CONTOUR TRACKING VIA ON-LINE DISCRIMINATIVE APPEARANCE MODELING BASED LEVEL SETS

Xin Sun, Hongxun Yao, Shengping Zhang

Harbin Institute of Technology, 92 West Dazhi Street, Harbin 150001, China
{sunxin, H.Yao, S.Zhang}@hit.edu.cn

ABSTRACT
A novel level set method based on on-line discriminative appearance modeling (DAMLMS) is presented for contour tracking. In contrast with traditional level set models which emphasize the intensity consistent segmentation and consider no priors, the proposed DAMLMS takes the context of tracking into account and use a discriminative patch based target model to guide the curve evolution. By modeling both the region and edge cues in a Bayesian manner, the proposed level set method can lead an accurate convergence to the candidate region with maximum likelihood of being the target. Finally, we update the target model to adapt to the appearance variation, enabling tracking to continue under occlusion. Experiments confirm the robustness and reliability of our method.

Index Terms— Contour tracking, level set, appearance modeling, curve evolution

1. INTRODUCTION
Tracking refers to the task of generating the trajectories of moving objects in a sequence of images. Most of existing algorithms use rectangle or oval to represent the tracking results while objects, however, in practice may have complex shapes that cannot be well described by simple geometric shapes.

Some attempts in literature have been made to use silhouette or contour, segmenting technique for dynamic tracking [1, 2, 3]. Level set technique [4, 5], as an implicit representation of contours, is widely used. The contour is represented as the zero level set of the graph of a higher dimensional function and deformed until it minimizes an image-based energy function. Binary level set model [6], due to its improved computational efficiency via using a two-valued level set function, is more suitable for tracking applications. However, from a performance perspective, the binary level set model is more inclined to segment out the region with consistent intensity, which is similar to the thresholding segmentation method.

Recently, many researchers apply level set models to visual tracking [7, 8, 9]. However, few refine them by the prior target knowledge and deal with the problem of multi-mode target segmentation. Bibby et al. [7] derive a probabilistic framework for robust tracking of multiple previously unseen objects where the shapes are implicit contours represented using level sets. In [9], the authors add Mumford-Shah model into the particle filter framework. Once the particle filter gives the candidate positions, the level set curve evolution is included, with no prior target knowledge, to give the candidate contours.

In this paper, we present a novel discriminative appearance modeling based level set method (DAMLMS) for non-rigid object tracking. Instead of acting towards intensity consistency direction, the curve evolution of the proposed DAMLMS is target-oriented and supervised by the on-line constructed target model. By considering both the region and edge cues, the proposed DAMLMS can achieve multi-mode target segmentation, and the curve finally converges to the candidate region with maximum likelihood of being target.

The rest of this paper is organized as follows: We analysis the general curve acting principle of the level set model and gives the antecedent of our improvement in Section 2. In Section 3, the proposed tracking algorithm is described in detail. Experiment results on challenging video sequences are shown in Section 4, followed by conclusion in Section 5.

2. CURVE ACTING PRINCIPLE
In binary level set model [6], a piecewise constant-valued function \( u \) is used to approximate the intensity distribution of image \( I \). The image is divided into two regions \( \Omega_1 \) and \( \Omega_2 \). In region \( \Omega_1 \), the level set function \( \phi = 1 \) and \( u = c_1 \) while in region \( \Omega_2 \), \( \phi = -1 \) and \( u = c_2 \). That is \( u = \frac{\phi}{2} + \frac{1}{2} \).

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

\[
E_{\text{image}}(c_1, c_2, \phi) = \frac{1}{2} \int_{\Omega_1} |u(c_1, c_2, \phi) - I|^2 \, dx \, dy + \mu \int_{\Omega_1} |\nabla \phi| \, dx \, dy + \frac{1}{2} \tau \int_{\Omega_1} W(\phi) \, dx \, dy
\]

where the first item is used to measure the similarity of the two-valued function \( u \) with the image \( I \); the second item refers to the length of the curve, 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 $c_1$, $c_2$ are obtained directly by minimizing the energy function:

$$c_1 = \frac{\int_{\Omega_1} I(1 + \phi) \, dx \, dy}{\int_{\Omega_1}(1 + \phi) \, dx \, dy}, \quad c_2 = \frac{\int_{\Omega_2} I(1 - \phi) \, dx \, dy}{\int_{\Omega_2}(1 - \phi) \, dx \, dy}$$

(2)

where, we can see, $c_1$ and $c_2$ are the average intensity 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, which is similar to the thresholding segmentation method. However, the object may consist of inconsistent intensity which occurs most often in practice. Additionally, in the context of tracking, we usually have a specific target of interest, which can be explored to supervise the evolution of the curve and refine its acting orientation.

3. THE PROPOSED LEVEL SET METHOD

3.1. Discriminative target appearance modeling

Given a target region learned from previous views, depending on the assumption that the most informative object region for tracking are the same region that best discriminate between object and background, we divide its enclosing rectangle into a number of patches from which we select the most discriminative one as the tracking basis. A larger ring of neighboring pixels is chosen to represent the background. Let $X_0$ denote the center location of the object and $Y_0$ the relative position between $Y_0$ and $X_0$.

We use the augmented variance ratio (AVR), the ratio of the between class variance to the within class variance, to measure the discriminative power of patches as in [10]. For each patch and the background, by normalizing their histograms, we can get a discrete probability density $p(j)$ for the patch, and density $q(j)$ for the background, where index $j$ ranges from 1 to $b$, the number of histogram buckets.

The log likelihood of an image value $j$ can be given by

$$L(j) = \log \frac{\max \{p(j), \delta\}}{\max \{q(j), \delta\}}$$

(3)

where $\delta$ is a small value (set to 0.001) that prevents dividing by zero. It is obvious that the log likelihood maps the region into positive values for colors distinctive to the object, and negative for colors associated with the background. Colors shared by both object and background tend towards zero.

Then the variance ratio of $L(j)$ can be computed to quantify the separability of the patch and background classes:

$$\text{VR}(L; p, q) = \frac{\text{var}(L; p + q) / 2}{\text{var}(L; p) + \text{var}(L; q)}$$

(4)

where

$$\text{var}(L; a) = \sum_j a(j) L^2(j) - \left(\sum_j a(j) L(j)\right)^2$$

(5)

defines the variance of $L(j)$ with respect to a discrete probability density function $a(j)$.

Since we would like log likelihood values of pixels on the object and background to both be tightly clustered while the two clusters should ideally be spread apart as much as possible, the denominator of the variance ratio enforces that the two clusters should ideally be spread apart as much as possible, the numerator rewards cases where values associated with object and background are widely separated.

After we got the most discriminative patch $Y_0$ with the largest variance ratio and its corresponding $R_0$, a target appearance model can be constructed base on it as follow

$$T_0 = (X_0, Y_0, R_0)$$

(6)

Fig.1 shows an example of the model on riding sequence.

3.2. The region based likelihood

Let $C(s) = [x(s) \ y(s)]^T$, $s \in [0, 1]$, denote a closed curve in $\mathbb{R}^2$. Within a new arriving frame, for each candidate curve and its respective location $X_C$, we can get the candidate model $T_C = (X_C, Y_C, R_C)$ where $Y_C$ is the corresponding patch to $Y_0$ in $T_0$, determined by employing particle filter procedure based on $Y_0$. Let $Y_C$ denote the patch in the new frame that has the same relative position, $R_0$, as $Y_0$. Then the region based likelihood of curve $C$ can be computed by measuring the similarity between the candidate region and target model:

$$R(C, T_0) = \exp(-\lambda DIS(Y_C, Y_0))$$

(7)

where $DIS$ function returns the distance between the two patches, $\lambda$ denotes the weighting parameter that is set to 25.

3.3. Level set formulation

For computational efficiency considerations, the proposed DAMLSM maintains the advantage of using two-valued level set function:

$$\phi(x, y, k) = \begin{cases} 1, & \text{if } [x \ y]^T \text{ inside } C_k \\ -1, & \text{if } [x \ y]^T \text{ outside } C_k \end{cases}$$

(8)

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.

Let $I_k : \mathbb{R}^2 \rightarrow \mathbb{R}^m$ denote the image at time $k$ that maps a pixel $x = [x \ y]^T \in \mathbb{R}^2$ to a value. Given all the observations $I_{0:k}$ up to time $k$, target model $T_0$, and the previous contours $C_{0:k-1}$, we model the probability of contour $C_k$ by considering both the region and edge cues in a Bayesian manner as

$$p(C_k | I_{0:k}, T_0, C_{0:k-1}) \propto p(T_0 | C_k) p(I_k | C_k) p(C_k | C_{0:k-1})$$

(9)
where \( p_T(T_0|C_k) \) presents the likelihood of the region inside \( C_k \) being the target object, and \( p_c(I_k|C_k) \) the likelihood that the contour is on image edge, \( p(C_k)|C_{0:k-1} \) the prior probability which we regard equally for all candidate curves. Here, the assumption we based on is that the measurements are independent of each other.

When we maximize the probability of (9), obviously, we expect to obtain the contour that surrounds the target region and just right converges to its edge.

The region based probability \( p_T(T_0|C_k) \) has been presented as

\[
p_T(T_0|C_k) \propto R(C_k, T_0)
\]

Under the objective of driving the contour to the target boundary, we use image gradient for edge detecting, see Fig.1, and the edge based probability \( p_c(I_k|C_k) \) can be computed as

\[
p_c(I_k|C_k) \propto \sum_{[x \ y]^T \in C_k} T(x, y)
\]

where

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

where \( \nabla \) denotes spatial gradient operator, * convolution and \( G_\sigma \) 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 (9), as follow:

\[
E(\phi, T_0) = -R(C, T_0) + \xi \int_C -T(x) dx + \mu |\nabla(C)| + \frac{1}{\tau} \int_\Omega W(\phi) dx
\]

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 term is for constraint of \( \phi^2 = 1 \), where \( W \) can be defined as \( (\phi^2 - 1)^2 \) and \( \Omega = \Omega_1 \cup \Omega_2 \) is the image domain.

Employing the binary level set function, we rewrite (13) as

\[
E(\phi, T_0) = -R(\phi, T_0) + \int_\Omega -\xi T(x)(1 - \phi^2) + \mu |\nabla(\phi)| + \frac{1}{\tau} W(\phi) dx
\]

where \( \ell(C) = \int_\Omega |\nabla(\phi)| dx \). The associated Euler-Lagrange equation for this function can be given by

\[
0 = -R'_\phi(\phi, T_0) + 2\xi T(x) - \mu div\left(\frac{\nabla(\phi)}{|\nabla(\phi)|}\right) + \frac{1}{\tau} W'(\phi)
\]

and implemented by the following gradient descent:

\[
\frac{\partial \phi}{\partial t} = R'_\phi(\phi, T_0) - 2\xi T(x) + \mu div\left(\frac{\nabla(\phi)}{|\nabla(\phi)|}\right) - \frac{1}{\tau} W'(\phi)
\]

where \( \text{div} \) is the divergence operator.

In contrast with conventional level set formulations, ours, instead of based upon intensity consistence, is supervised by the specific knowledge of the target. Therefore, the curve, in

\[\text{Fig. 1. Illustration of the proposed method. (a) shows the initial curve obtained from previous frame and the corresponding discriminative patch based target model. Then we use the target knowledge, conjunction with the edge cue shown on the left side of (b), to supervise the curve evolution and finally obtain the contour convergent to the target in frame #33.}\]

DAMLSTM, 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.1 illustrates the whole proposed algorithm.

4. EXPERIMENTAL RESULTS

In this section, the proposed method was tested on several challenging video sequences. The initial curve of the first frame was a rough polygon supplied manually while the subsequent ones were fed by the results of previous frame. We use HSV color space and 12 × 12 patch size.

The first sequence consists of 820 frames and describes a ship navigating on the river with moving waves behind and illumination changes. We can see that the performance of our method is good as shown in Fig.2.

In above test, many contour tracking approaches can give good results as the proposed method and the same phenomenon can be observed on other unicolor target sequences. Furthermore, in order to show the improvement of our approach, we compared the proposed method on two multi-mode target sequences with conventional level sets based method in [9], which adds Mumford-Shah model within the particle filter framework for contour tracking without considering any target information.

The first sequence describes a man in strip colorful clothes walking on the balcony, undergoing significant scale changes and shape deformation as he walked toward or deviating from the camera. As we can see in Fig.3, it is a challenge for traditional intensity consistence based level set method to represent the person accurately. In contrast, the proposed method shows pleasant results, demonstrating the effectiveness of the technical. The second test is on a gray scale sequence, which describes a toy dog held and swayed under a lamp. As shown in Fig.4, our method can perform well even with large appearance changes in gray scale images.

Next, we use another three sequences with different challenges to further evaluate the proposed method. The first se-
Fig. 2. Tracking results of the proposed DAMLSM method on boat sequence.

Fig. 3. Tracking results on man sequence for frames of #18, #46, #54, #70, #92.

The second sequence describes a toy lemming moving fast above the table with a clutter background behind as well as sheltering cases. The third sequence records a diving process, where dramatic appearance changes and occlusion occurs. Fig. 5 shows the tracking results of these three sequences, indicating the robustness of the proposed method in dealing with these challenging cases and its availability under severe occlusion, profited by on-line updating of the target model.

5. CONCLUSION

A novel level set method has been presented in this paper for non-rigid object tracking. In contrast with conventional intensity consistency based level set models, our approach is object-oriented and the curve evolution is refined by an on-line constructed discriminative model of target. With the consideration of both region and edge cues, the proposed method can lead an accurate convergence to the targets in tracking applications. Experimental results on several challenging video sequences have validated the effectiveness of the technique.

6. ACKNOWLEDGMENTS

This work is supported by the National Natural Science Foundation of China (No. 61071180) and NEC cooperative project (No. LC04-20101201-03).

7. REFERENCES


