shape model is set to 0.6. Besides, we set $\alpha = 1$, $\beta = 3$, $\mu = 1$, and $\tau = 2$ in our active contour model. During the evolution, we set number of the inner and outer iteration steps as 8 and 40, respectively.

Compared Algorithms and Evaluation Criteria: In our experiment, eight target contour tracking algorithms are compared: (a) our method with distance regularized level set evolution (DRLSE) [11]; (b) our method with region-based active contours (G-CV) [10]; (c) our method with edge-based active contours (GAC) [3]; (d) our method without shape information (w/o shape); (e) Scribble tracker which based on matting approach (Scribble tracker) [9]; (f) particle filter based mode tracker (deform PF-MT) [7]; (g) region tracking method based on background mismatch (Mismatch) [4]; and (h) our proposed method (MCAC). Moreover, To evaluate the segmentation performance, mis-tracked pixels rate (MPR) is defined: $MPR_t = |R_t^g \cup R_t - R_t^g \cap R_t| / |R_t^g|$, which indicates the coverage ratio between result and ground truth.

3.2 Qualitative and Quantitative Analysis

Complex Background: We test the methods on video *Lemming*, where the background is sophisticate during target moving, to verify the effectiveness of our method. As shown in Fig. 4, due to the sophisticated background and lacking of target shape information in DRLSE based method, the curve stops at the misplaced boundaries. Because of the accumulated errors during the tracking, Scribble tracker fails to segment the target correctly, yet. In our method, the rough target region extracted by the global and local based appearance model makes the segmentation environment more clear and provides important region information for the active contour model. By combing the region and shape information, our active contour model could cope with the interferences caused by complex background.

Various Appearance: For demonstrating the improvements of our method under various appearance tracking environment, we running the compared methods on video *Seq.sb*. During the tracking, as shown in Fig. 5, due to the changes of target pose and appearance, both deform PF-MT and Mismatch tracker lose the appearance information and fail to extract the target. On the contrary, profiting from the pre-learning procedure, our dynamic appearance model could capture

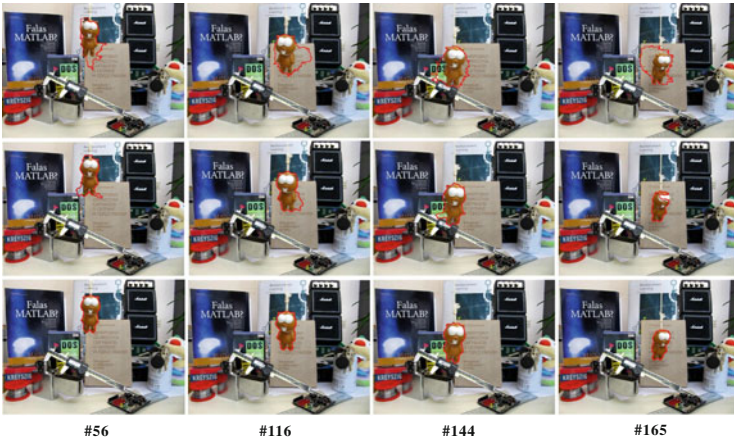


Fig. 4. Tracking results on *Lemming* with three methods (from top to bottom): Scribble tracker [9], ours with DRLSE [11], and the proposed method.

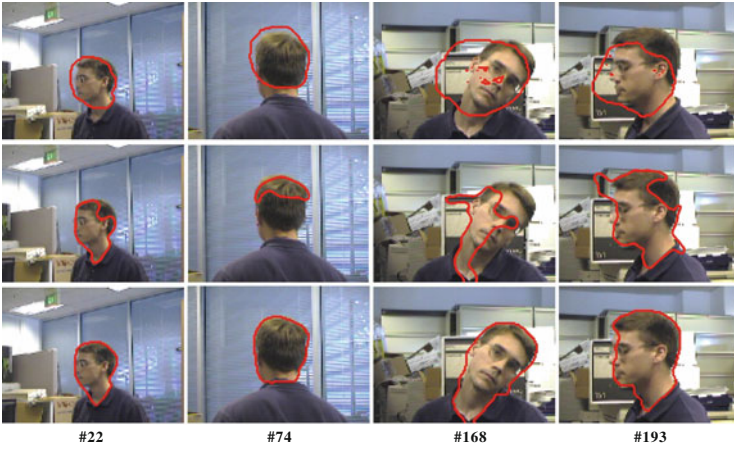


Fig. 5. Tracking results on *Seq.sb* with three methods (from top to bottom): Mismatch tracker [4], Deform PF-MT [7], and the proposed method.



Fig. 6. Tracking results on *Panda* with three methods (from top to bottom): Scribble tracker [9], ours w/o shape method, and the proposed method.

Table 1. The mis-tracked pixels rate (MPR) on seven video clips with eight compared methods (the second best results are labeled with red font).

Sequence	w/o shape	G-CV [10]	GAC [3]	DRLSE [11]	Mismatch [4]	Scribble [9]	PF-MT [7]	MCAC
<i>Lemming</i>	0.156	1.425	0.873	0.164	0.443	0.330	0.254	0.096
<i>Sufer</i>	0.432	1.995	0.829	0.497	0.664	0.284	0.322	0.262
<i>Pedxing1</i>	0.113	0.333	0.474	0.216	0.460	0.199	0.328	0.121
<i>Seq_sb</i>	0.269	2.111	0.911	0.504	0.661	0.656	0.459	0.231
<i>Seq_dhb</i>	0.439	1.727	0.603	0.326	0.881	0.490	0.405	0.253
<i>Seq_dt</i>	0.183	1.343	0.696	0.436	0.494	0.656	0.491	0.153
<i>Panda</i>	0.226	2.229	1.698	0.463	1.952	1.327	0.874	0.181

the appearance changes promptly, which enables the proposed active contour model to segment the target correctly.

Occlusion: As shown in Fig. 6, in video *Panda*, the target is occluded by a tree and also rotates during its moving. At frame 54, Scribble tracker cannot extract the target region due to the occlusion. Notice that here we also test our method without shape information, as shown in Fig. 6, where we can see that without the shape information, the curve crosses the target boundaries and stops at internal region of the target. Thanks to the shape information in our active contour model, our method could deal with the occlusion during the tracking.

Now we quantify our method. As shown in Table 1, the conventional active contour model based methods (G-CV [10], GAC [3], and DRLSE [11]) usually lose the target during the tracking. That is mainly because the noises from the various tracking environment interfere the curve motion during the evolution, which results in false segmentation. Due to lacking of the dynamic appearance information, Mismatch tracker [4] could not capture the various appearance, and fails to segment the target. Both Scribble [9] and deform PF-MT [7] tracker do better on several tested sequences than conventional tracker, however, these two methods could not cope with the occlusion, as tested on *Panda*. Integrating with edge, region, and shape information, the proposed method performs better than other state-of-the-art methods under various tracking environment.

4 Conclusion

In this paper, we propose a novel level set based target contour tracking method based on multi-cues active contours by combing edge, region, and dynamic shape information for segmenting the target. Qualitative and quantitative results show that our method performs better than other state-of-the-art methods. Further work will aim at developing a more powerful appearance model to represent the target, which may improve the segmentation performance.

References

1. Yilmaz, A.: Kernel-based object tracking using asymmetric kernels with adaptive scale and orientation selection. *Mach. Vision Appl.* **22**(2), 255–268 (2011)
2. Paragios, N., Deriche, R.: Geodesic active contours and level sets for the detection and tracking of moving objects. *IEEE Trans. Pattern Anal. Mach. Intell.* **22**(3), 266–280 (2000)
3. Caselles, V., Kimmel, R., Sapiro, G.: Geodesic active contours. In: *Proceedings of IEEE International Conference on Computer Vision*, pp. 694–699 (1995)
4. Zhang, T., Freedman, D.: Improving performance of distribution tracking through background. *IEEE Trans. Pattern Anal. Mach. Intell.* **27**(2), 282–287 (2005)
5. Niethammer, M., Tannenbaum, A., Angenent, S.: Dynamic active contours for visual tracking. *IEEE Trans. Autom. Control* **51**(4), 562–579 (2006)
6. Bibby, C., Reid, I.: Real-time tracking of multiple occluding objects using level sets. In: *Proceedings of IEEE Conference on Computer Vision Pattern Recognition*, pp. 1307–1314 (2010)
7. Vaswani, N., Rathi, Y., Yezzi, A., Tannenbaum, A.: Deform pf-mt: particle filter with mode tracker for tracking nonaffine contour deformations. *IEEE Trans. Image Process.* **19**(4), 841–857 (2010)
8. Cai, L., He, L., Yamashita, T., Yiren, X., Zhao, Y., Yang, X.: Robust contour tracking by combining region and boundary information. *IEEE Trans. Circuits Syst. Video Technol.* **21**(12), 1784–1794 (2011)
9. Fan, J., Shen, X., Ying, W.: Scribble tracker: a matting-based approach for robust tracking. *IEEE Trans. Pattern Anal. Mach. Intell.* **34**(8), 1633–1644 (2012)
10. Chen, L., Zhou, Y., Wang, Y., Yang, J.: GACV: geodesic-aided c-v method. *Pattern Recogn.* **39**(7), 1391–1395 (2006)
11. Li, C., Xu, C., Gui, C., Fox, M.D.: Distance regularized level set evolution and its application to image segmentation. *IEEE Trans. Image Process.* **19**(12), 3243–3254 (2010)