Non-Rigid Object Contour Tracking via a Novel Supervised Level Set Model

Xin Sun, Member, IEEE, Hongxun Yao, Member, IEEE, Shengping Zhang, Member, IEEE, and Dong Li, Member, IEEE

Abstract—We present a novel approach to non-rigid objects contour tracking in this paper based on a supervised level set model (SLSM). In contrast to most existing trackers that use bounding box to specify the tracked target, the proposed method extracts the accurate contours of the target as tracking output, which achieves better description of the non-rigid objects while reduces background pollution to the target model. Moreover, conventional level set models only emphasize the regional intensity consistency and consider no priors. Differently, the curve evolution of the proposed SLSM is object-oriented and supervised by the specific knowledge of the targets we want to track. Therefore, the SLSM can ensure more accurate convergence to the exact targets in tracking applications. In particular, we firstly construct the appearance model for the target in an online boosting manner due to its strong discriminative power between the object and the background. Then, the learnt target model is incorporated to model the probabilities of the level set contour by a Bayesian manner, leading the curve converge to the candidate region with maximum likelihood of being the target. Finally, the accurate target region qualifies the samples fed to the boosting procedure as well as the target model prepared for the next time step. We firstly describe the proposed mechanism of two-phase SLSM for single target tracking, then give its generalized multi-phase version for dealing with multi-target tracking cases. Positive decrease rate is used to adjust the learning pace over time, enabling tracking to continue under partial and total occlusion. Experimental results on a number of challenging sequences validate the effectiveness of the proposed method.

Index Terms—Object tracking, level sets, curve evolution, boosting, appearance modeling.

I. INTRODUCTION

O

BJECT tracking, which refers to the task of generating the trajectories of the moving objects in a sequence of images, is a challenging research topic in the field of computer vision. The problem and its difficulty depend on several factors, such as the amount of prior knowledge about the target object and the number and type of parameters being tracked, e.g., location, scale, detailed contour. Although there has been some success with building trackers for specific object classes, tracking generic real-world objects has remained challenging due to unstable lighting condition, pose variations, scale changes, view-point changes, and camera noise etc.

Early tracking methods [1], [2] use fixed appearance model to describe the target, which are unable to successfully track the target over long time. To overcome this drawback, some tracking algorithms try to update the target appearance over time in an online manner. The appearance models adopted by these methods include histogram [3], subspace models [4] as well as sparse representation models [5], [6]. Besides, some researchers resort to adopting discriminative learning methods to make the trackers easy to distinguish the target from its background. The methods based on boosting [7], [8] and SVMs [9], [10] show impressive performance and attract much attention. In contrast with constructing two separate models for the target and background respectively, classifier learning based approaches are more inclined to seize properties of most discrimination between them.

Despite having the promising performance, these traditional trackers face a practical problem that they use rectangular bounding box or oval to approximate the tracked target. However, objects in practice may have complex shapes that cannot be well described by simple geometric shapes, see Fig. 1(a) for some examples. Since the rectangle box used for presenting the tracked target directly determines the samples to be extracted in the subsequent target appearance modeling/update step, it is a critical factor to tracking performance. Inaccurate target presentation easily results in performance loss due to the pollution of non-object regions residing inside the rectangle box. In order to better fit the object shape, some methods adopt the scale selection mechanism that aims to search for the best scale that covers the target accurately. An intuitive idea is to run the algorithm in different scales, then select the one maximizing the object function of the tracking algorithm. Further, this selection mechanism is also extended to orientation. By simultaneously controlling both the scale and orientation, the statistic bias for the target distribution can be controlled, see Fig.1(a), and this, to some extent, makes better target description and tracking estimation. Nevertheless, all these scale/orientation adjustments are still based on simple geometric shapes (such as rectangle and oval), which will inevitably introduce a large number of background pixels when used for presenting real-world object with complex shapes.

Ideally, a better manner to describe the target is to use the accurate contour along the target’s surface. Some attempts in
Then evolve the graph so that this level set moves according to the zero level set of the graph of a higher dimensional function. The idea of the level set approach is to embed the contour as zero level sets and able to deal with changes in topology. The basic level set technique [14]–[17] is an implicit representation of parametric active contour models, such as snake model [13], in contrast with explicit representation of contours in segmenting technique for dynamic tracking [11], [12]. In the literature have been made to use silhouette or contour sequences. whose frame numbers are 511, 14, 108, 151, 418, 205 respectively in their tracking examples of the proposed method in various challenging cases. Fig. 1. Motivation: (a) shows the typical bounding box presentation on complex object with scale/orientation adaption. (b)-(g) gives some contour tracking examples of the proposed method in various challenging cases, whose frame numbers are 511, 14, 108, 151, 418, 205 respectively in their sequences.

In this paper, we present a novel supervised level set model in Section III, and propose its generalized multi-phase version in Section IV. Section V presents dense experiments conducted on a number of challenging video sequences. Section VI concludes the paper.

II. RELATED WORK

A. Tracking Methods With Online Appearance Learning

Han and Davis [20] deal with the variations of lighting condition, pose, scale, and view-point over time by approximately estimating the pixel-wise color density in a sequential manner. The work of Grabner and Bischof [21] learn a binary classifier as implicit appearance model and apply it in each new frame to locate the position of the target. Babenko et al. [22] introduce multiple instance learning into online tracking where samples are grouped into positive and negative bags or sets. Recently, a semi-supervised learning approach [23] is developed in which positive and negative samples are selected via an online classifier with structural constraints. In [24], the authors focus on the problem of long-term object tracking and propose a detection-based approach which learns appearance models from a large negative training set. In [25], the authors propose an online learning method using an incremental linear discriminant analysis for discriminating the appearances between multiple tracked objects. All these tracking methods, however, use bounding box to describe the tracked target. Scale and orientation selection mechanism have been adopted in this kind of trackers to better fit the object shapes.

B. Scale/Orientation Selection for Better Fit to Object Shape

Instead of acting towards intensity consistent direction, the curve evolution of the SLSM is target-oriented and supervised by the knowledge of the specific targets in tracking application. Boosting approach is used for online construction of the target appearance model due to its strong ability in distinguishing the target from its background. Then the learned target model is incorporated to model the level set contour probabilities by a Bayesian manner, leading the curve converge to the candidate region with maximum likelihood of being the tracked target. Finally, samples extracted from accurate target region are fed back to the boosting procedure for target appearance update. We use the positive decrease rate to adjust the target learning pace over time, which enables tracking to continue under partial and total occlusion. In this paper, we firstly describe the proposed mechanise of 2-phase SLSM for single target tracking, whose preliminary results were also presented in the early conference paper [19]. Then we novelly propose the generalized multi-phase SLSM for dealing with multi-target tracking cases. Fig.1(b-g) shows some tracking examples of our method in various challenging cases.

The remainder of this paper is organized as follows. In Section II, we review the current state of the related work. Then we introduce the proposed 2-phase supervised level set model in Section III, and propose its generalized multi-phase version in Section IV. Section V presents dense experiments conducted on a number of challenging video sequences. Section VI concludes the paper.
tracking method where the EM estimation in conjunction with KL-divergence are used to develop a target-center and kernel bandwidth update scheme. In [29], the authors extend the original mean shift approach to handle orientation space and scale space. The method estimates the motion, including the location, orientation and scale, of the interested object window simultaneously. Although these adjustments, to some extent, can help with better target presentation, these methods still use simple geometric shapes to specify the target region. Hence it is difficult for them to ideally therewith the tracked object shape. In contrast, the proposed method is capable of tracking the accurate contours of the targets.

C. Dynamic Tracking Methods for Non-Rigid Objects

Zhao et al. in [30] focus on the human shape modeling and track humans in crowded environments where occlusion persistently occurs. Ramanan et al. in [31] firstly construct an appearance model for the target, which is then used as detecting model for tracking. They report good results on articulated person tracking. In [32], the authors model human body as a combination of singleton parts and symmetric pairs of parts, and treat the human pose tracking as a multi-target tracking problem where the “targets” are associated by an underlying articulated structure. In [33], the authors propose a strategy that learns online key poses of the tracked person as multiple reference models to drive a shape tracking method of human. These methods, however, basically assume that specific classes of the targets are given. Kwon and Lee [34] use a patch-based dynamic appearance model in conjunction with an adaptive Basin Hopping Monte Carlo sampling method to track a non-rigid object, where no specific class information of the target is requested, neither does the off-line training phase. Similarly, our method is applicable to general real-world objects and utilizes the accurate contours to describe the target as well as qualified samples for learning the target appearance.

D. Close Work on Segmentation Based Tracking Methods

Bibby and Reid [35] derive a probabilistic framework for robust tracking of multiple previously unseen objects. The observed image data is used to compute a posterior over the object poses, shapes and relative depths where the shapes are implicit contours represented using level sets. In [36], Godec et al. present a novel tracking-by-detection approach to non-rigid object tracking based on the generalized Hough-transform. They couple the voting based detection and back-projection with a rough segmentation based on GrabCut [37]. Afterwards, Stefan et al. in [38] improve the above HoughTrack to a faster version by using pixel-based descriptors. In [39], the authors introduce the Mumford-Shah model [16], [17] into the particle filter framework. Once the particle filter gives the candidate positions in prediction step, the active curve evolution is included to give the candidate contours. Shape information has also been considered in the form of static [40], [41] or dynamic [42] priors. Those try to constrain the active curve using the statistics of a set of training shapes, either by performing the optimization in a subspace [40] or by bringing in the shape constraints on the variational level [41]. However, they need to encompass the entire shape variability, which is impractical for real-world applications. In contrast, our method encodes the tracked target knowledge by the manner of boosting classifier learning and supervises the curve acting direction by the current learned target model.

III. THE 2-PHASE SUPERVISED LEVEL SET MODEL

In this section, we analyze the general curve acting principle of the level set model which gives the antecedent of our improvements. Then we describe the proposed 2-phase supervised level set model in detail, which is competent in dealing with real-world object tracking problems.

A. Curve Acting Principle

In binary level set model [18], a piecewise constant-valued function \( u \) is used to approximate the intensity distribution of image \( I \). The contour \( C \), embedded as the zero level set of the level set function \( \phi \), divides the image into two regions \( \Omega_1 \) and \( \Omega_2 \). In region \( \Omega_1 \), \( \phi = 1 \) and \( u = c_1 \) while in region \( \Omega_2 \), \( \phi = -1 \) and \( u = c_2 \). So the piecewise constant-valued function can be defined as

\[
\begin{align*}
\phi & = 1 \\
\phi & = -1
\end{align*}
\]

\[
\begin{align*}
u & = \frac{c_1}{2}(\phi + 1) - \frac{c_2}{2}(\phi - 1)
\end{align*}
\]

where \( c_1 \) and \( c_2 \) are positive constants.

Then the energy function of the active contour model can be defined as

\[
\begin{align*}
E_{image} & = E_B(c_1, c_2, \phi) \\
& = \frac{1}{2} \int_{\Omega} |u(c_1, c_2, \phi) - I|^2 \, dxdy \\
& + \mu \int_{\Omega} |\nabla \phi| \, dxdy + \frac{1}{\tau} \int_{\Omega} W(\phi) \, dxdy
\end{align*}
\]

where \( \mu \) and \( \tau \) are the proportional coefficients. The first item is used to measure the similarity of the two-valued function \( u \) with the image \( I \), and makes the function \( u \) more close to the intensity distribution of image \( I \). The second item is used to measure the length of the curve \( C \), playing the role of smoothing region boundaries. The last item is for the binary constraint.

In conventional level set methods, there is no any prior knowledge taken into account and the positive constants \( c_1, c_2 \) can be obtained directly by minimizing the energy function

\[
\begin{align*}
c_1 & = \frac{\int_{\Omega_1} I(1 + \phi) \, dx \, dy}{\int_{\Omega}(1 + \phi) \, dx \, dy}, \\
c_2 & = \frac{\int_{\Omega_2} I(1 - \phi) \, dx \, dy}{\int_{\Omega}(1 - \phi) \, dx \, dy}
\end{align*}
\]

where, we can see, \( c_1 \) and \( c_2 \) are the average intensities of image \( I \) in region \( \Omega_1 \) and \( \Omega_2 \).

So when we minimize the energy function \( E_{image} \), we want the function \( u \) more close to the image \( I \), that is, the region with average intensity is close to the original image. As a result, this definition of \( u \) makes the level set model more inclined to segment out the region with consistent
intensity (see Fig. 2), which is similar to the threshold segmentation method. However, objects may consist of inconsistent intensities which occurs most often in practice. Additionally, in the context of tracking, we usually have a specific target of interest, which can be exploited to supervise the evolution of the curve and refine its acting direction.

B. Online Appearance Modeling

We construct the online appearance model in an implicit manner of boosting classification as in [7]. In contrast with distinguishing between the object and background by modeling two separate models respectively, treating the separation as a binary classification problem is more inclined to seize properties of most discrimination between them, as well as decrease pollution from the similar background pixels.

The boosting-based appearance modeling procedure can be summed up with several keys (Algorithm 1).

We apply this implicit target model as a detector in the arriving frame. For each sample evaluated, within the search region, we can obtain a confidence score indicating the likelihood it derives from the target object. In [7], mean shift algorithm is implemented for searching the peak and a rectangle is used for presenting the result. However, although the tracker locks onto most competitive samples, it may ignore the contribution of the other. Additionally, the inevitable background pixels mixed inside the rectangle may pollute the subsequent target update process. Differently, we introduce these scores into the proposed SLSM (see Fig. 3) as prior target knowledge to supervise the curve evolution and give a global optimal result of target contour, which can also supply the boosting procedure with qualified samples as well as deal with the re-scale problem in mean shift algorithm.

C. Level Set Formulation

Our goal is to estimate the target contour from a sequence of images. Let $I_k : x \rightarrow \mathbb{R}^m$ denote the image at time $k$ that maps a pixel $x = [x \ y]^T \in \mathbb{R}^2$ to a value, where the value is a scalar in the case of a grayscale image ($m = 1$) or a three-element vector for an RGB image ($m = 3$).

![Fig. 2. Illustration of the role of function $u$ used in level set energy function. (a) is the original image in size of 84×84; (b) represents the inaccurate contour while (c) represents its corresponding piecewise constant-valued function $u$, which we can see has a great difference compared to the original image; (d) represents the accurate contour and (e) represents its function $u$ respectively, which is more close to the original image.](image)

Let $C(s) = [x(s) \ y(s)]^T, s \in [0, 1]$, denote a closed curve in $\mathbb{R}^2$. An implicit function $\phi(x, y)$ is defined such that the zeroth level set of $\phi$ is $C$, that is, $\phi(x, y) = 0$ if and only if $C(s) = [x \ y]^T$ for some $s \in [0, 1]$. In response to the low efficiency of the traditional level set models, the proposed SLSM maintains the advantage of using two-valued level set function $\phi$ to replace the traditional signed distance function $\phi(x, y, k) = \begin{cases} 1, & \text{if } [x \ y]^T \text{ inside } C_k \\ -1, & \text{if } [x \ y]^T \text{ outside } C_k \end{cases} (4)$

Using this simple form can avoid the re-initialized process of the level set function in each iteration as well as the cumbersome numerical realization.

Given all the observations $I_{0:k}$ up to time $k$, boosting score map $S_k$, we model the probability of contour $C_k$ at time $k$ by considering both the region and edge cues in a Bayesian manner as

$$p(C_k | I_{0:k}, S_k) \propto p_{ib}(S_k | C_k) \ p_e(I_k | C_k) \ p(C_k) \tag{5}$$

where $p_{ib}(S_k | C_k)$ presents the likelihood that the regions inside and outside $C_k$ are the target object and background, respectively; and $p_e(I_k | C_k)$ gives the likelihood that the contour is on image edge; $p(C_k)$ is the prior probability of the contour, where we encode the length prior for smoothing region boundary. Here, the assumption we depend on is that the measurements are independent of each other.

When we maximize the probability of (5), obviously, we expect to obtain the contour that surrounds the target region and exactly converges to its edge.

Let $R^+$ be the region of the image inside the curve and $R^-$ the region outside the curve. The region-based probability $p_{ib}(S_k | C_k)$ in (5) can be decomposed as

$$p_{ib}(S_k | C_k) \propto p_t(S_k | R^+) \ p_b(S_k | R^-) \tag{6}$$

where $p_t(S_k | R^+)$ captures the target probabilities inside $C_k$, and $p_b(S_k | R^-)$ captures the background probabilities.
outside \( C_k \). Let \( S_k(x) \) denote the score value of pixel \( x \) and is positive for pixels that are more likely to belong to the target than to the background, and vice versa for negative. Assuming that \( S_k(x) \) is independent of \( S_k(y) \) for \( x \neq y \) and taking log, we have

\[
\log(p_t(S_k|R^+)) \propto \sum_{x \in R^+} S_k(x) \\
\log(p_b(S_k|R^-)) \propto \sum_{x \in R^-} -S_k(x)
\]

Under the objective of driving the contour to the target boundary, we use image gradient for edge detecting, see Fig. 3, and the edge-based probability \( p_e(I_k|C_k) \) in (5) can be computed as

\[
\log(p_e(I_k|C_k)) \propto \sum_{[x,y] \in C_k} B(x,y)
\]

where

\[
B(x,y) = |\nabla[G_\sigma(x,y) * I_k(x,y)]|^2
\]

where \( \nabla \) denotes spatial gradient operator, \( * \) denotes convolution and \( G_\sigma \) is the Gaussian filter with standard deviation \( \sigma \).

We define the energy function, minimizing which over the level set function is equivalent to maximizing the probability of (5)

\[
E(\phi, S_k) = \int_{R^+} -S_k(x)dx + \int_{R^-} S_k(x)dx + \frac{\xi}{\tau} \int_C -B(x)dx + \frac{\mu}{\tau} \int_{\Omega} \int W(\phi)dx
\]

where \( \xi, \mu \) and \( \tau \) are the coefficients that weight the relative importance of each item. \( \ell(C) \) is the length of the curve. The last item is for constraining \( \phi^2 = 1 \), where \( W \) can be defined as \( (\phi^2 - 1^2)^2 \) and \( \Omega = R^+ \cup R^- \) is the image domain.

Employing the binary level set function as a differentiable threshold operator, we unify the integral region and rewrite (11) as

\[
E(\phi, S_k) = \int_{\Omega} \int -\frac{1}{2} S_k(x)(1 + \phi)dx + \int_{\Omega} \frac{1}{2} S_k(x)(1 - \phi)dx + \frac{\xi}{\tau} \int_B -B(x)|\nabla \phi|dx + \frac{\mu}{\tau} \int |\nabla \phi|dx + \frac{1}{\tau} \int W'(\phi)dx
\]

The associated Euler-Lagrange equation for this function can be given by

\[
0 = -S_k(x) + \xi B(x)\frac{\nabla \phi}{|\nabla \phi|} - \mu \frac{\nabla \phi}{|\nabla \phi|} + \frac{1}{\tau} W'(\phi)
\]
and implemented by the following gradient descent
\[
\frac{\partial \phi}{\partial k} = S_k(x) - \xi B(x) \text{div} \left( \frac{\nabla \phi}{|\nabla \phi|} \right) + \mu \text{div} \left( \frac{\nabla \phi}{|\nabla \phi|} \right) - \frac{1}{\tau} W'(\phi)
\]
where div is the divergence operator.

In contrast with conventional level set formulations, ours, instead of being based upon intensity consistency, is supervised by the prior knowledge of the tracked target. Therefore, the curve, in SLMS, can be steered to the target from a wide variety of states, without any request of the initial curve that must be inside or outside the target completely. Fig. 3 illustrates the tracking principle of the proposed algorithm.

IV. THE MULTI-PHASE SUPERVISED LEVEL SET MODEL

In above 2-phase supervised level set model, with only one level set function, we can represent and track only one target. In this section, we generalize the 2-phase SLMS to the multi-phase version with which we can deal with multi-target tracking cases. We use several active contours to enclose and represent the multiple tracked targets. Similarly, we firstly construct the appearance models for these tracked targets, then use them as prior knowledge to supervise and refine the evolution of the active contours.

One can employ any existing appearance modeling method to construct these target models. In this work, for convenience, we use the same way as in part B of Sec. III to learn the appearance model for each tracked target. Specifically, let us consider \( N \) tracked targets. For each target \( i \), we construct its implicit appearance model \( T(i) \) by the manner of boosting classifier learning using Algorithm 1. In the next arriving frame, a larger ring of neighboring pixels surrounding the initial target region is included as an extension to form the search region of the \( i \)th target. Then the learnt target model \( T(i) \) is applied as a detector within the search region so that each sample evaluated gets a confidence score \( S(i)(x) \), indicating the likelihood of the pixel \( x \) belonging to the target \( i \). We denote the appearance models of all \( N \) targets as \( T = \{T(i)\}_{i=1}^{N} \), and the score maps of all targets as \( S = \{S(i)(x)\}_{i=1}^{N} \). We include them as the prior targets knowledge into the level set formulation to supervise the active contours evolution and obtain the multi-target contour tracking results. This is explained next.

Our goal is to estimate the \( N \) contours of the \( N \) targets from a sequence of images. Let \( \{R_i\}_{i=1}^{N} \) denote the candidate regions of the \( N \) targets. We consider that each pixel can belong to only one target. Then the tracking task in each image frame consists of finding a partition \( \{R_i^*\}_{i=1}^{N} \) of the image domain \( \Omega \), so that each region is homogeneous with respect to the corresponding target model. In this case, it is convenient to cast the tracking task in a Bayesian framework. The problem would then consist of finding a partition \( \{R_i^*\}_{i=1}^{N} \) which maximizes a posteriori probability over all possible \( N \)-region partitions of \( \Omega \)

\[
\{R_i^*\}_{i=1}^{N} = \arg \max_{R_i \subset \Omega} p((R_i)_{i=1}^{N} | T)
\]

\[
= \arg \max_{R_i \subset \Omega} p(T|R_i)_{i=1}^{N} p((R_i)_{i=1}^{N})
\]

Here, for simplicity, we do not consider the edge cue as in (5) (it may have only a very small effect in most of the cases, because the regional fitting term is dominant). Assuming that \( T(x) \) is independent of \( T(y) \) for \( x \neq y \) and taking \(-\log\) of (15), we have

\[
\{R_i^*\}_{i=1}^{N} = \arg \min_{R_i \subset \Omega} F[[R_i]_{i=1}^{N}]
\]

where

\[
F[[R_i]_{i=1}^{N}] = \sum_{i=1}^{N} \int_{x \in R_i} - \log p(T(i) | I(x)) dx - \log p((R_i)_{i=1}^{N})
\]

The first term in (17), referred to as the data term, measures the conformity of the image data \( I(x) \), in region \( R_i \), with respect to the corresponding target model \( T(i) \), \( i = 1 \ldots N \). This conformity can be naturally estimated by the confidence score \( S(i)(x) \) which indicates how likely a pixel \( x \) belongs to the target \( i \). Here note that this confidence score is only computed within a local search region surrounding the initial target region, not on the whole image domain. We fill the rest of the image region with zero score so that the score map \( S(i)(x) \) is defined over the whole image. In this case, the data term can be expressed as follows:

\[
D = \sum_{i=1}^{N} \int_{x \in R_i} - \log p(T(i) | I(x)) dx \propto \sum_{i=1}^{N} \int_{x \in R_i} -S(i)(x) dx
\]

The second term in (17) embeds prior information on the target regions. Here, we encode the length prior for smoothing regions boundaries

\[
L = - \log p((R_i)_{i=1}^{N}) = \mu \sum_{i=1}^{N} \int_{R_i} ds
\]

where \( \mu \) is a positive factor.

We use \( N \) active curves to enclose the \( N \) target regions to obtain the contour tracking results. Using several active curves can lead to ambiguity when two or more curves intersect. The main issue is to guarantee that the curves converge to the target regions, let \( \{C_i\}_{i=1}^{N} \) be \( N \) simple closed plane curves and \( \{R(C_i)\}_{i=1}^{N} \) the regions they enclose. Then, the following regions \( \{R_i\}_{i=1}^{N} \) form a partition: \( R_1 = R(C_1); R_2 = R(C_1) \cap R(C_2); \ldots; R_i = R(C_1) \cap \ldots \cap R(C_{i-1}) \cap R(C_i); \ldots; \) and \( R_N = R(C_1) \cap \ldots \cap R(C_{N-1}) \cap R(C_N) \), where \( R(C_i) \) is the complementary of \( R_i \) in image domain \( \Omega \). The target regions defined in this way form a disjoint decomposition of \( \Omega \), and each pixel \((x, y) \in \Omega \) will belong to at most one target. This is illustrated in Fig.4(a) for three target regions.

We use level set manner to represent the active curves. Let us consider \( N \) level set functions \( \phi_i : \Omega \rightarrow \mathbb{R}, i = 1 \ldots N \).
The set of curves \( \{ C_i \}_{i=1}^{N} \) is represented by the union of the zero-level sets of the functions \( \phi_i \). Here we also use the binary level set function of (4), and denote \( \Phi = \{ \phi_i \}_{i=1}^{N} \).

Employing the binary level set function as a differentiable threshold operator, we can now define the target regions or phases in the domain, in the following way:

\[
\begin{align*}
R_1 &= \{(x, y)| \phi_1(x, y) = 1 \} \\
R_2 &= \{(x, y)| \phi_1(x, y) = -1, \phi_2(x, y) = 1 \} \\
\vdots \\
R_i &= \{(x, y)| \phi_1(x, y) = -1, \ldots, \phi_{i-1}(x, y) = -1, \phi_i(x, y) = 1 \} \\
\vdots \\
R_N &= \{(x, y)| \phi_1(x, y) = -1, \ldots, \phi_{N-1}(x, y) = -1, \phi_N(x, y) = 1 \}
\end{align*}
\]

Based on above region partition, we define the energy function, minimizing which over the level set function is equivalent to minimizing the equation (17)

\[
E_N(\Phi, S) = \sum_{i=1}^{N} \int_{\Omega} -S^{(i)}(x) \chi_i d\mathbf{x} + \mu \sum_{i=1}^{N} \int_{\Omega} \left| \nabla \chi_i \right| d\mathbf{x} + \frac{1}{r} \sum_{i=1}^{N} \int_{\Omega} W(\phi_i) d\mathbf{x}
\]

where \( \chi_i \) is the characteristic function for each target region \( R_i \)

\[
\begin{align*}
\chi_1 &= \frac{1}{2}(1 + \phi_1) \\
\chi_2 &= \frac{1}{2}(1 - \phi_1) \frac{1}{2}(1 + \phi_2) \\
\vdots \\
\chi_i &= \frac{1}{2}(1 - \phi_1) \ldots \frac{1}{2}(1 - \phi_{i-1}) \frac{1}{2}(1 + \phi_i) \\
\vdots \\
\chi_N &= \frac{1}{2}(1 - \phi_1) \ldots \frac{1}{2}(1 - \phi_{N-1}) \frac{1}{2}(1 + \phi_N)
\end{align*}
\]

and \( \int_{\Omega} \left| \nabla \chi_i \right| d\mathbf{x} \) is the length of the region boundary. The last item is for the binary constraint of the level set functions.

In above description, we only consider the conformity of the image data within each region \( R_i \) with respect to its corresponding target model \( T^{(i)} \) (intra class). However, in fact, the \( N \) search regions (local regions surrounding the initial target regions) for the \( N \) targets may have overlaps in terms of spatio position between each others (see Fig. 6(b)), i.e. the confidence scores within these \( N \) search regions may have spital overlaps. Further, after we extend the confidence maps to the whole image domain (by using zero value to make up the rest parts), we actually get \( N \) confidence maps \( \{ S^{(i)}(x) \}_{i=1}^{N} \) for the \( N \) targets, each of which is defined over the whole image domain (see Fig.6(c),(d)). In this way, each pixel \( x \) in the image corresponds to \( N \) scores (can be regarded as \( N \) channels of the score map) indicating how likely it comes from the \( N \) targets. Therefore, one can comprehensively consider the unconformity of the image data within \( R_i \) with respect to other target models \( T^{(j)} \) \( j=1, \ldots, N, j \neq i \) (inter class), as well as the background region with respect to all target models.

Besides the \( N \) target regions, we define the rest of the image region as the background, \( R_{N+1} = \Omega \setminus (R_1 \cup R_2 \cup \ldots \cup R_N) = R_{C_1} \cap R_{C_2} \cap \ldots \cap R_{C_N} \), see Fig.4(b). Employing the level set function, we have

\[
R_{N+1} = \{(x, y)| \phi_1(x, y) = -1, \ldots, \phi_N(x, y) = -1 \}
\]

Considering the image data within the \( N+1 \) regions \( \{ R_i \}_{i=1}^{N+1} \), with respect to all \( N \) target models \( \{ T^{(i)} \}_{i=1}^{N} \), we rewrite the probability equation (17) as

\[
F([R_i]_{i=1}^{N}) = \sum_{i=1}^{N} \int_{x \in R_i} -\log p(T^{(i)}|I(x))d\mathbf{x}
\]

\[
+ \sum_{j=1, \ldots, N, j \neq i} \int_{x \in R_i} \log p(T^{(j)}|I(x))d\mathbf{x}
\]

\[
- \log p([R_i]_{i=1}^{N})
\]

The first term measures the conformity of the image data in region \( R_i \) with respect to its corresponding target model \( T^{(i)} \), while the second term measures its unconformity with respect to other target models \( T^{(j)} \) \( j=1, \ldots, N, j \neq i \); the third term measures the unconformity of the background region \( R_{N+1} \) with respect to all target models. We define the corresponding level set energy function as

\[
E_N(\Phi, S) = \int_{\Omega} -S^{(1)}(x) \chi_1 d\mathbf{x} + \sum_{j=2, \ldots, N} \int_{\Omega} S^{(j)}(x) \chi_1 d\mathbf{x}
\]

\[
+ \int_{\Omega} -S^{(2)}(x) \chi_2 d\mathbf{x} + \sum_{j=1, \ldots, N, j \neq 2} \int_{\Omega} S^{(j)}(x) \chi_2 d\mathbf{x}
\]

\[
+ \cdots + \int_{\Omega} -S^{(N)}(x) \chi_N d\mathbf{x}
\]

\[
+ \sum_{j=1, \ldots, N-1} \int_{\Omega} S^{(j)}(x) \chi_{N-1} d\mathbf{x}
\]

\[
+ \sum_{j=1, \ldots, N} \int_{\Omega} S^{(j)}(x) \chi_{N+1} d\mathbf{x}
\]

\[
+ \mu \sum_{i=1}^{N} \int_{\Omega} \left| \nabla \chi_i \right| d\mathbf{x} + \frac{1}{r} \sum_{i=1}^{N} \int_{\Omega} W(\phi_i) d\mathbf{x}
\]

where \( \chi_{N+1} \) corresponds to the characteristic function of region \( R_{N+1} \)

\[
\chi_{N+1} = \frac{1}{2}(1 - \phi_1) \ldots \frac{1}{2}(1 - \phi_N)
\]
We use \( \sum_{i} \int_{\Omega} |\nabla \phi_i| \), the sum of the length of the zero-level sets of \( \phi_i \), to simplify the length term \( \sum_{i} \int_{\Omega} |\chi_i'\| \).

Clearly, for \( N = 1 \) target and using \( N = 1 \) level set function, we obtain the 2-phase energy (12) (without the edge term). For the purpose of illustration, let us write the above energy for \( N = 2 \) targets and using \( N = 2 \) level set functions (see Fig. 5(a)):

\[
E_2(\Phi, S) = \int_{\Omega} \frac{1}{2} S^{(1)}(x)(1 + \phi_1)dx + \int_{\Omega} \frac{1}{2} S^{(2)}(x)(1 + \phi_2)dx
+ \int_{\Omega} \frac{1}{4} S^{(1)}(x)(1 - \phi_1)(1 + \phi_2)dx
+ \int_{\Omega} \frac{1}{4} S^{(1)}(x)(1 - \phi_1)(1 - \phi_2)dx
+ \int_{\Omega} \frac{1}{4} S^{(2)}(x)(1 - \phi_2)(1 - \phi_2)dx
+ \mu \int_{\Omega} |\nabla \phi_1| dx + \frac{1}{\tau} \int_{\Omega} W(\phi_1)dx
+ \mu \int_{\Omega} |\nabla \phi_2| dx + \frac{1}{\tau} \int_{\Omega} W(\phi_2)dx
(27)
\]

The Euler-Lagrange equations obtained by minimizing (27) with respect to \( \Phi = \{\phi_1, \phi_2\} \) are:

\[
\frac{\partial \phi_1}{\partial k} = S^{(1)}(x) - \frac{1}{2} S^{(2)}(x)(1 + \phi_2)
+ \mu \text{div}(\frac{\nabla \phi_1}{\nabla \phi_1}) - \frac{1}{\tau} W'(\phi_1)
(28)
\]

\[
\frac{\partial \phi_2}{\partial k} = \frac{1}{2} S^{(2)}(x)(1 - \phi_1) + \mu \text{div}(\frac{\nabla \phi_2}{\nabla \phi_2}) - \frac{1}{\tau} W'(\phi_2)
(29)
\]

We show in Fig. 5(b), the \( N = 3 \) targets case using \( N = 3 \) level set functions. Fig. 6 illustrates the tracking principle of the proposed multi-phase SLSM on a 2-target sequence.

Although our framework can accommodate additional information for more complex cases, such as considering the targets previous trajectories and intersections information, in this paper, our goal is to present the whole supervised framework and we only use the boosting based appearance information for illustration.

V. EXPERIMENTAL RESULTS

In this section, the proposed SLSM method was tested on a number of video sequences which correspond to different challenges for visual tracking. In all cases, we use five weak classifiers and employ the local histogram of oriented gradients as well as the color cues to construct the 11D feature vector as in [7]. The initial curve of a target in the first frame was a rough bounding box supplied manually while the subsequent ones were fed by the result of previous frame.

A. Comparisons With Bounding Box Trackers

Firstly, we compare the proposed contour tracking method with regular bounding box trackers on a plane sequence to show its advantage. This is a very low contrast sequence and captures a flying plane which looks quite small because of the long distance air shooting. These challenges make the tracking task difficult. We employ two prevalent trackers to give the regular tracking results: a) standard particle filter based on HSV histogram [2], with 30 particles; b) the DF tracker in [45] where a distribution field is proposed as the image descriptor. A DF is an array of probability distributions that defines the probability of a pixel of taking each feature value. As can be seen in Fig. 7(a) and Fig. 7(b), it is a challenge for the bounding box trackers to accurately locate the target, since the long and thin wings of the plane make the bounding box contain a lot of background pixels. As a result, the established target model can not provide accurate information to better distinguish between the target and its local background, which may weaken the judgement and result in deviation from the ground truth. Drift occurs when the background pollution passes down to the subsequent frames. In contrast, the proposed algorithm, integrating the boosting classifier and active contour model, extracts the accurate contours to describe the target as well as qualified samples to establish the appearance model, shown in Fig. 7(c).

B. Comparisons With Segmentation-Based Trackers

In this section, to further demonstrate the performance of our tracking approach, we evaluate it using two public sets of challenging video sequences that are commonly used in the literature, and compare it to two segmentation-based tracking methods on these datasets in terms of both the tracking accuracy and speed.

The first dataset is VOT2014\(^1\) [46] which comprises 25 sequences (an overall size of more than 10,000 frames) showing various objects in challenging backgrounds. Most of the objects undergo large shape deformations and some rather large lighting variations as well as occlusions. The sequences were annotated using rotated bounding boxes which provide highly accurate ground truth values.

The second dataset is from the visual tracker benchmark 2013\(^2\) [47]. The full benchmark contains 100 sequences from recent literatures. These test sequences are manually tagged with 9 attributes, which represent the challenging aspects in visual tracking, including illumination variation, scale variation, occlusion, fast motion and so on. Some of the sequences in this dataset overlap with above VOT2014 dataset, and for the consideration of space, here

\(^1\)http://www.votchallenge.net/vot2014/dataset.html

\(^2\)http://cvlab.hanyang.ac.kr/tracker_benchmark/datasets.html
Fig. 6. Tracking principle of the proposed multi-phase SLSM. We firstly in (a) shows the tracking result of the previous frame, which is used as the initial curves in current frame (b). The respective search region in (b) is set as a local extension of the initial target region by half of the width. (c) and (d) show the two channels of the confidence score map, correspond to target 1 and target 2 respectively, where we can see have an overlap in terms of spatial positions. Then we introduce the score map as target knowledge into the proposed SLSM to supervise the curves evolution and finally obtain the contours enclosing the tracked targets (e).

Fig. 7. Comparison results of the proposed method with bounding box trackers on plane sequence: First row: standard particle filter [2], second row: DF tracker [45], third row: the proposed SLSM method, bottom row: ground truth.

We compared our algorithm to two state-of-the-art methods, which are also based on segmentation technique for dynamic tracking. The first method is HoughTrack (HT) proposed by...
TABLE I

**Evaluation Results of the Three Compared Methods on VOT2014 Dataset: Percentage of Correctly Tracked Frames (Score > 0.5)**

<table>
<thead>
<tr>
<th>Sequence</th>
<th>HoughTrack [36]</th>
<th>PixelTrack [38]</th>
<th>SLSM</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. ball</td>
<td>15.12</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>2. basketball</td>
<td>9.10</td>
<td>41.79</td>
<td>37.43</td>
</tr>
<tr>
<td>3. bicycle</td>
<td>63.43</td>
<td>1.49</td>
<td>96.27</td>
</tr>
<tr>
<td>4. bolt</td>
<td>1.14</td>
<td>10.00</td>
<td>2.29</td>
</tr>
<tr>
<td>5. car</td>
<td>64.68</td>
<td>65.87</td>
<td>65.48</td>
</tr>
<tr>
<td>6. david</td>
<td>72.34</td>
<td>91.69</td>
<td>78.57</td>
</tr>
<tr>
<td>7. diving</td>
<td>0.46</td>
<td>35.16</td>
<td>100</td>
</tr>
<tr>
<td>8. drunk</td>
<td>3.14</td>
<td>4.13</td>
<td>3.72</td>
</tr>
<tr>
<td>9. fernando</td>
<td>2.05</td>
<td>33.56</td>
<td>16.10</td>
</tr>
<tr>
<td>10. fish1</td>
<td>1.15</td>
<td>1.61</td>
<td>6.65</td>
</tr>
<tr>
<td>11. fish2</td>
<td>5.81</td>
<td>24.19</td>
<td>18.06</td>
</tr>
<tr>
<td>12. gymnastics</td>
<td>9.66</td>
<td>72.46</td>
<td>100</td>
</tr>
<tr>
<td>13. hand1</td>
<td>100</td>
<td>17.43</td>
<td>20.76</td>
</tr>
<tr>
<td>14. hand2</td>
<td>47.57</td>
<td>19.48</td>
<td>48.69</td>
</tr>
<tr>
<td>15. jogging</td>
<td>80.78</td>
<td>2.28</td>
<td>22.15</td>
</tr>
<tr>
<td>16. motocross</td>
<td>100</td>
<td>6.71</td>
<td>18.29</td>
</tr>
<tr>
<td>17. polarbear</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>18. skating</td>
<td>85.50</td>
<td>9.25</td>
<td>53.75</td>
</tr>
<tr>
<td>19. sphere</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>20. sunshade</td>
<td>100</td>
<td>10.06</td>
<td>68.60</td>
</tr>
<tr>
<td>21. surfing</td>
<td>100</td>
<td>98.57</td>
<td>100</td>
</tr>
<tr>
<td>22. torus</td>
<td>100</td>
<td>80.46</td>
<td>100</td>
</tr>
<tr>
<td>23. trellis</td>
<td>72.93</td>
<td>87.70</td>
<td>39.72</td>
</tr>
<tr>
<td>24. tunnel</td>
<td>59.67</td>
<td>1.78</td>
<td>25.03</td>
</tr>
<tr>
<td>25. woman</td>
<td>18.43</td>
<td>17.59</td>
<td>88.78</td>
</tr>
<tr>
<td>average</td>
<td>51.7184</td>
<td>41.3004</td>
<td>56.4132</td>
</tr>
</tbody>
</table>

TABLE II

**Evaluation Results of the Three Methods on 2013 Benchmark Dataset: Percentage of Correctly Tracked Frames (Score > 0.5)**

<table>
<thead>
<tr>
<th>Sequence</th>
<th>HoughTrack [36]</th>
<th>PixelTrack [38]</th>
<th>SLSM</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Bikerc</td>
<td>44.37</td>
<td>47.18</td>
<td>45.07</td>
</tr>
<tr>
<td>2. Bird2</td>
<td>95.96</td>
<td>27.27</td>
<td>90.91</td>
</tr>
<tr>
<td>3. BlurCar2</td>
<td>100</td>
<td>5.98</td>
<td>100</td>
</tr>
<tr>
<td>4. Car1</td>
<td>18.63</td>
<td>2.75</td>
<td>20.20</td>
</tr>
<tr>
<td>5. CarDark</td>
<td>27.99</td>
<td>0.25</td>
<td>2.04</td>
</tr>
<tr>
<td>6. CliBar</td>
<td>29.87</td>
<td>41.74</td>
<td>44.28</td>
</tr>
<tr>
<td>7. Coke</td>
<td>61.86</td>
<td>76.98</td>
<td>90.38</td>
</tr>
<tr>
<td>8. Dancer</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>9. David3</td>
<td>33.33</td>
<td>98.02</td>
<td>100</td>
</tr>
<tr>
<td>10. Deer</td>
<td>100</td>
<td>100</td>
<td>46.48</td>
</tr>
<tr>
<td>11. Dog</td>
<td>77.95</td>
<td>39.37</td>
<td>85.83</td>
</tr>
<tr>
<td>12. Dog1</td>
<td>100</td>
<td>51.78</td>
<td>100</td>
</tr>
<tr>
<td>13. DragonBaby</td>
<td>63.72</td>
<td>82.30</td>
<td>38.05</td>
</tr>
<tr>
<td>14. Human4</td>
<td>10.04</td>
<td>2.69</td>
<td>18.59</td>
</tr>
<tr>
<td>15. Ironman</td>
<td>16.87</td>
<td>2.41</td>
<td>2.41</td>
</tr>
<tr>
<td>16. Jumping</td>
<td>92.33</td>
<td>79.55</td>
<td>94.57</td>
</tr>
<tr>
<td>17. Matrix</td>
<td>31.00</td>
<td>7.00</td>
<td>5.00</td>
</tr>
<tr>
<td>18. MountainBike</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>19. Panda</td>
<td>9.20</td>
<td>9.50</td>
<td>100</td>
</tr>
<tr>
<td>20. RedTeam</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>21. Skiing</td>
<td>41.98</td>
<td>100</td>
<td>14.81</td>
</tr>
<tr>
<td>22. Surfer</td>
<td>95.48</td>
<td>58.78</td>
<td>78.99</td>
</tr>
<tr>
<td>23. Tiger2</td>
<td>90.41</td>
<td>25.75</td>
<td>95.89</td>
</tr>
<tr>
<td>24. Walking</td>
<td>100</td>
<td>23.06</td>
<td>100</td>
</tr>
<tr>
<td>25. Walking2</td>
<td>37.40</td>
<td>11</td>
<td>37.40</td>
</tr>
<tr>
<td>average</td>
<td>63.1356</td>
<td>47.7344</td>
<td>64.4360</td>
</tr>
</tbody>
</table>

Godec et al. [36], where the authors proposed a patch-based voting algorithm with Hough forests [48]. By back-projecting the patches that voted for the object centre, the authors initialise a graph-cut algorithm to segment foreground from background. The second method is PixelTrack proposed by Duffner and Garcia [38], where the authors combined two components, a detector that makes use of the generalised Hough transform with pixel-based descriptors and a probabilistic segmentation method based on global models for foreground and background, in a co-training manner to track objects that undergo rigid and non-rigid deformations. We use the original configuration and standard settings provided by the implementation of these two methods. The output of the HoughTrack method in each frame is an estimated object contour as well as an object region surrounded by the contour. The PixelTrack selects to output a cross at the center position of the target or a bounding box, and we choose the bounding box output as it is reported in its original literature [38].

For quantitative analysis, for each video, we determine the percentage of frames in which the object is correctly tracked. Since the ground truth annotation included in the datasets is represented by a simple bounding box, and to let the contour trackers be compared fairly with other bounding box trackers, we measure the tracking accuracy using the Agarwal-criterion [49] as in [36]. It is defined as

\[
\text{score} = \frac{R_T \cap R_{GT}}{R_T},
\]

where \( R_T \) is the output target region from the tracking algorithm and \( R_{GT} \) the ground truth. In each video frame, the tracking is considered correct if the Agarwal overlap measure is above a threshold (we set to 0.5).

Table I summarises the evaluation results on the VOT2014 dataset. As we can see, the proposed method performs better than the other two segmentation based methods on many sequences of the dataset and also on average. Fig. 8 shows some tracking results from the ball sequence. This sequence describes a red/white ball rolling between two people. HoughTrack loses the target after frame 99 and stops to detect the object and removes the track after frame 183. Though the PixelTrack could follow the target throughout the sequence, it can not provide accurate target region since the segmentation incorporated in the algorithm is quite rough and not capable to provide a contour tracking result. In contrast, the proposed algorithm, using the boosting scores to supervise curve evolution of the active contour model, can present accurate contours to describe the target as well as extract qualified samples to establish the appearance model.

Table II illustrates the results of the three compared methods on the 2013 Benchmark dataset. For 16 out of 25 video sequences the proposed method outperforms the other algorithms, and also the average of correct tracking. Fig. 9 presents some tracking results of the three compared methods on Panda sequence. This sequence records a panda moving around in a yard. Both the HoughTrack and the PixelTrack lose the target after frame 107 when the panda passed a small tree and occlusion occurred. Unlike the explicit target modeling methods, the classifier based implicit representation has the ability of accommodating information of target in the past, thus the proposed method can deal with these occlusion cases. We use decrease rate of the detected positive samples as an occlusion detector. Once detected, we slow down the speed of updating the weak classifiers as in [7], allowing tracking to resume when the target reappears.
see [19] for more details and results involving partial and total occlusions. Fig. 10 shows some more tracking results of the proposed method on the two datasets.

We measured the average processing speed of each algorithm for all the 25 videos of the VOT2014 dataset on a double 2.6 GHz Intel Core i7 processor. The codes of HoughTrack and PixelTrack are implemented in C++ using OpenCV library while ours uses Matlab. We use 15 iterations for each image to evolve the level set curve. In terms of speed, see Table III, the proposed method performs faster than the HoughTrack but slower than the PixelTrack. The reason why the proposed method and the HoughTrack are relatively slow is their contour extraction procedure, which is computationally demanding. On the other hand, however, these

---

**TABLE III**

<table>
<thead>
<tr>
<th></th>
<th>HT [36]</th>
<th>PixelTrack [38]</th>
<th>SLSM</th>
</tr>
</thead>
<tbody>
<tr>
<td>AVERAGE OVERALL PROCESSING SPEED IN SECONDS PER FRAME</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MEASURED WITHOUT SCREEN DISPLAY</td>
<td>0.4926</td>
<td>0.0104</td>
<td>0.2632</td>
</tr>
</tbody>
</table>

---

Fig. 8. Tracking results of the compared methods on *ball* sequence from the VOT2014 dataset: Top row: HoughTrack [36], middle row: PixelTrack [38], bottom row: the proposed SLSM method.

Fig. 9. Tracking results of the compared methods on *Panda* sequence from the 2013 Benchmark dataset: Top row: HoughTrack [36], middle row: PixelTrack [38], bottom row: the proposed SLSM method.
two methods achieve better segmentation of the tracked object and therefore get a more accurate contour tracking result while the PixelTrack aims at fast tracking of the position of an object with a rough segmentation as an auxiliary. We adopted the binary level set method, with advantages in speed and simplicity for the reinitialization procedure as well as
numerical computations. Besides, the binary level set segmentation is not applied to the whole image but only to the local search region, which also greatly reduces the computation. We expect, still, that we can reduce this runtime with optimized programming and a better convergence criterion, since convergence seems to have been achieved before the 10th iteration.

C. Multi-Target Tracking

So far, all above tests are implemented on the uni-target sequences using 2-phase SLSM method. Here, we show the ability of the proposed multi-phase SLSM method to work on multi-target sequences. The first test sequence contains two unicolored targets: a green bowl and a purple glassbox, which are held to move, rotate and interact with each other. The second test is implemented on a complex scenario from [50] with two multi-colored targets: two children are running quickly in a circular path in a cluttered room, undergoing significant shape deformation. The third test is conducted on a 3-target sequence: a toy bear, a cup and a glassbox are held and swayed round each other. Fig. 11 shows the tracking results of the proposed SLSM method on these three sequences, indicating its robustness in dealing with these multi-target contour tracking cases.

VI. CONCLUSION

We have presented a novel supervised level set method (named SLSM) in this paper for non-rigid objects contour tracking. By considering the context of tracking, we refined the curve evolution of the SLSM by the specific knowledge of the targets we want to track, which is learned in an online boosting manner. Hence, in contrast with conventional intensity consistency based level set methods, our approach is object-oriented and can lead a more accurate convergence to the exact targets in tracking applications. We firstly proposed the mechanism of 2-phase SLSM for single target tracking, then proposed the generalized multi-phase SLSM for dealing with multi-target cases. Experimental results on a number of challenging video sequences have verified that the proposed method is effective in many complicate scenes.

REFERENCES


Xin Sun received the B.S. and M.S. degrees from the School of Computer Science and Technology, Harbin Institute of Technology, China, in 2008 and 2010, respectively, and the Ph.D. degree from the Harbin Institute of Technology in 2015. Her research interests include object tracking, image segmentation, and machine learning.

Hongxun Yao received the B.S. and M.S. degrees in computer science from the Harbin Institute of Technology, in 1987 and 1990, respectively, and the Ph.D. degree in computer science from the Harbin Institute of Technology, in 2003. She is currently a Professor with the School of Computer Science and Technology, Harbin Institute of Technology. Her research interests include computer vision, pattern recognition, multimedia computing, and human-computer interaction technology. She has five books and over 200 scientific papers published, and won both the honor title of the New Century Excellent Talent in China and enjoy special government allowances expert in Heilongjiang Province, China.

Hongxun Yao

Shengping Zhang received the Ph.D. degree in computer science from the Harbin Institute of Technology, in 2013. He is currently an Associate Professor with the School of Computer Science and Technology, Harbin Institute of Technology at Weihai, China. He has been a Post-Doctoral Research Associate at Brown University, and a Visiting Student Researcher at UC Berkeley. He has authored or co-authored over 30 research publications in refereed journals and conferences. His research interests focus on moving object detection, tracking, and action recognition. He is an Associate Editor of *Signal Image and Video Processing* journal and was the leading Guest Editor of special issues in *Signal Image and Video Processing* and *IET Computer Vision* journals.

Hongxun Yao

Shengping Zhang

Dong Li received the B.S. and M.S. degrees from the School of Computer Science and Technology, Harbin Institute of Technology, in 2008 and 2010, respectively, and the Ph.D. degree from the Harbin Institute of Technology, in 2015. His research interest focuses on machine learning algorithms and their applications in information diffusion, recommendation system, and computer vision.

Dong Li

Xin Sun

Hongxun Yao

Shengping Zhang

Dong Li

Xin Sun

Hongxun Yao

Shengping Zhang

Dong Li