Tracking Non-Rigid Target via Dynamic Discriminative Geodesic Active Contours

Peng Lv, Qingjie Zhao

Beijing Lab of Intelligent Information Technology, School of Computer Science, Beijing Institute of Technology, Beijing 100081, P.R. China

Email: p1v@bit.edu.cn, zhaoqj@bit.edu.cn

Dongbing Gu

School of Computer Science and Electronic Engineering, University of Essex, Colchester, UK

Email: dgu@essex.ac.uk

Abstract

In this paper, we propose a novel dynamic framework for tracking non-rigid target in video sequence. To solve the problems caused by various situation and sophisticated background during tracking of a target, a discriminative appearance model based on dynamic SVMs is first proposed to extract the rough target region, which provides discriminative target region information clues. In addition, we design a new discriminative geodesic active contour model (DGACM) that combines both edge and discriminative region information, thereby enabling the extraction of more precise object boundaries compared to conventional active contour models. By adjusting the proportion between edge and region constraint terms, the proposed model can adapt to appearance variation, sophisticated background and occlusion. Experiments on synthesized and real image sequences demonstrate that the proposed method achieves more robust and effective deformable target tracking in challenging situations compared to other competitive contour tracking methods.

Index Terms

Active contours, Contour tracking, Segmentation, Discriminative model, Level set method

I. INTRODUCTION

In computer vision, the problem of detecting and tracking moving objects has a wide range of applications, including navigation [15], robotics [40], video surveillance [21] and human computer interaction [32].

Over the last two decades, a large variety of detection and tracking methods have been proposed. For a survey of these early algorithms, we refer the reader to [42]. Generally, object tracking algorithms can be categorized as generative and discriminative. In generative tracking methods, a generative model is built to estimate a potential target location according to previous target observations. The target observations, such as shapes, colors, and motions, are often modeled as Gaussian distributions [34], Markov Random Fields [38], or other mixture models [37]. Because the generative models should be carefully built under different conditions and because the parameters are difficult to estimate, discriminative models have recently begun to receive increasingly more attention. Discriminative tracking methods [14] attempt to maximize the inter-class separability between foreground and background regions using discriminative learning techniques such as Support Vector Machines [2], [24], [33], boosting [17], random forest [30], metric learning [36], and multiple instance learning [1]. To adapt to the variations in the tracking environment (e.g., shape deformation, illumination changes, or occlusion), discriminative models usually learn the target information online and dynamically update the tracker.

However, most such tracking approaches are limited to rectangular bounding boxes or other stationary shapes [41], whereas objects, such as hands, legs, and desks, in practice, may have complex

shapes that cannot be described accurately by rigid shapes. To avoid this problem, authors use contours or silhouettes to dynamically represent the targets during the tracking process. Contour-based methods, which use curves to model the boundaries of targets, can provide more information about a target, especially shape and pose information. Early studies on contour-based tracking often use Kalman filters to track a set of marked points. Later, authors tend to use active contour models to optimize the region- or boundary-based functional of the targets. Originally, active contour model is introduced by Kass, Witkin, and Terzoponlos [5] as a boundary method for segmenting objects. In Snakes [5] and geodesic active contours [8], the curve is driven to move toward the target boundaries by minimizing a boundary integral functional of features depending on edges. Introduced by [9], [10], active contours driven by region functionals achieve a substantially better performance compared to edge-based models when the edge clues of the target are much too indistinct. Moreover, methods that integrated with both region and boundary clues [35] or other target information, such as shape and motions, appear to achieve exceptional performance.

Most recently, studies on active contour tracking [4], [27] utilize the implicit level set evolution to automatically optimize the target boundaries. The level set method is independently introduced early on for image segmentation that can represent complex topological contours and is able to address topological changes such as splitting and merging. However, the conventional level set method needs to be reinitialized after several iterations to ensure the accuracy of the segmentation result, which requires substantial extra computation time. To overcome this difficulty, Li et al. [23] presents a level set formulation with a distance regularization term that requires no reinitialization step.

In this paper, we propose a novel discriminative framework for non-rigid object tracking that combines the active contour model and dynamic appearance discriminative learning model. Our work on non-rigid object tracking mainly provides the following three contributions:

1) We propose a dynamic, discriminative, non-rigid object tracking framework that combines an online SVM appearance model and discriminative active contours. The target is located using a color- and texture-based particle filter tracker. Then, a discriminative model is trained to separate the target region from the background, thereby obtaining a good performance in cluttered and various other situations. By applying the variational level set formulation to our geodesic active contour model, the accurate target contour is finally obtained. To solve the drifting problem and obtain a more accurate location, we use the information of the target segmentation result to correct the particle filter tracker.
2) We propose a discriminative appearance model for separating the target region. In our model, a series of histogram-based feature descriptors are extracted from the blocks in divided images to train the SVM classifier. In addition, our discriminative model is updated online to adapt to various tracking conditions, including illumination change and occlusion. Thus, the proposed appearance model can (a) capture changes of the target as it moves, and (b) effectively extract the rough target region from the complex background.
3) We also propose a novel non-reinitialization discriminative geodesic active contour model (DGACM). In our model, both region information and edge information are integrated into the energy functional to obtain a more accurate contour of the target, thereby achieving the multi-mode target segmentation in different challenging situations. Motivated by previous work, we add the non-reinitialization term to our formulation, thereby reducing the number of iterations and computation time while ensuring sufficient numerical accuracy during the segmentation.

The overview of our contour tracking scheme is shown in Fig. 1. First, the target is manually initialized and located by the particle filter tracker. Then, the rough target region can be obtained using our discriminative appearance model. After obtaining the edge information and discriminative region information of a target, the proposed discriminative geodesic active contour model (DGACM) drives the curve toward target boundaries. Finally, we use the new tracking result to update the appearance model and correct the particle filter tracker to enable more robust tracking. Some related studies

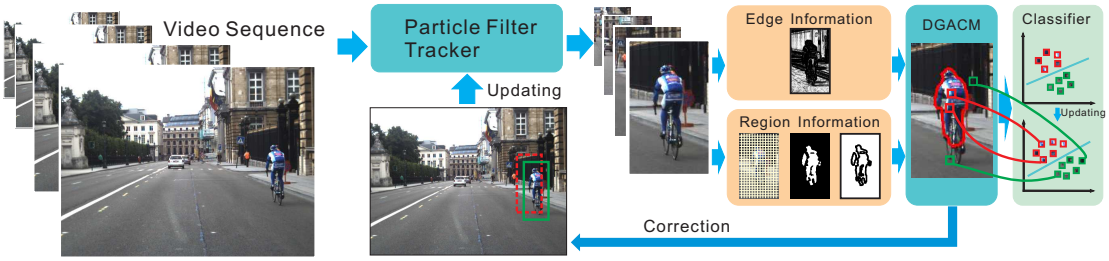

Fig. 1. The overview of the proposed our contour tracking scheme..

are listed in the following section. We describe our proposed dynamic discriminative geodesic active contour tracking approach in Sect. III. Our experimental result and its analysis are presented in Sect. V. Sect. VI summarizes the paper.

II. RELATE WORK

In this section, we will review a number of related studies that focus on detecting and tracking non-rigid target, and we also note their advantages and disadvantages.

Early papers tended to use active contour models or snakes [5] to segment a target based on the edge information of the image. The geodesic active contour model, which improves the object segmentation performance over that of the traditional parametric active contour model, is independently introduced by Caselles et al. [7] and Malladi et al. [25]. In [28], the authors apply geodesic active contours [8], [20] to an object boundary detection framework for tracking. However, the edge-based active contour method may not be able to drive the curve to the target boundaries when the target is subject to a cluttered background. Compared with edge-based contour tracking models, region-based active contour models can incorporate the color and texture information of the target into the curve evolution procedure [11], [12], [16], [19]. Chan et al. [9] develop a region-based active contour model that improves the segmentation performance on objects whose boundaries are not clearly defined by gradient. In [26], the authors track the target region by finding the correspondence of the pixels between successive frames and the upcoming frame. Yilmaz et al. [43] present a region-based contour tracking method that adopts the features of both foreground and background to achieve robust target region detection. They also implement contour-based tracking for addressing occlusion by minimizing the region-based energy functional in [44]. A probabilistic real-time contour tracking framework based on the color histogram of pixels that performs well when the background is approximately constant is introduced by Bibby et al. in [3]. Zhang et al. [45] develop a distribution contour tracking model by maximizing the density mismatching between the target region and the density sampled from background. Nevertheless, current region-based methods also have some limitations, e.g., the contour is sensitive to the similarities of colors or textures between foreground and background, which may lead to mis-segmentation. Shape-based contour tracking methods [18] can segment a target under shape information constraints. However, when the target shape changes drastically, the tracker may produce mis-segmentation. Later, various authors tend to combine both region and edge information in their generative contour tracking framework; however, when the target undergoes variations caused by camera noise, illumination changes or self-shadowing, the performance of the tracker may decline due to the lack of dynamic information about the target. In [13], an image matting based tracking scheme is proposed to extract the target boundaries on a scribble trimap, which is generated by the trackable points surrounding the boundaries.

As we will explain in the following section, compared with prior studies, the key differences in our work are as follows: (a) we integrate both edge information and discriminative region information into our dynamic tracking framework, which enables our tracker to be robust against cluttered background,

occlusion, pose and illumination variation; (b) our tracker is corrected frame by frame based on target segmentation result, which avoids the drifting problem during tracking; and (c) we add a non-reinitialization term to our active contour model to achieve more effective object segmentation.

III. DISCRIMINATIVE CONTOUR TRACKING SCHEME

In this section, we will describe our dynamic discriminative contour tracking scheme based on SVMs and the geodesic active contour model. First, we introduce the initialization procedure and the SVM-based target appearance model in Sects. III-A and III-B, respectively. Our discriminative geodesic active contour model is described in Sect. III-C. Sect. III-D explains how our tracker locates the target and how our tracker is corrected using the segmentation result frame by frame during tracking.

A. Target Initialization

Compared to conventional visual tracking methods, which usually use a rectangular bounding box or an elliptic shape that contains the entire target to represent the initial state of the target, contour tracking methods can describe the initial state with more complicated non-rigid shape. To initialize the first frame of the image sequence, in our scheme, we use GrabCut [29] to interactively and efficiently extract the target from a complex environment based on graph cuts and a Gaussian Mixture Model (GMM). After segmenting the target from the first frame, we obtain the initial contour $\mathcal{C}_0$ of the target, where the region inside the contour $\mathcal{C}_0$ is considered as the foreground and the remainder is the background. The initial state of the target can be represented as a binary matrix S_0, where the foreground is denoted by 1s and the background is denoted by 0s.

B. Feature Selection and Classification

Because the tracking problem can be considered as a binary classification problem [4] in the field of object tracking, discriminative techniques are widely used in many tracking methods. In machine learning, discriminative techniques focus on the problem of classifying positive and negative samples by modeling the conditional probability distribution. Such techniques first learn the parameter of the model using positive and negative samples extracted from the known training set and then train a discriminative model, which can classify further samples without initializing the model parameters of the specific problem.

Contour tracking is also an example of binary classification problems and can be considered as segmenting the foreground (object) from the background. Compared to conventional discriminative technique based tracking, in contour tracking, more shape information of the target must be considered due to its deformable shape. To track the target contours, we must determine if regions or pixels belong to the deformable target, especially those close to the target boundaries. Early contour tracking methods [16], [28] usually employ the active contour model in segmenting targets in video sequences. However, when the target background is complex or its appearance changes during tracking, it is difficult to accurately separate the foreground from the background in conventional tracking methods.

Considering the above discussion, we integrate the discriminative model into our contour tracking framework, which reduces the effects of sophisticated background, blurred targets or other conditions. Motivated by the work in [39], we split the region containing the target into a group of $l \times l$ blocks every two pixels, where l is the length of a side of the block. For each block $\mathbf{b}$, a histogram-based m-dimensional feature descriptor, $\mathbf{s} = \{x_1, x_2, \cdots, x_m\}$, will be extracted in the RGB or HSV color space. The feature descriptors $\mathbf{s}$ sampled from the target region are labeled by positive $\mathbf{s}^+$ or negative $\mathbf{s}^-$ according to the following criteria:

$$\mathbf{s} = \begin{cases} \mathbf{s}^+, & \text{if } \dfrac{|\mathbf{b} \cap Target|}{|\mathbf{b}|} \geq \xi, \\ \mathbf{s}^-, & \text{if } \dfrac{|\mathbf{b} \cap Target|}{|\mathbf{b}|} < \xi, \end{cases} \tag{1}$$

(a) (b) (c) (d)

Fig. 2. The influence on the segmentation performance of the opening operation: **a** the classified image I_{SVM} (note that the false negative region is very large); **b** the segmentation result by our discriminative geodesic active contour model; **c** the classified image I_d after applying the opening operation; **d** the accurate segmentation result under the constraint of I_d in our model. We can see that after applying Eq. (2), the false negative regions shrink a lot and a better segmentation result can be obtained.

where $\mathbf{b} \cap Target$ denotes the intersection region between block **b** and the target. In our model, we set $\xi = 0.4$. To classify the features extracted from the next frame in the video sequence, we train the SVM classifier based on the set of known samples. After classifying the features in the upcoming frame, an approximate region I_{SVM} of the target will be obtained, as shown in Fig. 2(a). This rough region ignores most of sophisticated background and provides a better environment for the curve evolution process to be applied in our proposed geodesic active contour model.

However, when our discriminative model classifies the features, it may generate some false positives and false negatives in I_{SVM} (as shown in Fig. 2(a)), which interfere with the curve motion during its evolution. To avoid the deviation, we use the *opening* operation to reduce the noise:

$$I_d = G_{dilation}(G_{erosion}(I_{SVM})). \tag{2}$$

The erosion operation $G_{erosion}$ can eliminate large number of false positive regions or pixels. In addition, the dilation operation $G_{dilation}$ can collect the false negative regions, which improves the accuracy of the segmentation in our geodesic active contour model.

As shown in Fig. 2, without the opening operation, the target cannot be correctly segmented due to the false negative region. After applying Eq. (2) to the classified image, the false negative region decreases in size to a small block, and finally, we obtain good segmentation performance by applying our geodesic active contour model. The erosion and dilation operations are shown to be needed in our contour tracking system and remedy the deviations caused by the two types of errors in I_{SVM}. To better demonstrate the system's influence on our model, detailed experiment result on tested video sequences are presented in Sect. V-B.

To allow our model to adapt to various conditions during tracking, the SVM classifier is dynamically updated. Assuming that we have obtained the tracking result at frame t, the positive and negative samples denoted by $\mathbf{S}_t^+$ and $\mathbf{S}_t^-$ can be extracted. To adapt to various tracking environments, a forgetting factor ρ is introduced to update the training set $\mathbf{S}$ every k frames. We use the following formula to update the training set:

$$\begin{cases} \mathbf{S}^+ = \rho \cdot \mathbf{S}^+ + \dfrac{Numb((1-\rho)\cdot \mathbf{S}^+)}{Numb(\mathbf{S}_t^+)} \cdot \mathbf{S}_t^+, \\ \mathbf{S}^- = \rho \cdot \mathbf{S}^- + \dfrac{Numb((1-\rho)\cdot \mathbf{S}^-)}{Numb(\mathbf{S}_t^-)} \cdot \mathbf{S}_t^-, \end{cases} \tag{3}$$

where $\rho \cdot \mathbf{S}$ means selecting samples from set $\mathbf{S}$ with the probability ρ and $Numb(\mathbf{S})$ denotes the number of elements in a sample set $\mathbf{S}$.

C. Proposed Geodesic Active Contour Model

This section introduces our proposed geodesic active contour model, which integrates both region information and boundary information into the level set evolution procedure. First, we will introduce the framework of level-set-based active contours.

The active contour model attempts to automatically detect contours of interesting regions or targets in an image. In traditional models, authors have used an energy functional for detecting target boundaries. By minimizing the energy functional through some optimization methods, the curve can move toward the target boundaries. Let $\mathbb{C}$ be a family of curves in $\mathbf{R}^2$ satisfying the following definition:

$$\mathbb{C} = \left\{ \mathcal{C} : [a,b] \to \Omega,\ \mathcal{C} \text{ piecewise } \mathbb{C}^1,\ \mathcal{C}(a) = \mathcal{C}(b) \right\};$$

then, an energy functional is defined for $\mathcal{C} \in \mathbb{C}$:

$$\mathcal{E}(\mathcal{C}) \quad = \quad \alpha \int_a^b |\mathcal{C}'(q)|^2 dq + \beta \int_a^b |\mathcal{C}''(q)|^2 dq - \lambda \int_a^b g^2(|\nabla I(\mathcal{C}(q)|) dq. \tag{4}$$

The first two terms, called the internal energy, ensure a smooth curve while optimizing the energy functional. The third term, called the external energy, is responsible for attracting the curve moving toward the object boundaries. In the energy functional $\mathcal{E}(\mathcal{C})$, g is called an edge-detector function, which is defined as

$$g(|\nabla I(x)|) = \frac{1}{1 + |\nabla \hat{I}(x)|^p} \quad (p = 1 \text{ or } 2). \tag{5}$$

To optimize the energy functional in the active contour model, Osher and Sethian offer a robust and effective implementation using level sets for curve evolution in [31]. The level set method is a numerical technique for tracking topological shapes in PDEs that can perform numerical computations involving curves and surfaces on a higher dimensional space without having to parameterize the points on a contour as in classic parametric active contour models. Therefore, the curve evolution equation can be converted into the following PDE:

$$\begin{cases} \dfrac{\partial \varphi(t, \mathcal{C}(t,q))}{\partial t} = \mathbf{F} |\nabla \varphi(t, \mathcal{C}(t,q))|, \\ \varphi(0, \mathcal{C}_0(p)) = 0, \end{cases} \tag{6}$$

where $|\nabla \varphi(t, \mathcal{C})|$ denotes the gradient norm vector. Then, a connection between the curves $\mathcal{C}(t,q)$ and the family of the surfaces $\varphi(t, \mathcal{C})$ is established by the curve evolution equation in (6). After several iterations, the final contour can be obtained. In our contour tracking model, we propose a new geodesic active contour model using level sets that does not require a reinitialization term based on the work in [23].

1) **Edge Term:** In our model, when the target boundaries are obvious, we want the curve to converge to those boundaries. To avoid unnecessary iterations during the evolution, we simply let the curve move on the extended rough target region $I'(t) = I_d(t) \cdot I(t)$. Additionally, we use an edge detector defined by $g(|\nabla I|) = 1/(1 + |\nabla \hat{I}|^2)$ to detect the borderline $g(|\nabla I'(t)|)$ of this extended region:

$$\begin{aligned} g(|\nabla I'(t)|) = \frac{1}{1 + |\nabla \widehat{I_d(t) \cdot I(t)}|^2} &= \frac{I_d(t)}{1 + |\nabla \hat{I}(t)|^2} + 1 - I_d(t) \\ &= I_d(t) \cdot g(|\nabla I(t)|) - I_d(t) + 1. \end{aligned} \tag{7}$$

We then define an edge term to drive the curve to the target boundaries:

$$\begin{aligned} \mathcal{Z}_e(\varphi) &\triangleq \int_{\mathbf{R}^+ \times \mathbf{R}^2} g(|\nabla I'(t)|) \cdot \delta(\varphi) |\nabla \varphi| d\mathbf{x} \\ &\triangleq \int_{\mathbf{R}^+ \times \mathbf{R}^2} (I_d(t) \cdot g(|\nabla I(t)|) - I_d(t) + 1) \cdot \delta(\varphi) |\nabla \varphi| d\mathbf{x}. \end{aligned} \tag{8}$$

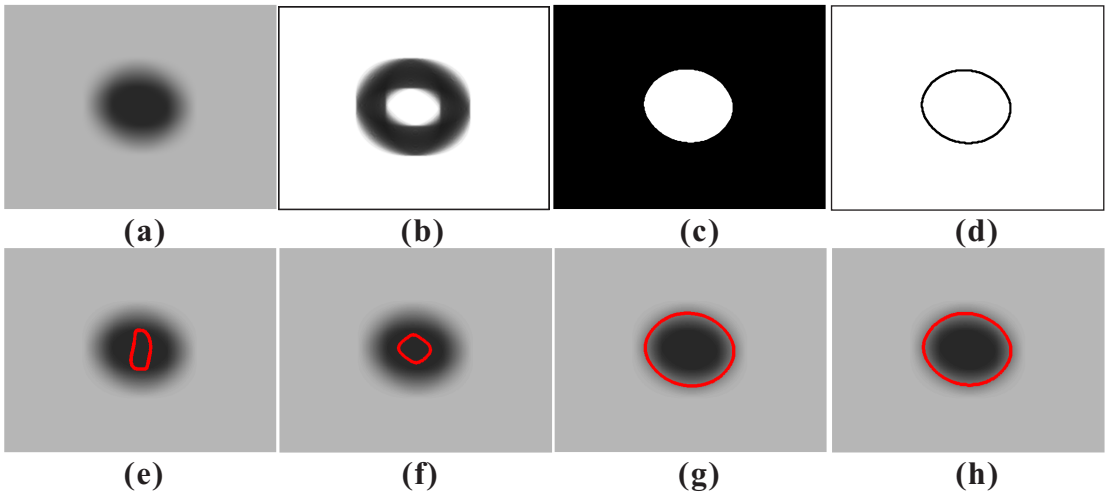

Fig. 3. Illustration of the segmentation performance of our model and several other competitive models on the blurred target: **a** The tested blurred target image I; **b** The edge information I^g of the blurred image extracted by edge-detector function; **c** The classified image using our discriminative model, as described in Sect. III-B; **d** The edge information I^g_{SVM} of the classifier image I_{SVM}; segmentation performance of **e** GAC [8]; **f** DRLSE [23]; **g** C-V [9]; **h** our proposed method.

2) **Region Term:** In conventional edge-based active contour models (GAC [8], DRLSE [23]), the curve moves toward the object boundaries until reaching the image borderline detected by the edge detector. However, in some situations, the target boundaries may be ignored because the curve may stop at the background boundaries or pass through the blurred boundaries of the target, as shown in Figs. 3(e) and 3(f). To overcome restrictions caused by sophisticated background or blurred target, we embed region information obtained from our discriminative model into the level set evolution procedure.

Assuming that we have obtained the tracking result of prior frames from frame 1 to frame t of the video sequence, we want to segment the target from frame $t+1$. Now, we reconsider our discriminative model in Sect. III-B. After classifying the image by the SVM-based discriminative appearance model, we can obtain the rough region of the target even if the background is sophisticated or the target is seriously blurred, as shown in Fig. 4(b) and Fig. 3(c). Note that this approximate region provides important boundary information of the target that is lost in the DRLSE geodesic active contour model. However, this region information cannot be embedded into the edge-based geodesic active contour model in a straightforward manner. To address this constraint, we transform the region information into homologous edge information, which allows the geodesic active contour model to incorporate more information of the target. The homologous edge information can be represented as follows:

$$\begin{aligned} g_{region} =& \overbrace{g(|\nabla I_{SVM}(t)\cdot I(t)|)}^{T_1} + \overbrace{g(|\nabla I_{SVM}(t)|)-1}^{T_2} \\ =& I_{SVM}(t)\cdot g(|\nabla I(t)|) - I_{SVM}(t) + 1 + g(|\nabla I_{SVM}(t)|) - 1 \\ =& I_{SVM}(t)\cdot g(|\nabla I(t)|) + g(|\nabla I_{SVM}(t)|) - I_{SVM}(t), \end{aligned} \tag{9}$$

where T_1 and T_2 denote the homologous edge information of the discriminative region I_{SVM} and its boundaries, respectively. After the transformation, the discriminative region information can be

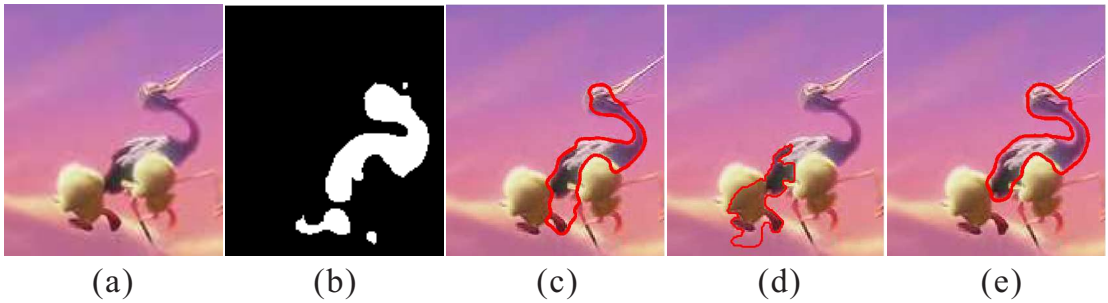

Fig. 4. Illustration of the segmentation performance with occlusion of our model and several other competitive models: **a** original image with occlusion; **b** discriminative region information; segmentation performance of **c** DRLSE [23] **d** G-CV [10] **e** the proposed method.

embedded into our active contour model:

$$\begin{aligned}\mathcal{Z}_r(\varphi) &\triangleq \int_{\mathbf{R}^+\times\mathbf{R}^2} g_{region}\cdot\delta(\varphi)|\nabla\varphi|d\mathbf{x}\\ &\triangleq \int_{\mathbf{R}^+\times\mathbf{R}^2} (I_{SVM}(t)\cdot g(|\nabla I(t)|)+g(|\nabla I_{SVM}(t)|)-I_{SVM}(t))\cdot\delta(\varphi)|\nabla\varphi|d\mathbf{x}.\end{aligned} \tag{10}$$

3) **Non-reinitialization Term:** In traditional level-set-based active contour models, periodically reinitializing the level set function during the curve evolution is necessary for maintaining the accuracy of the result but requires excessive computation time. To avoid this problem, we add an intrinsic distance regularization term [23] to our model as follows:

$$\mathcal{R}(\varphi) \triangleq \int_{\mathbf{R}^+\times\mathbf{R}^2} p(|\nabla\varphi|)d\mathbf{x}, \tag{11}$$

where the potential function p is defined as

$$p(\theta) = \begin{cases} \dfrac{1}{(2\theta)^2}(1-\cos(2\pi\theta)) & \text{if } \theta \le 1,\\ \dfrac{1}{2}(\theta-1)^2 & \text{if } \theta \ge 1.\end{cases} \tag{12}$$

Li et al. apply the distance regularization term to the edge-based contour model (DRLSE), which performs quite well when the boundaries of a target are sufficiently clear to recognize the target. However, in some situations, such as with blurred target or sophisticated background, the model may fail to segment the target due to the vague edge information (e.g., for a blurred target, its edge information is difficult to recognize, as shown in Fig. 3(b)). Compared to DRLSE, our model can effectively address these constraints, as shown in the following section.

4) **Energy Functional and Curve Evolution:** In the proposed discriminative geodesic active contour model (DGACM), the edge term Eq. (8), discriminative region term Eq. (10), and non-initialization term Eq. (11) are integrated into a new energy functional:

$$\mathcal{E}(\varphi) = \mu\mathcal{R}(\varphi)+\alpha\mathcal{Z}_e(\varphi)+\beta\mathcal{Z}_r(\varphi)+\lambda\mathcal{T}(\varphi), \tag{13}$$

where the term $\mathcal{T}(\varphi)$ is an area acceleration term defined by

$$\mathcal{T}(\varphi) \triangleq \int_{\mathbf{R}^+\times\mathbf{R}^2} g(|\nabla I(t)|)H(-\varphi)d\mathbf{x}. \tag{14}$$

The function H is called the Heaviside function, which is the integral of the Dirac delta function θ defined in [23].

To optimize the energy functional $\mathcal{E}(\varphi)$, we apply the finite difference calculation framework to the level set method:

$$\frac{1}{\Delta t}(\varphi_{t+1}-\varphi_t)=\mathcal{D}(\varphi_t), \quad k=1,2,\ldots, \tag{15}$$

where Δt is the step size of the finite difference calculations and the function $\mathcal{D}(\varphi_t)$ is the approximation of the level set function. By substituting the Dirac delta function $\delta_\varepsilon(\varphi)$ and the Heaviside function $H_\varepsilon(\varphi)$ into the energy function $\mathcal{E}(\varphi)$, we obtain the new gradient flow as follows:

$$\begin{aligned}\frac{\partial\varphi}{\partial t}=\mu\mathrm{div}(d_p(|\nabla\varphi|)\nabla\varphi)+\delta_\varepsilon(\varphi)\Bigg[\alpha\mathrm{div}\left(g(|\nabla I'|)\cdot\frac{\nabla\varphi}{|\nabla\varphi|}\right)\\ +\ \beta\mathrm{div}\left(g_{region}\cdot\frac{\nabla\varphi}{|\nabla\varphi|}\right)\Bigg]+\lambda g(|\nabla I|)\delta_\varepsilon(\varphi),\end{aligned} \tag{16}$$

where d_p is defined by $d_p(\theta)\triangleq p'(\theta)/\theta$. Using the finite difference calculation framework [22], the energy $\mathcal{E}(\varphi)$ will impede the shrinking or expanding of the zero level contour when the curve arrives at the target boundaries.

In our model, the terms $\mathcal{Z}_e(\varphi)$ and $\mathcal{Z}_r(\varphi)$ denote the boundary constraints of the original image $I(t)$ and the classified image $I_{SVM}(t)$, respectively. When the boundary information of the image is sufficiently manifested, the term $\mathcal{Z}_e(\varphi)$ enables the curve to converge to the boundaries of the target, identical to the procedure in conventional geodesic active contour models. In addition, the boundary information of the classified image provides discriminative information between the target and the background, which can improve the accuracy of our contour tracking result. In some situations, even the boundary information is very rough, our proposed geodesic active contour model can also accurately obtain the target contour under the constraint of the term $\mathcal{Z}_r(\varphi)$. Fig. 3 shows that both DRLSE- and GAC-based active contour models fail to segment the blurred ellipse due to the lack of target region information. Fig. 3(h) shows that our proposed discriminative geodesic active contour model can segment the target under the region constraint of term $\mathcal{Z}_r(\varphi)$. Note that the region-based C-V model performs quite well due to the strong contrast between foreground and background; however, this model fails to correctly segment the foreground in most tracking situations, which will be discussed in more detail in Sect. V. With both the edge information and discriminative region information, our model can also address occlusion, as shown in Fig. 4, whereas both the edge- and region-based models fail to accurately segment the occluded target.

Compared with conventional active contours based on edge or region information, our discriminative geodesic active contour model (DGACM) combines both edge information and region information and also does not require a reinitialization procedure during the curve evolution, which avoids numerical errors while maintaining sufficient accuracy of the segmentation result.

D. Target Location and Correction

To reduce the impact of a sophisticated background and the computational cost of our discriminative appearance model, the target should be located before the rough region is extracted. In our model, we use a particle filter to locate the rectangular target region R_t in the upcoming frame $I(t)$. In the particle filter tracking method, particles $\mathbf{x}_t^i$ are sampled from $I(t)$ to estimate the maximum posterior density of the target state:

$$p(\mathbf{x}_t|\mathbf{y}_{0:t})\propto p(\mathbf{y}_t|\mathbf{x}_t)\int_{\mathbf{x}_t}p(\mathbf{x}_t|\mathbf{x}_{t-1})p(\mathbf{x}_{t-1}|\mathbf{y}_{0:t-1})d\mathbf{x}_t, \tag{17}$$

which can be approximated as

$$p(\mathbf{x}_t|\mathbf{y}_{0:t})\propto\sum_{i=1}^{N}\omega_t^i\delta(\mathbf{x}_t-\mathbf{x}_t^i). \tag{18}$$

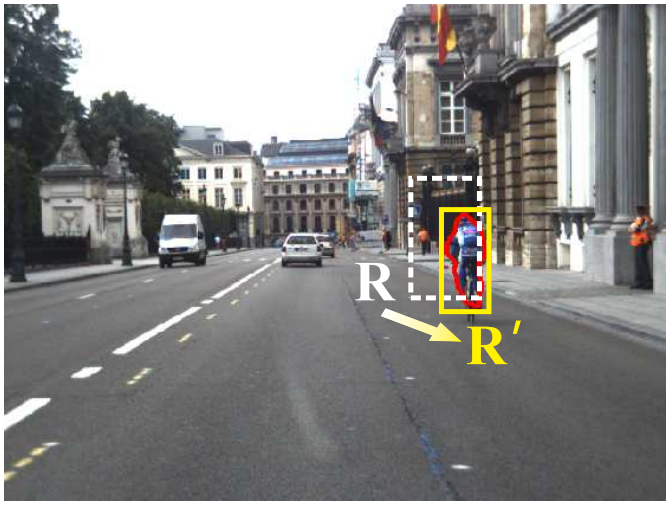

Fig. 5. The target location is corrected according to the segmentation result.

The importance weights ω_t^i are defined by

$$\omega_t^i \propto \omega_{t-1}^i \frac{p(\mathbf{y}_t|\mathbf{x}_t^i)p(\mathbf{x}_t^i|\mathbf{x}_{t-1}^i)}{q(\mathbf{x}_t^i|\mathbf{x}_{t-1}^i,\mathbf{y}_t)} \qquad i = 1,2,\ldots,N, \tag{19}$$

where $q(\cdot)$ is called the importance density and N is the number of particles. For every sample $\mathbf{x}_t^i$, a 128-dimensional RGB color histogram and 30-dimensional LBP histogram-based feature are extracted. Additionally, as in many other studies, we use the Bhattacharyya distance to approximate $p(\mathbf{y}_t|\mathbf{x}_t^i)$:

$$p(\mathbf{y}_t|\mathbf{x}_t^i) \propto \exp -\varepsilon B^2(\mathbf{y}_t,\mathbf{x}_t^i). \tag{20}$$

As discussed in many other papers [37], [42], the accumulated errors may cause drifting problem under various conditions during tracking. Note that after segmenting the target by the proposed discriminative active contour model, a more accurate target location R_t' is obtained, as shown in Fig. 5. To solve the drifting problem, we use R_t' to correct the target location and update the target template in our particle filter tracker.

However, the target template may be affected by accidental segmentation failure. To avoid this problem, we only update the target template when the following criterion is satisfied:

$$\min\left\{\frac{|R_t \cap R_t'|}{|R_t|}, \frac{|R_t \cap R_t'|}{|R_t'|}\right\} \geq 0.5, \tag{21}$$

where $|R_t \cap R_t'|$ denotes the intersection of regions R_t and R_t'. After updating the template, all particles are resampled.

To provide a better overview of our proposed method, we present our complete approach in Algorithm 1.

IV. PARAMETER ANALYSIS

A. Choosing an Appropriate Size for a Block

In the feature selection and classification procedure presented in Sect. III-B, we divide the region that contains the target into a group of $l \times l$ overlapped blocks. Then, features are extracted from those blocks to train the classifier. There are generally two types of evaluation criteria for these features in our discriminative model: discriminative ability and anti-various ability. The anti-various

Algorithm 1 Framework of SVM-based Dynamic Discriminative Geodesic Active Contour Tracking

Input: The set of the video sequence for tracking, $\mathbf{I}$. The update frequency of our SVM classifier, k. The forgetting factor ρ in the classifier updating procedure.

Output: The contour set of the target in the video sequence, $\mathcal{C}$.

Initialize the initial curve $\mathcal{C}_0$ of the target in frame $I(1)$ using GrabCut.
for frame index $t = 2, 3, \ldots$ **do**
  Use the particle filter method to estimate the target location R_t and expand its region to obtain more robust segmentation.
  if t is a multiple of k **then**
    1) $p \leftarrow t/k$
    2) Extract the p-th feature bag $\mathbf{S}_p = \{\mathbf{S}_p^+, \mathbf{S}_p^-\}$ from the blocks of the frame $I(t-1)$.
    3) Update the training set using Eq. (21).
    4) Train the SVM classifier according to the training set $\mathbf{S}_p$.
  end if
  Extract features $\mathbf{S}$ from the blocks in the target region R_t of frame $I(t)$.
  Classify the features $\mathbf{S}$ by the SVM classifier to obtain the classified region I_{SVM}.
  Erode and dilate the classified region I_{SVM} and initialize the active contour model with I_d.
  Segment the target from the region R_t using our proposed discriminative geodesic active contour model (DGACM) in Eq. (13), obtaining target contour $\mathcal{C}_t$ and accurate location R_t'.
  Use R_t' to correct the target location and update the particle filter tracker.
end for

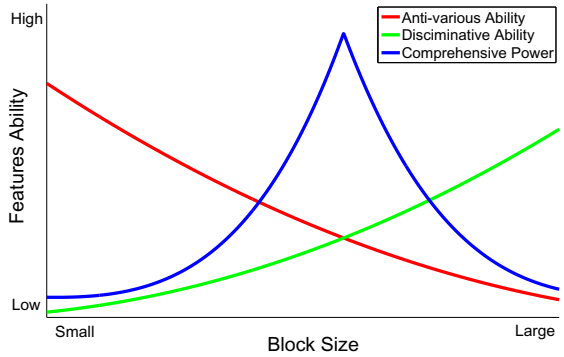

Fig. 6. The trade-off between discriminative (*green*) and anti-various (*red*) abilities of the features. The comprehensive ability (*blue*) is also shown in the figure.

ability indicates the feature's capacity to maintain discriminative power when the target appearance and texture vary. However, as shown in Fig. 6, there is a trade-off between those two criteria that when the block size is too large, features would contain more discriminative information but would heavily rely on the appearance texture, which may lead to poor classification result when the target appearance varies, as shown in the *Seq_sb* sequence in Table 1. In contrast, if the block size is too small, features would be independent of the target texture; nevertheless, similar to pixel-wise features, they have low discriminative ability. By balancing the trade-off between the two evaluation criteria, the features would have both discriminative and anti-various abilities. Here, we define the comprehensive

TABLE I
MIS-TRACKED PIXELS RATE (MPR) ON TEN VIDEO SEQUENCES WITH DIFFERENT BLOCKS.

Sequence	1×1	3×3	5×5[a]	7×7	9×9	11×11
Sunny Day	0.1076[b]	0.1102	**0.1068**[c]	0.5654	0.8512	0.9717
Brussels	0.9196	0.3339	**0.2054**	0.5658	0.3378	0.9324
Blurred	0.1059	0.1071	0.1053	**0.0988**	0.1054	0.2866
Seq_sb	1.0005	0.3854	**0.1010**	2.4574	3.3214	4.7912
Bird_2	0.8848	0.4828	**0.2264**	1.5117	1.6710	2.0418
Panda	0.9994	1.4176	**0.1222**	1.6184	0.8543	0.8022
Surfer	1.1182	0.4221	**0.1256**	0.2468	0.2791	0.3549
Pedxing1	0.9998	0.3175	**0.0690**	0.1291	0.1304	0.1603
Pedxing2	1.0697	0.7756	**0.1108**	0.6734	0.7911	0.7773
Browse1	1.0579	0.4826	**0.0436**	0.1807	0.6068	0.3883
overall	0.8263	0.4835	**0.1216**	0.8048	0.8948	1.1507

[a] We can see that when $l=5$ the MPRs reach the minimum on different video sequences.
[b] The second best result is labeled with red font for each video sequence.
[c] The top best result is labeled with bold font for each video sequence.

power of the features as follows:

$$F_{power} = \exp\left\{\min\{F_{dist}, F_{anti}\}^2 + \frac{|F_{dist} - F_{anti}|}{\theta}\right\}, \tag{22}$$

where F_{dist} and F_{anti} denote the discriminative ability and anti-various ability of the features, respectively. To find an appropriate value of the block side length to achieve a better performance, we test six different types of blocks on ten video sequences in our implementations: 1×1, 3×3, 5×5, 7×7, 9×9 and 11×11. Table 1 shows the MPRs of our proposed contour tracking method when implemented on the video sequences with different block sizes. We can see that the MPRs are almost minimized when we divide the image using 5×5 blocks.

B. Importance of Morphological Operation

In our discriminative appearance model, we use the morphological operation (opening) to eliminate the noise. To demonstrate the importance of the morphological operation defined in Eq. (2), we compared the proposed method and DGACM without the MOR method. Comparing with the proposed method, DGACM without MOR is unable to efficiently extract the target, except on the video *Blurred*, as shown in Fig. 7(a). It is shown that by incorporating the morphological erosion and dilation operations, our discriminative contour tracker can achieve more robust tracking.

C. The Ratio Between $\mathcal{Z}_e(\varphi)$ and $\mathcal{Z}_r(\varphi)$

In our energy functional $\mathcal{E}(\varphi)$ defined in Eq. (13), both the edge constraint term $\mathcal{Z}_e(\varphi)$ and the region constraint term $\mathcal{Z}_r(\varphi)$ are embedded into the geodesic active contour model. Those two terms control the convergence speed and magnitude during the curve evolution. The term $\mathcal{Z}_e(\varphi)$ allows the curve to converge to the boundaries of the image I, and the term $\mathcal{Z}_r(\varphi)$ guides the curve to converge according to the boundary information of the classified image I_d. In our model, the values for α and β will affect the convergence performance in some experimental situations. For example, in Fig. 8, our dynamic SVM classifier cannot precisely extract the foreground region due to the similarities between the target and background. If τ is too large, then the curve tends to stop at the boundaries of image I, which may contain background regions. Otherwise, the curve may simply converge to the foreground of classified image I_d, as shown in Fig. 8. In our experiments, we test different values of the parameters α and β, as shown in Fig. 7(b), and we find that more stable and accurate result can usually be obtained when $\alpha \in [1.5, 5]$ and $\beta \in [1, 2.5]$. Using those two terms, our model can describe the target with greater detail compared to the conventional model [6]. As shown in Fig. 9, comparing

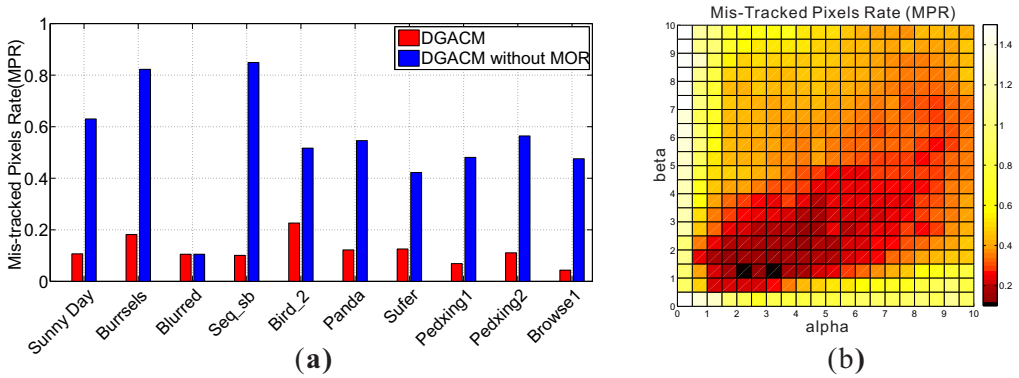

Fig. 7. (a) MPRs of DGACM without MOR and DGACM methods on tested video sequences. In most sequences, the opening operation significantly reduce the mis-tracked pixel rates. (b) We test different values for α and β on the video clips. When $\alpha \in [1.5, 5]$ and $\beta \in [1, 2.5]$, the proposed method makes good performance.

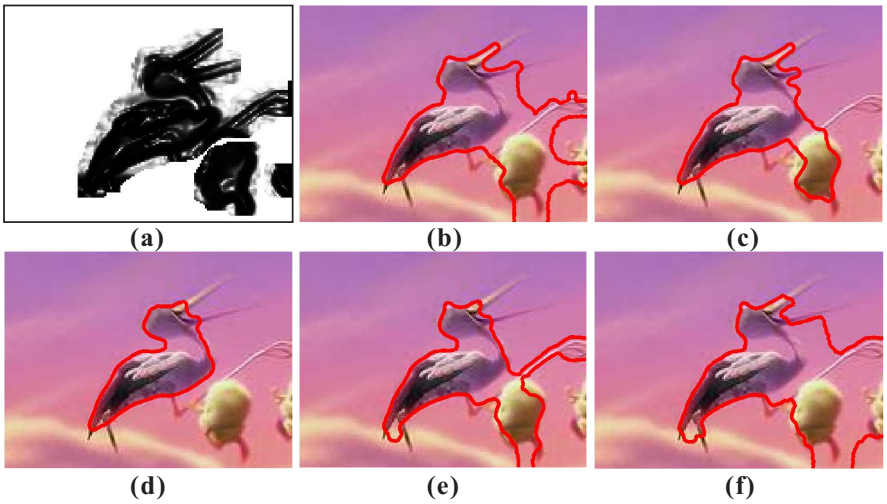

Fig. 8. The segmentation performance on the image from *Bird_2* with different values of α and β: **a** the edge information $g_d(|\nabla(I)|)$ of the classified image; **b** $\alpha = 1,\ \beta = 20,\ \tau = 0.05$; **c** $\alpha = 1,\ \beta = 10,\ \tau = 0.1$; **d** $\alpha = 3,\ \beta = 1,\ \tau = 3$; **e** $\alpha = 10,\ \beta = 1,\ \tau = 10$; **f** $\alpha = 20,\ \beta = 1,\ \tau = 20$. As we can see in this figure, when $\tau = 3$, $\alpha = 3$, and $\beta = 1$, good segmentation result is obtained.

with [6], the segmentation performance of our model is more clearly improved in the gray video clip.

V. EXPERIMENTAL RESULTS

This section consists of three parts. In the first part, we introduce the experimental setup of our method. To demonstrate the improvement obtained using our approach, in the last two parts, we evaluate the performance of our proposed method on ten video sequences with different challenges for object contour tracking. The challenges in these videos in our experiments include a blurred target, illumination variations, appearance variations, pose variations, and sophisticated background.

Fig. 9. Tracking performance on STC-3 sequence. Top row: result of [6]; bottom row: our tracking result.

A. Experimental Setup

We implement the proposed method in MATLAB R2010b under 2.6.18 kernel based Red Hat Enterprise Linux platform on an Intel(R) Core(TM) i7-2600 3.4 GHz processor with 4 GB memory. In our method, each frame costs 5.23 s on average.

Implementation details: As discussed in Sect. IV, in our discriminative appearance model, we set a block size $l = 5$. In the morphological opening operation, we erode and dilate the region by 5 and 12 pixels, respectively. As for our classifier updating procedure, we set the update frequency as $k = 3$. To obtain stable tracking result, we suggest readers set $\rho \in [0.5, 0.8]$. For our level-set-based discriminative active contour model, we set $\mu = 1$ and $\lambda = 1.5$. In particular, to allow our readers to better understand our experiment, we also provide the values for α and β on different video clips. *Sunny Day*: $\alpha = 5$, $\beta = 2$; *Brussles*: $\alpha = 4.5$, $\beta = 1.5$; *Blurred*: $\alpha = 5$, $\beta = 2$; *Seq_sb*: $\alpha = 3$, $\beta = 1$; *Bird_2*: $\alpha = 3.6$, $\beta = 1.5$; *Panda*: $\alpha = 3$, $\beta = 1$; *Surfer*: $\alpha = 4$, $\beta = 1.5$; *Pedxing1*: $\alpha = 4.5$, $\beta = 1.5$; *Pedxing2*: $\alpha = 3$, $\beta = 1$; *Browse1*: $\alpha = 4$, $\beta = 1.5$. When optimizing the energy functional $\mathcal{E}(\varphi)$, we set the number of inner and outer iteration steps as 8 and 40, respectively.

Compared algorithms: Seven object contour tracking algorithms are compared in our implementations: (a) tracking framework with distance regularized level set evolution in [23] (DRLSE-based method); (b) tracking framework with edge-based geodesic active contour model in [8] (GAC-based method); (c) tracking framework with region-based geodesic active contour model in [9] (C-V-based method); (d) scribble tracker based on matting approach in [13]; (e) our static contour tracking method (SDGACM); (f) our tracking framework without the morphological operations (DGACM without MOR) described in Sect. III-B; (g) our proposed method (DGACM). In algorithm (a), we replace our discriminative geodesic active contour model by the edge-based distance regularized level set evolution proposed in [23]. Algorithms (b) and (c) are also based on our contour tracking framework; however, the differences compared to our method are that they apply the edge-based [8] and region-based [9] active contour model to the curve evolution procedure. To demonstrate the improvement of our method in changing circumstances, such as illumination variations, appearance variations, and occlusion, we also test our contour tracking method without the dynamic updating procedure in algorithm (e). In addition, to illustrate the importance of the morphological operations in our model, we compare the proposed algorithm with algorithm (f), which neglects the opening operation described in Sect. III-B.

Evaluation criteria: To evaluate the performance of the implemented algorithms, we use the mis-tracked pixel rate (MPR) to represent the difference between tracking result and groundtruth. At

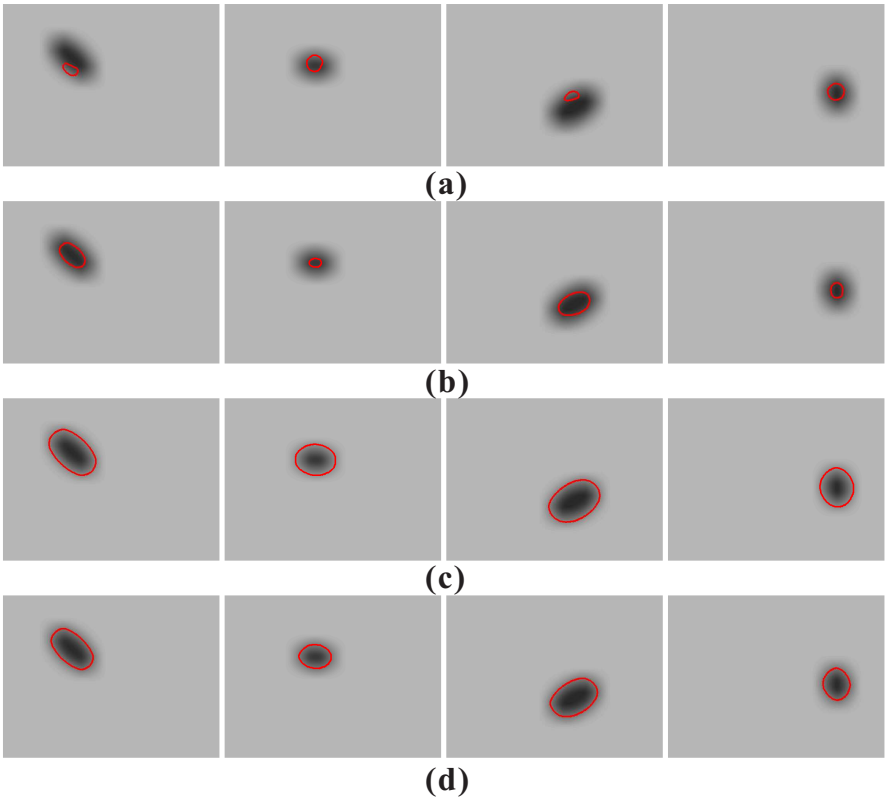

Fig. 10. Tracking performance on *Blurred* sequence: From left to right, the frame numbers are 88, 150, 260, and 380, respectively. Four competitive methods are compared: **a** GAC based method [8]; **b** DRLSE based method [23]; **c** C-V based method [9]; **d** DGACM.

time t, the mis-tracked pixels rate (MPR) can be described as

$$MPR_t = \frac{|R_t^g \cup R_t^a - R_t^g \cap R_t^a|}{|R_t^g|}, \tag{23}$$

where R_t^g denotes the groundtruth of the target, R_t^a denotes the tracking result of the implemented algorithms, and $|R_t^g \cup R_t^a - R_t^g \cap R_t^a|$ denotes the size of false negative and false positive regions. The lower the MPR is, the better the algorithm performs and vice versa.

B. Qualitative Evaluation

1) **Blurred Target:** To test the performance of our method on a blurred target video sequence, we create a 640×480 synthesized video containing 400 frames that simulates a blurred ellipse smoothed by a Gaussian filter moving from left to right along a sinusoidal curve. During its motion, the blurred ellipse also rotates and scales by itself. Fig. 10 shows that because the edge information of the blurred ellipse is not significant, as described in Sect. III-C, the curve exhibits overconvergence, and the edge-based contour tracking methods (DRLSE-based method and GAC-based method) fail to segment the target from the image. Our framework combined with the C-V model is also implemented on this video, as shown in Fig. 10(c), which performs quite well in segmenting the target. However, due to the lack of the prior knowledge of the target, the result obtained using the C-V-model-based tracker have some deviations as well.

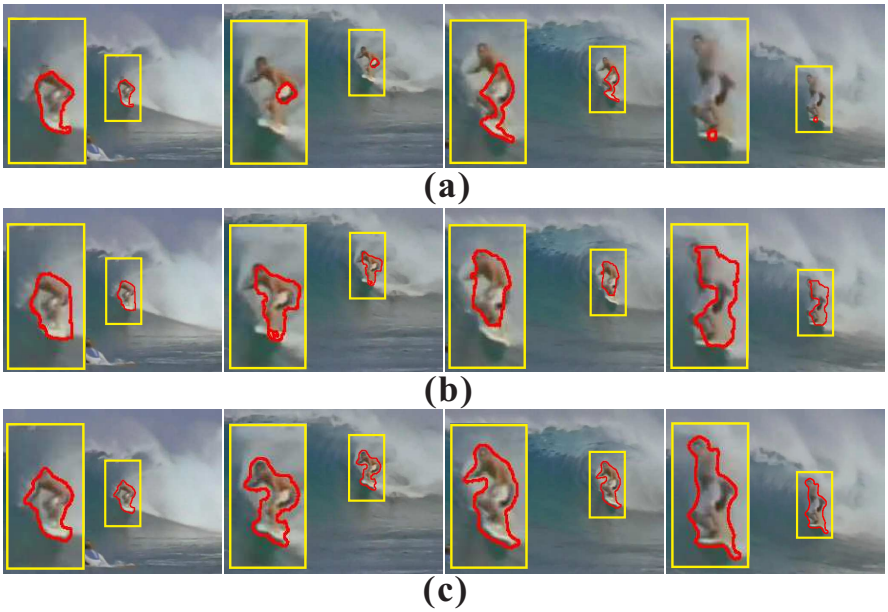

Fig. 11. Tracking performance on *Surfer* sequence: From left to right, the frame numbers are 156, 238, 257, and 313, respectively. Three competitive methods are compared: **a** DRLSE based method [23]; **b** Scribble Tracker [13]; **c** DGACM.

A real video clip, *Surfer*, in which the target is rather blurred during surfing, is also tested. Due to the lack of target region information, the DRLSE-based method fails to correctly extract the blurred limb, as shown in Fig. 11. For the same reason, the scribble tracker also fails to precisely segment the target at frame 313. In contrast, our method can track the blurred target more accurately compared to other methods. This is because our dynamic SVM classifier can precisely extract the target region and provides important clues concerning the target boundaries for the discriminative active contour model.

2) **Illumination and Appearance Variation:** Another advantage of our method is the robustness to mutative target appearance. We test two video sequences wherein the targets undergo appearance and illumination variations. Here, we show four comparative experiment results: DGACM , SDGACM, GAC-based method, scribble tracker, and DRLSE-based method. The first video sequence comes from *Sunny Day* of *ETHZ* in *ICCV 2007*, which records a couple walking along the street with a moving camera. To demonstrate the improvement of our proposed method in various situations, we tend to track the man's jacket, where the surface illumination changes when the man walks into the shadow.

As shown in Fig. 12(c), when the target moves into the shadow at frames 128, 170, and 240, the SDGACM-based method fails to extract the target when the illumination of the jacket changes. This is the reason that the SVM classifier in the SDGACM model does not update dynamically during tracking; therefore, it cannot correctly classify the features when the illumination changes, which leads to segmentation failure. In Figs. 12(a) and 12(b), because of the lack of region information of the target and interference by the background, neither the C-V- nor the DRLSE-based methods could accurately extract the target. The scribble tracker loses the target when the target appearance changes after moving into the shadow at frame 170.

In Fig. 13, the video sequence *Seq_sb* comes from *Stanford University*, where the appearance of the man's head changes when he rotates himself on a chair with a moving camera. In the SDGACM method, when the man rotates with a large amplitude at frames 41 and 56, the target appearance changes substantially, and the tracker only extracts a part of the target, as shown in Fig. 13(c). The

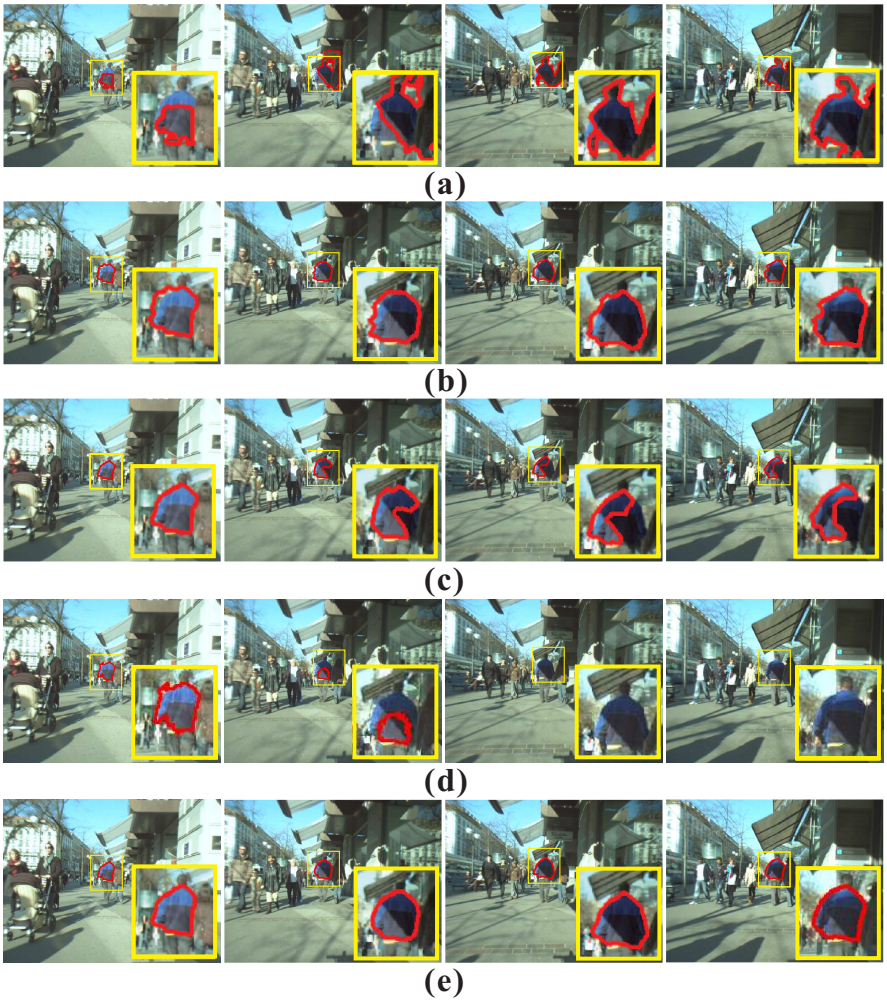

Fig. 12. Tracking performance on *Sunny Day* sequence: From left to right, the frame numbers are 50, 128, 170, and 240, respectively. Four competitive methods are compared: **a** C-V based method [9]; **b** DRLSE based method [23]; **c** SDGACM; **d** Scribble Tracker [13]; **e** DGACM.

C-V-based method and scribble tracker fail to segment the target such that the final contours in their result only converge to the background boundaries, not the true target boundary. As we can see in Fig. 13(b), the DRLSE-based method obtains a better performance compared to the C-V-based method. However, when the target boundaries are less obvious, the tracker is also prone to mis-segmenting the target region due to the lack of the region constraint term $\mathcal{Z}_r(\varphi)$. In the proposed method, the target variations can be captured and learned by the discriminative appearance model periodically, which ensures that the target can be accurately extracted under illumination variations and appearance changes.

3) **Cluttered Background:** We implement our method on two challenging video sequences to demonstrate that our method performs better on sophisticated and occluded situations compared to other competitive methods. The first video sequence comes from *TUD Brussels* and describes a man riding a bike on a street. In addition to the cluttered background, this video clip also exhibits several other challenges such as abrupt motion, scale variation, and camera variation. The second video, *Panda*,

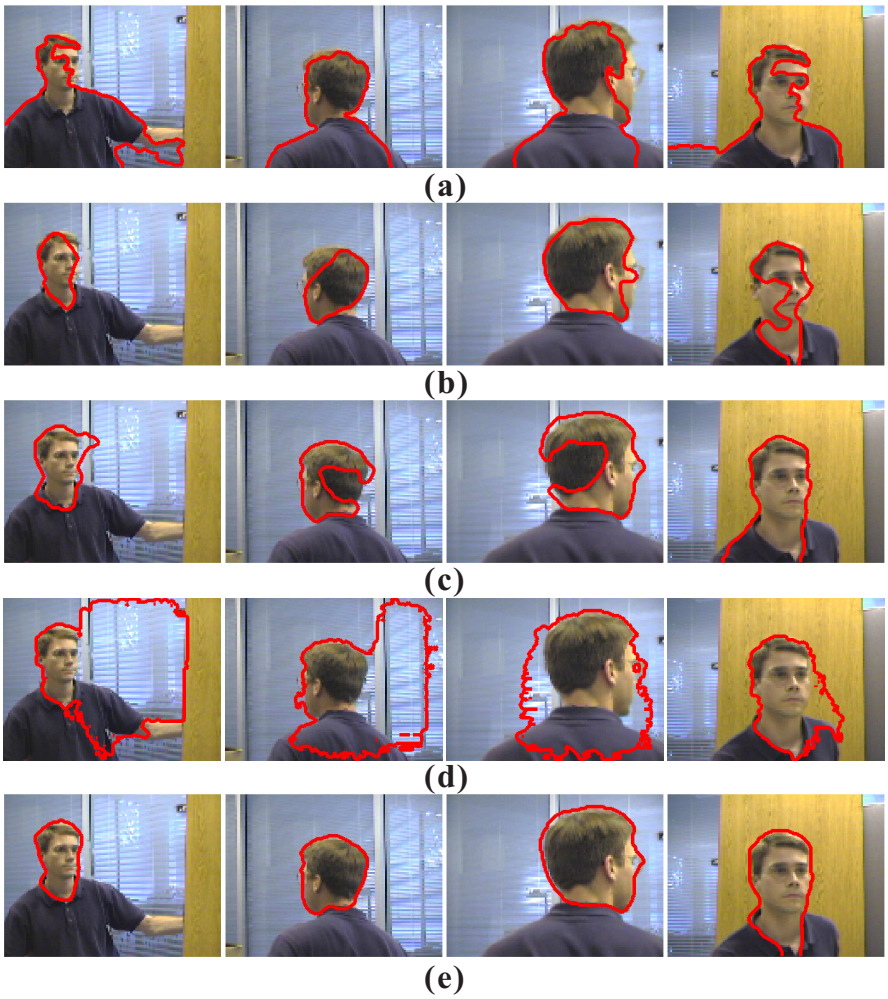

Fig. 13. Tracking performance on *Seq_sb* sequence: From left to right, the frame numbers are 23, 41, 56, and 83, respectively. Four competitive methods are compared: **a** C-V based method [9]; **b** DRLSE based method [23]; **c** SDGACM; **d** Scribble Tracker [13]; **e** DGACM.

as shown in Fig. 15, also has a sophisticated background.

As shown in Figs. 14(a) and 14(b), in the GAC-based method, due to the sophisticated situation, the curve barely converges to the edges of the background, therein completely mis-matching the target boundaries. Comparing with GAC, the DRLSE-based method obtains better tracking performance. Nevertheless, because no region information is considered in the curve evolution procedure, in some situations, such as frames 25, 52, and 70 in Fig. 15(c), the DRLSE-based method is unable to accurately drive the curve to converge to the target boundaries. In the *Brussels* video clip, the scribble tracker loses the target due to the sophisticated background, and in *Panda*, it performs better because the target appearance is obviously different from the background such that the scribble points can be effectively tracked. However, the tracker still simply roughly extracts part of the target. Due to the contributions of the edge constraint term $\mathcal{Z}_e(\varphi)$ and the region constraint term $\mathcal{Z}_r(\varphi)$, our proposed method can achieve more robust and effective contour tracking under a cluttered environment, as shown in the last rows of Figs. 14 and 15.

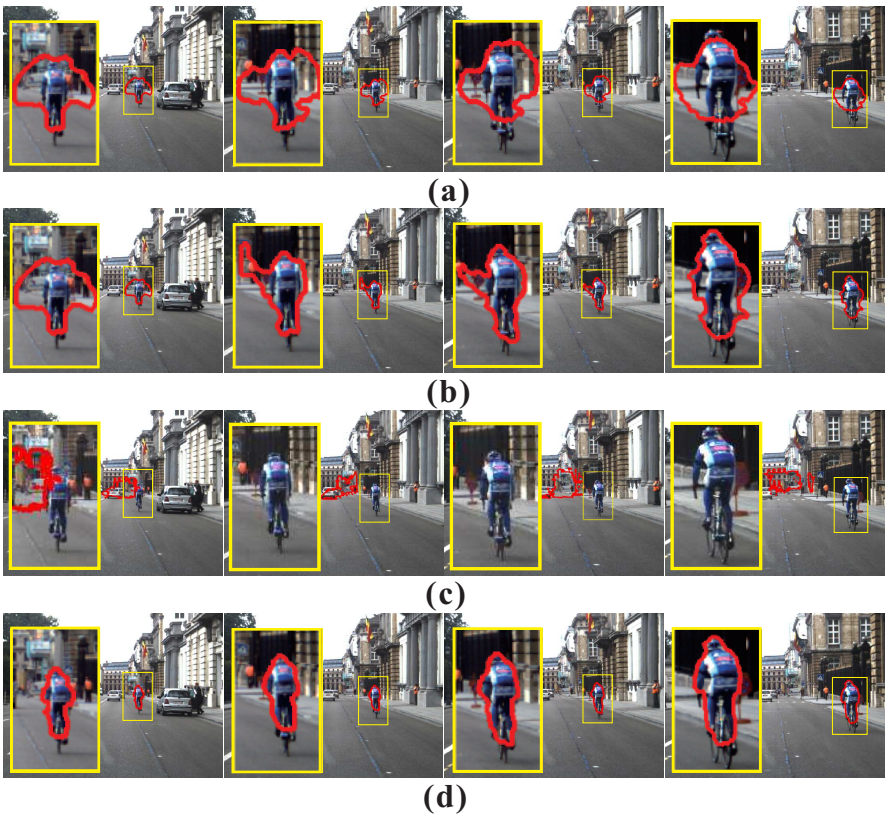

Fig. 14. Tracking performance on *Brussels* sequence: From left to right, the frame numbers are 13, 19, 22, and 27, respectively. Four competitive methods are compared: **a** GAC based method [8]; **b** DRLSE based method [23]; **c** Scribble Tracker [13]; **d** DGACM.

4) **Occlusion:** Furthermore, we also implement the proposed method, GAC, CV, DRLSE, and the scribble tracker to verify the effectiveness of the proposed method in occluded situations. In this video clip, namely, *bird_2*, a bird is heavily occluded by other small birds during walking. Figs. 16(a) and 16(c) show the contour tracking performance of the GAC- and DRLSE-based methods, respectively. We can see that without the prior region information of the target, those two methods cannot precisely extract the moving target boundaries either. The GAC-based method fails to distinguish the target from the background that the final contour contains some background regions. In comparison with GAC, the DRLSE-based method performs better that the contour could converge to the boundaries that enclose the target. Note that the C-V-based method and the scribble tracker only segment part of the bird, as shown in Figs. 16(b) and 16(d), which is caused by the uneven surface and occlusion. Compared with the other methods, it is observed that the proposed method performs better and that the bird is successfully tracked and segmented under occlusion. More tracking performance is shown in Fig. 17.

C. Quantitative Evaluation

In this section, we more clearly show our implementations. We present the MPRs of all six compared methods, which are implemented on both synthesized and real video sequences in Fig. 18. As shown in Fig. 18(a), the C-V-based method [9] performs well on the blurred sequence *Blurred*. However, due to the lack of prior target information, the obtained result still have some deviations

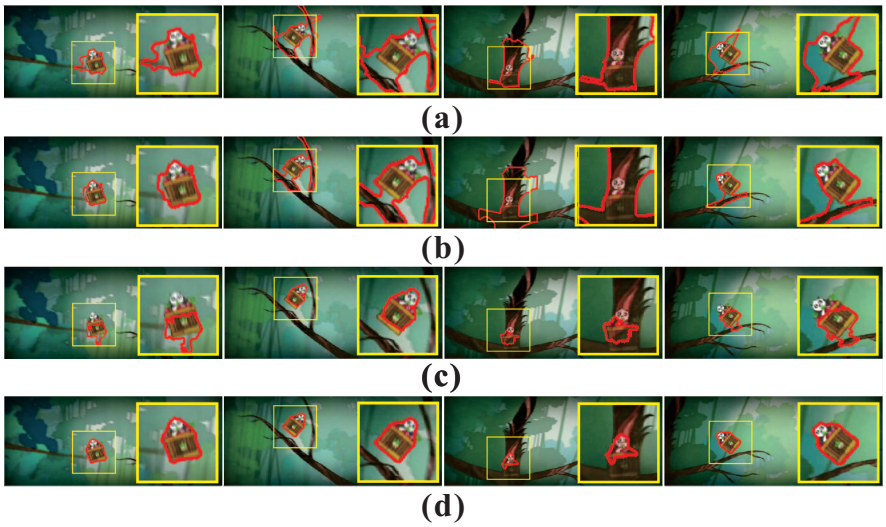

Fig. 15. Tracking performance on *Panda* sequence: From left to right, the frame numbers are 5, 25, 52, and 70, respectively. Four competitive methods are compared: **a** GAC based method [8]; **b** DRLSE based method [23]; **c** Scribble Tracker [13]; **d** DGACM.

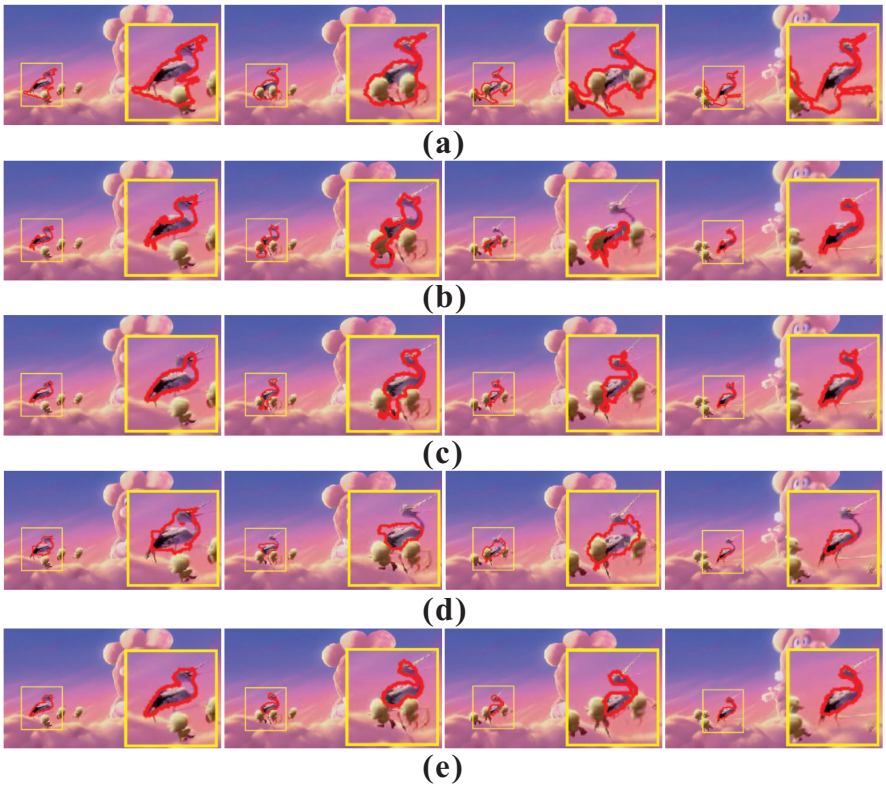

Fig. 16. Tracking performance on *Bird_2* sequence: From left to right, the frame numbers are 4, 8, 12, and 20, respectively. Four competitive methods are compared: **a** GAC based method [8]; **b** C-V based method [9]; **c** DRLSE based method [23]; **d** Scribble Tracker [13]; **e** DGACM.

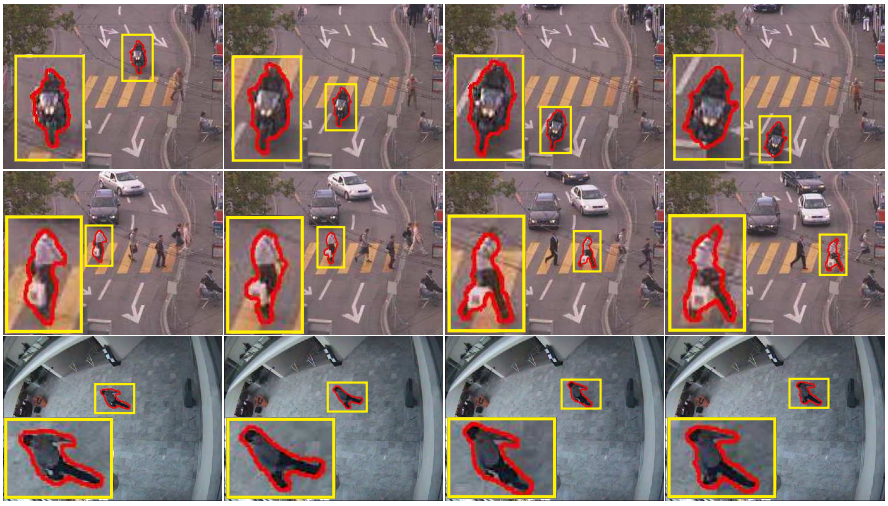

Fig. 17. Tracking performance on other sequence. From top to bottom are: *Pedxing1*, *Pedxing2* and *Browse1*, respectively.

compared to the proposed method. In other video sequences, as a result of the sophisticated background, the C-V-based method cannot precisely segment the moving target, especially in sequences *Sunny Day* and *Panda*. Comparing with the region-based tracking method, the edge-based methods, GAC [8] and DRLSE [23], can drive the curve to the edges of the image. In the conventional GAC model, the motion of the curve is vulnerable to the sophisticated background, which leads to the unstable result. As shown in Figs. 18(d) and 18(f) of sequences *Brussels* and *Bird_2*, respectively, the MPRs of the GAC-based tracker are higher than those of the implemented methods. In contrast, the DRLSE-based tracker performs better than the GAC-based tracker does in most cases. When the target background is not sophisticated and when the target boundaries are clear, as in *Seq_sb*, the DRLSE-based tracker could precisely segment the moving target, similar to our proposed tracker, as shown in Fig. 18(b). However, without the region constraint term, the DRLSE-based tracker will mis-segment the target under occlusion, as shown in Fig. 18(e) on the video *Panda*.

The SDGACM tracker also obtains competitive tracking performance when the target appearance is static (*Blurred*, *Brussels*, and *Panda*). However, in some situations, such as *Sunny Day*, where the target surface changes substantially during its motion, the SDGACM tracker failed to accurately extract the target. As shown in Fig. 18(b), the MPRs of the static tracker sharply increase when the target moves into the shadow (frames 112 to 130, 158 to 180, and 220 to 250). In the DGACM without MOR method, the tracker can drive the curve to the inner boundaries of the target, as described in Sect. III-B, which leads to some deviations. Due to the updating procedure, the accumulated error would increase during tracking. As we can see in Figs. 18(b), 18(d), and 18(e), the MPRs of the DGACM without MOR tracker gradually increase until the entire target is lost. In the scribble tracker, when the target undergoes a sophisticated background or illumination variations, the scribble points are prone to be mis-tracked, which can lead to large deviations, especially in video clips *Sunny Day*, *Seq_sb*, and *Brussels*. Moreover, as shown in Fig. 18, by combining both edge information and discriminative region information, our proposed method could obtain stable tracking performance under different challenges.

VI. CONCLUSION

In this paper, we have proposed an efficient dynamic object contour tracking framework based on the SVM appearance model and the proposed discriminative geodesic active contour model. To obtain

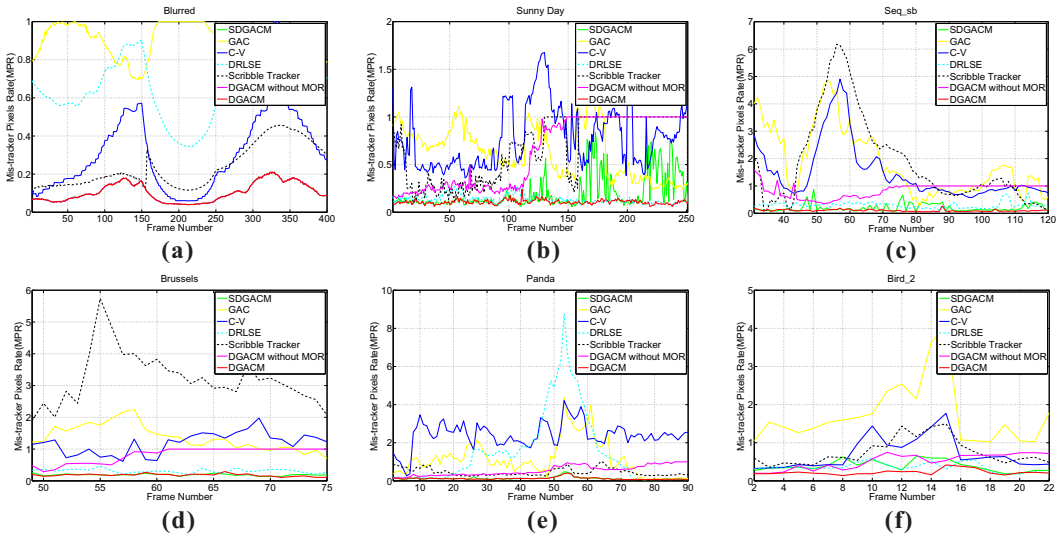

Fig. 18. MPRs of the implemented methods (Our proposed method, SDGACM, DGACM without MOR, DRLSE based method [23], GAC based method [8], C-V based method [9], and Scribble tracker [13]) on tested video sequences: **a** *Blurred*; **b** *Sunny Day*; **c** *Seq_sb*; **d** *Brussels*; **e** *Panda*; **f** *Bird_2*.

the rough target region, we use a dynamic SVM classifier to discriminate positive and negative feature descriptors extracted from the image, which performs quite well on sophisticated background and appearance variation videos. The opening operation is introduced to correct the deviations and fetch the false positives generated by the classifier, thereby providing a better environment for the proposed active contour model. In Sect. III-C, we also define a new non-reinitialization energy functional to drive the curve to the target boundaries by integrating both the edge constraint term $\mathcal{Z}_e$ and the discriminative region constraint term $\mathcal{Z}_r$, therein obtaining a better performance compared to several other state-of-the-art active contour models. To solve the drift problem and to obtain more accurate target location, the tracker is dynamically updated and corrected during tracking. Further, to demonstrate the improvement of the proposed method, we also implement several competitive algorithms on both synthesized and real video sequences. The experimental result presented in the paper demonstrates that our proposed method can achieve robust and effective target contour tracking by minimizing the energy functional under different challenging conditions such as illumination variations, appearance variations, blurred target, sophisticated background, and occlusion.

There are three main directions for our future research. The next direction is to find a way to automatically adjust the block size when tracking different sized targets to achieve an improved tracking performance. To further enhance the classification, the target texture feature [12] will be considered, which would provide more precise target information in various situations, especially when the illumination changes intensely. Another direction for future work is that we plan to integrate target shape information into the active contour evolution, which would be beneficial for addressing heavily cluttered situations and improve the overall performance of our method.

VII. ACKNOWLEDGMENTS

This work was supported by the National Natural Science Foundation of China (No. 61175096 and No. 61273273) and the Specialized Fund for Joint Building Program of Beijing Municipal Education Commission. The authors would like to thank Chunming Li, Xu Jia, and Ming-Hsuan Yang et al. for providing their source codes for comparisons in our experiments.

REFERENCES

[1] Babenko, B., Yang, M.H., Belongie, S.: Visual tracking with online multiple instance learning. In: Proc. IEEE Conf. Comput. Vision Pattern Recogn., pp. 983–990 (2009)

[2] Bai, Y., Tang, M.: Robust tracking via weakly supervised ranking svm. In: Proc. IEEE Conf. Comput. Vision Pattern Recogn., pp. 1854–1861 (2012)

[3] Bibby, C., Reid, I.: Robust real-time visual tracking using pixel-wise posteriors. In: Proc. Eur. Conf. Comput. Vision: Part II, ECCV '08, pp. 831–844 (2008)

[4] Bibby, C., Reid, I.: Real-time tracking of multiple occluding objects using level sets. In: Proc. IEEE Conf. Comput. Vision Pattern Recogn., pp. 1307–1314 (2010)

[5] Butenuth, M., Heipke, C.: Network snakes: graph-based object delineation with active contour models. Mach. Vision Appl. **23**(1), 91–109 (2012)

[6] Cai, L., He, L., Yamashita, T., Xu, Y., Zhao, Y., Yang, X.: Robust contour tracking by combining region and boundary information. IEEE Trans. Circuits and Systems for Video Technology, **21**(12), 1784–1794 (2011)

[7] Caselles, V., Catté, F., Coll, T., Dibos, F.: A geometric model for active contours in image processing. Numerische Mathematik **66**(1), 1–31 (1993)

[8] Caselles, V., Kimmel, R., Sapiro, G.: Geodesic active contours. In: Proc. IEEE Int. Conf. Comput. Vision, pp. 694–699 (1995)

[9] Chan, T.F., Vese, L.A.: Active contours without edges. IEEE Trans. Image Process. **10**(2), 266–277 (2001)

[10] Chen, L., Zhou, Y., Wang, Y., Yang, J.: Gacv: geodesic-aided c–v method. Pattern Recognition **39**(7), 1391–1395 (2006)

[11] Chen, Q., Sun, Q.S., Heng, P.A., Xia, D.S.: Two-stage object tracking method based on kernel and active contour. IEEE Trans. Circuits Syst. Video Technol. **20**(4), 605–609 (2010)

[12] Cremers, D., Rousson, M., Deriche, R.: A review of statistical approaches to level set segmentation: integrating color, texture, motion and shape. Int. J. Comput. Vision **72**(2), 195–215 (2007)

[13] Fan, J., Shen, X., Wu, Y.: Scribble tracker: a matting-based approach for robust tracking. IEEE Trans. Pattern Anal. Mach. Intell. **34**(8), 1633–1644 (2012)

[14] Fan, J., Wu, Y., Dai, S.: Discriminative spatial attention for robust tracking. In: Proc. Eur. Conf. Comput. Vision, pp. 480–493 (2010)

[15] Fossati, A., Schönmann, P., Fua, P.: Real-time vehicle tracking for driving assistance. Mach. Vision Appl. **22**(2), 439–448 (2011)

[16] Freedman, D., Zhang, T.: Active contours for tracking distributions. IEEE Trans. Image Process. **13**(4), 518–526 (2004)

[17] Grabner, H., Leistner, C., Bischof, H.: Semi-supervised on-line boosting for robust tracking. In: Proc. Eur. Conf. Comput. Vision: Part I, pp. 234–247. Springer-Verlag, Berlin, Heidelberg (2008)

[18] Hu, W., Zhou, X., Li, W., Luo, W., Zhang, X., Maybank, S.: Active contour-based visual tracking by integrating colors, shapes, and motions. IEEE Trans. Image Process. **22**(5), 1778–1792 (2013)

[19] Huang, S.S., Fu, L.C., Hsiao, P.Y.: Region-level motion-based foreground segmentation under a bayesian network. IEEE Trans. Circuits Syst. Video Technol. **19**(4), 522–532 (2009)

[20] Kichenassamy, S., Kumar, A., Olver, P., Tannenbaum, A., Yezzi, A.: Gradient flows and geometric active contour models. In: Proc. IEEE Int. Conf. Comput. Vision, pp. 810–815 (1995)

[21] Krotosky, S.J., Trivedi, M.M.: Person surveillance using visual and infrared imagery. IEEE Trans. Circuits Syst. Video Technol. **18**(8), 1096–1105 (2008)

[22] LeVeque, R.J.: Finite difference methods for ordinary and partial differential equations: steady-state and time-dependent problems. SIAM (2007)

[23] Li, C., Xu, C., Gui, C., Fox, M.: Distance regularized level set evolution and its application to image segmentation. IEEE Trans. Image Process. **19**(12), 3243–3254 (2010)

[24] Li, X., Dick, A., Wang, H., Shen, C., van den Hengel, A.: Graph mode-based contextual kernels for robust svm tracking. In: Proc. IEEE Int. Conf. Comput. Vision, pp. 1156–1163 (2011)

[25] Malladi, R., Sethian, J., Vemuri, B.: Shape modeling with front propagation: a level set approach. IEEE Trans. Pattern Anal. Mach. Intell. **17**(2), 158–175 (1995)

[26] Mansouri, A.R.: Region tracking via level set pdes without motion computation. IEEE Trans. Pattern Anal. Mach. Intell. **24**(7), 947–961 (2002)

[27] Niethammer, M., Tannenbaum, A., Angenent, S.: Dynamic active contours for visual tracking. IEEE Trans Autom. Control **51**(4), 562–579 (2006)

[28] 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)

[29] Rother, C., Kolmogorov, V., Blake, A.: Grabcut: Interactive foreground extraction using iterated graph cuts. In: ACM Trans. Graph., vol. 23, pp. 309–314 (2004)

[30] Santner, J., Leistner, C., Saffari, A., Pock, T., Bischof, H.: Prost: Parallel robust online simple tracking. In: Proc. IEEE Conf. Comput. Vision Pattern Recogn., pp. 723–730 (2010)

[31] Sethian, J.A.: Level set methods and fast marching methods: evolving interfaces in computational geometry, fluid mechanics, computer vision, and materials science, vol. 3. Cambridge university press (1999)

[32] Stenger, B., Thayananthan, A., Torr, P.H.S., Cipolla, R.: Model-based hand tracking using a hierarchical bayesian filter. IEEE Trans. Pattern Anal. Mach. Intell. **28**(9), 1372–1384 (2006)

[33] Tang, F., Brennan, S., Zhao, Q., Tao, H.: Co-tracking using semi-supervised support vector machines. In: Proc. IEEE Int. Conf. Comput. Vision, pp. 1–8 (2007)

[34] Tao, H., Sawhney, H., Kumar, R.: Object tracking with bayesian estimation of dynamic layer representations. IEEE Trans. Pattern Anal. Mach. Intell. **24**(1), 75–89 (2002)
[35] Tian, Y., Duan, F., Zhou, M., Wu, Z.: Active contour model combining region and edge information. Mach. Vision Appl. **24**(1), 47–61 (2013)
[36] Tsagkatakis, G., Savakis, A.: Online distance metric learning for object tracking. IEEE Trans. Circuits Syst. Video Technol. **21**(12), 1810–1821 (2011)
[37] Tu, J., Tao, H., Huang, T.: Online updating appearance generative mixture model for meanshift tracking. Mach. Vision Appl. **20**(3), 163–173 (2009)
[38] Vasconcelos, N., Lippman, A.: Empirical bayesian em-based motion segmentation. In: Proc. IEEE Conf. Comput. Vision Pattern Recogn., pp. 527–532 (1997)
[39] Wang, J., Chen, X., Gao, W.: Online selecting discriminative tracking features using particle filter. In: Proc. IEEE Conf. Comput. Vision Pattern Recogn., vol. 2, pp. 1037–1042 vol. 2 (2005)
[40] Wren, C., Azarbayejani, A., Darrell, T., Pentland, A.: Pfinder: real-time tracking of the human body. IEEE Trans. Pattern Anal. Mach. Intell. **19**(7), 780–785 (1997)
[41] Yilmaz, A.: Kernel-based object tracking using asymmetric kernels with adaptive scale and orientation selection. Mach. Vision Appl. **22**(2), 255–268 (2011)
[42] Yilmaz, A., Javed, O., Shah, M.: Object tracking: A survey. ACM Comput. Surv. **38**(4), 13 (2006)
[43] Yilmaz, A., Li, X., Shah, M.: Object contour tracking using level sets. In: Asian Conf. Comput. Vision (2004)
[44] Yilmaz, A., Xin, L., Mubarak, S.: Contour-based object tracking with occlusion handling in video acquired using mobile cameras. IEEE Trans. Pattern Anal. Mach. Intell. **26**(11), 1531–1536 (2004)
[45] Zhang, T., Freedman, D.: Improving performance of distribution tracking through background mismatch. IEEE Trans. Pattern Anal. Mach. Intell. **27**(2), 282–287 (2005)

Peng Lv received the B.S. degree in School of Mathematical Sciences, Shandong University, in 2011. He won bronze in the ACM International Collegiate Programming Contest in 2010 and 2011, respectively. Now He is currently a Ph.D. candidate in School of Computer Science and Technology, Beijing Institute of Technology, China. His main research focused on target tracking, active contour model, computer vision and image processing.

Qingjie Zhao received the B.S. and M.S. degrees in Control Science and Engineering, from China University of Petroleum in 1988, and Shandong University in 1996, respectively. She received the Ph.D. degree in Computer Science and Technology from Tsinghua University, in 2003. She was a visiting fellow in the School of Computer Science and Electronic Engineering, University of Essex, U.K., from Sep. 2008 to Aug. 2009. She is currently an professor in the School of Computer Science and Technology, Beijing Institute of Technology, China. Her current research interests include computing intelligence, machine vision, system theory and application.

Dongbing Gu received the B.Sc. and M.Sc. degrees in control engineering from Beijing Institute of Technology, Beijing, China, in 1985 and 1988, respectively, and the Ph.D. degree in robotics from the University of Essex, Colchester, U.K., in 2004. In 1988, he joined the Department of Electronic Engineering, Changchun Institute of Optics and Fine Mechanics, Changchun, China, as a Lecturer, where he became an Associate Professor in 1993 and a Professor in 1999. From October 1996 to October 1997, he was an Academic Visiting Scholar with the Department of Engineering Science, University of Oxford, Oxford, U.K. He is currently a Senior Lecturer with the School of Computer Science and Electronic Engineering, University of Essex. His current research interests include multiagent systems, wireless sensor networks, distributed control algorithms, distributed information fusion, cooperative control, reinforcement learning, fuzzy-logic- and neural-network-based motion control, and model predictive control.