Shape Prior Based Foreground Segmentation with Local Rotation and Structural Changes

Xiaojun Wu, Jia Chen, Michael Yu Wang, Xiongwei Wang

Abstract—Shape prior based segmentation method has the advantage of robustly tracking moving object. But, it can not deal with local rotations or changes of object structure. In this paper, we propose an algorithm that can accurately segment the moving foreground object with local rotation and object structure change from the low resolution gray scale images based on Schoenemann’s method. Two new terms, turning angle and a distance function, are added to the energy function. When estimating the similarity between the prior shape template and the object contour, turning angle is used to measure the curve match, which is rotational invariance and translational invariance. To deal with the object structure change, a distance function is added to the energy function. The experimental results show that the improved new energy function can robustly segment and track the foreground object with local rotations and structure change.

I. INTRODUCTION

Foreground segmentation refers to extract object of interest from an image or image sequences, which has extensive applications in video surveillance, human-machine interaction, virtual reality, motion capture, etc. In the past two decades, a bunch of methods were proposed to extract the object of interest. Among them, background subtraction is a fundamental and most important class of techniques. The basic idea of the approach is detecting active objects from the difference between the current frame and a reference image. All of these methods try to effectively estimate the background model from the temporal sequence of the frames.

Wren et.al propose a running Gaussian average method to model the background independently at each pixel location [1]. The model is based on ideal fitting a Gaussian probability density function on the last $n$ pixels. In order to improve the robustness of conventional background subtraction, Stauffer et.al suggest a method called mixture of Gaussian (MoG) to model complex background [2]. This model has a deadlock. For a low learning rate, the model adapts to the changes of the scene slowly so that many background pixels may be classified as foreground object. For a high learning rate, slowly moving foreground may be absorbed into the background model. Methods employing MoG have been widely incorporated into algorithms that utilize Bayesian Framework [3], color and gradient information [4], and region based information [5]. Elgammal et al in[11] propose to model the background distribution by a non-parametric model based on Kernel Density Estimation(KDE) on the buffer of the last $n$ background values. The codebook (CB) model is a compact and compressed model and can handle non-static background for a long period of duration [6], [7]. Moving objects are allowed in the training period. Codebook based model can work well for color images but has many false detections with low contrast of gray images.

Second category of foreground extraction methods are based on variational deformable models (VDM), such as level set [8], [9] or active contour [10], [11]. The core idea of VDM is to evolve a curve, subject to constraints from a given image. For instance, starting with a curve around the object to be detected, the curve moves toward its interior normal and has to stop on the boundary of the object. The internal forces serve to impose a piecewise smoothness constraint. The image forces push the snake toward salient image features like edges, and subjective contours. The external constraint forces are responsible for putting the active contour near the desired minimum. Grenander [12] and Cootes [13] propose a shape prior based segmentation method through training a large number of shape samples. Schoenemann et al. utilize a closed parameterized curve to represent the contour to achieve the target boundary from a prior template shape through minimizing a energy function [14], [15]. The similarity of target curve and prior curve is measured with elastic energy of tangential angle through finding the cycles of minimal ratio.

Though Schoenemann’s method can segment foreground object robustly and completely in very cluttered background, this approach is only a translational and globally rotational invariant, which can not deal with the problem of local rotation or partial structure bifurcation of a foreground object. That is their method only can track a rigid object, rather than a flexible body, see the tracking failure in Fig. 1. In Fig. 1, after the actor’s right leg rotation or gesture change, the actors can not be segmented and tracked. So the application of this method is strictly limited in a narrow scope. In this paper, we propose a new method based on Schoenemann’s method to overcome the above limitations. Firstly we use a turning angle based energy function to replace the old one. Using this rotational and translational invariant to measure the similarity between the prior shape template and the target contour, the energy function is also rotational and translational invariant. When the shape of the object is changed, such as bifurcations or branches
appear on the object, the discrepancy between the template shape and object contour becomes large, the tangential angle based energy function can not follow the shape change of the foreground object. In our algorithm, we amend a new distance term in the energy function to track the shape deformation, which can be used to solve the problem of partial structure change of a moving object.

II. ENERGY FUNCTION

The algorithm of single shape prior based foreground segmentation method can be described as: Given a video sequence and a prior shape template of object of interest in the first frame, an object with closest boundary with the prior shape template will be tracked in the following frames. For each current frame, the boundary of a segmented object in the previous frame looks as the prior shape template, demonstrated in Fig.2.

Fig. 2: Prior shape based object segmentation, (a)Prior shape template, (b)Segmentation of 3D rotation and occlusion, (c)Segmentation with rotational invariance [14]

Given a prior shape curve $S$, the contour of the tea port in Fig.2(a), and the gray level image $I$. The tracking task is finding the closest boundary $C$ with contour curve $S$ in image $I$ according to some predefined criteria. Both curves are parameterized with arc length, denoted as

$$\begin{align*}
S : [0, l(S)] &\rightarrow R^2 \\
C : [0, l(C)] &\rightarrow R^2
\end{align*}$$

(1)

where $l(\cdot)$ represents the arc length.

A match function $m : [0, l(C)] \rightarrow [0, l(S)]$ represents the correspondence between the points on curve $C(s)$ and $S(m(s))$. Then we can use similarity between the corresponding points on both curve to measure the similarity of the two curves [16]. In order to find the closed shape in current image $I$, the energy function includes the similarity measurement function of prior shape template $S$ and target contour $C$. Meanwhile, if the algorithm converges to the target contour, a gradient detection term should be added in the energy function. Schoenemann et al employ the following energy function [14], [15].

$$\begin{align*}
\min_{C, m} \int_0^{l(C)} g(C(s))ds &+ \frac{\int_0^{l(C)} \Theta(C(s), S(m(s)))ds}{l(C)} + \\
\int_0^{l(C)} \Psi(m'(s)) + \nu |\alpha_C(s) - \alpha_S(m(s))|^2ds &\quad (2)
\end{align*}$$

where the denominator is the length of curve $C$, numerator part includes the edge detection function $g(x)$, similarity measurement $E_{shape}$ and searching range function and tangential angle $\alpha$. In Schoenemann energy function, the similarity measurement function roots from Basri’s cost function[17].

$$E_{shape} = \int_0^{l(C)} (k_C(s) - m'(s)k_S(s))ds + \lambda \int_0^{l(C)} \Psi(m'(s))ds$$

$E_{shape}$ computes the deformation cost from curve $S$ to $C$ in case of object structure unchanged. However, due to curve $C$ unknown, it is difficult to directly compute the integral of curvature difference. Schoenemann utilizes tangential angle to replace the curvature. The tangential angle is translational invariant, rather than rotational invariant. So we should find new cost function to represent rotational invariant and object flexible change.

III. ROTATIONAL INVARIANCE

Curve match is extensively studied in computational geometry, such as Hausdorff distance [18], Fréchet distance [19], and turning angle[20]. The first two methods can calculate the similarity of two known curves. Only turning angle can be used in local feature match, which is rotational invariant and can be used in our foreground segmentation algorithm. The turning angle $\beta(x, y, z)$, is included angle to two neighbor line segments, $(x, y)$ and $(y, z)$, see Fig. e.g.

$$\beta(x, y, z) = \alpha(y, z) - \alpha(x, y)$$

(4)
Accordingly, the similarity measurement function becomes

$$E_{\text{shape}} = \int_0^{l(C)} |\beta_C(s) - \beta_S(m(s))|^2 ds + \int_0^{l(C)} \Psi(m'(s)) ds$$

where $\beta_C(s)$ expresses the turning angle formed by two neighborhood points of $C(s)$. After introducing the turning angle function, the total energy function becomes

$$E_{\text{newshape}} = \frac{\int_0^{l(C)} g(C(s)) ds}{l(C)} + \lambda \frac{\int_0^{l(C)} \Psi(m'(s)) ds}{l(C)} + \frac{\int_0^{l(C)} \Theta(C(s), S(m(s))) ds}{l(C)} + \mu \frac{\int_0^{l(C)} (\beta_C(s) - \beta_S(m(s)))^2 ds}{l(C)}$$

Fig. 3: Definition of turning angle, $x, y, z$ are three adjacent pixels.

When the turning angle is utilized, we will take the turning angle as curve match unit, instead of single pixel. So the structure of product graph and edge weight should be adjusted. The node of product graph becomes $((x, y), t)$. The first class of edge in graph is defined as

$$((x, y), i \cdot K + k) \rightarrow ((y, z), j \cdot K) \quad i < j < i + K$$

The second class has the definition as

$$((x, y), i \cdot K + k) \rightarrow ((y, z), j \cdot K + k + 1) \quad k + 1 < K$$

To test the validity of this new energy function, we test two video sequences. Images in Fig. 4 come from HumanEva database [21]. We manually extract the silhouette of the first frame as the prior shape template, then over sixty frames are tracked. The new energy function can well track the actor’s right leg rotation, see the results in Fig. 4. Fig. 5 shows images captured in our lab. Initially, the first frame prior shape includes the actor’s left hand. With the movement of the actor, the left hand is gradually invisible in the 7th frame, which means this edge map does not include any foreground information. Finally, we compute the edge map of adjacent frame difference, $DE_i$. This edge map contains not only

![Fig. 4: When the local rotation of the object contour occurs such as the actor’s leg rotation, the improved method can obtain good results.](image1)

![Fig. 5: The new energy $E_{\text{newshape}}$ cannot handle the changes of object structure.](image2)

But (a) and (c) do not have the same object structure, because the two circles of (c) tangential contact to each other, while the two circles in (a) not. The curve (d) has only two circles, it does not has the same component with (a). So only the curve (a) and (b) have the same object structure.

When the object structure changes, the target contour will have large deformation. We can not obtain a minimized $E_{\text{shape}}$ or $E_{\text{newshape}}$. In order to drive the prior shape to the deformed object boundary, we should increase the weight of target contour in energy function.

A. Moving Edge Detection

Moving edge refers to the changed motion edge caused by the foreground movement of two adjacent frames. Moving edge detection extracts the changed edges of a foreground object, which is used in moving objects extraction and tracking from a video. Assume $E_i = \Phi(I_i)$ is an edge map of current frame, where $\Phi$ indicates edge detection operators, such as Canny, Sobel etc. $E_b = \Phi(I_b)$ is the background edge map, and $DE_i = \Phi(|I_i - I_{i-1}|)$ is the edges of frame differences, where $E_b$ is the absolute background edge map, which means this edge map does not include any foreground information. Firstly, we compute the edge map of adjacent frame difference, $DE_i$. This edge map contains not only
the motion edges for current frame, but also the background edges due to foreground movement. The edge map of current frame is \(E_i = e_1, e_2, \ldots\), where \(e\) is the edge pixel. As the edges in the previous frame contained in \(DE_i\) do not included in current frame edge map \(E_i\). But the moving edge of foreground object is in \(E_i\). So we can use the difference of \(E_i\) and \(DE_i\) to compress the edge of previous frame. The pixels of moving edge are the pixels whose distance with \(DE_i\) is less than a predefined threshold \(T_{change}\), and these pixels are also not on background edges, e.g.

\[
ME_i = \{ e \in E_i \mid \|e - x\| < T_{change}, e \notin E_b, x \in DE_i \}
\] (7)

Next, we should extract the static edge map of adjacent frames, \(SE_i\). The target object edge of \((i - 1)th\) frame, \(OE_{i-1}\), is the pixel collection of shape prior template \(S\), while the edge map of \(ith\) frame is contained in \(E_i\). We can use the following formulation to calculate \(SE_i\),

\[
SE_i = \{ e \in E_n \mid \|e - x\| < T_{still}, x \in OE_{i-1} \}
\] (8)

where \(T_{still}\) is a constant, often set to 1 or 2. Generally, the summation of \(ME_i\) and \(SE_i\) is the target edge map. But, a portion of edge may be lost because of noise or nonuniform illumination. For a video sequence, the object edge map of two adjacent frame \(OE_{i-1}\) and \(OE_i\) is within a small threshold \(T_{lost}\). For a pixel in \(OE_{i-1}\), if its \(T_{lost}\) neighborhood does not contain pixels belonging to \(SE_i\) or \(ME_i\), Then, this pixel should be the lost edge points. The lost edge in \(ith\) frame is defined as

\[
LE_i = \{ e \in OE_{i-1} \mid \|e - x\| > T_{lost}, x \in SE_i \cup ME_i \}
\] (9)

Finally, the target object edge is defined as follows.

\[
OE_i = SE_i \cup ME_i \cup LE_i
\] (10)

B. Distance Function and New Energy Function

Through aforementioned method, we can get a rough edge map of the foreground object, which can offer a strong hint for searching the real boundary of the target. We can use a distance function with the shape prior template as a new constraint in energy function. The distance function is defined as

\[
f_{dist}(x) = \min \|x - y\|, y \in OE_i
\] (11)

For efficiency, we only compute the distance in a narrow band around the rough edge map \(OE_i\). After taking account the distance function, the final energy function becomes Eqn. (12).

\[
E_{newshape} = \int_0^{l(C)} g(C(s))ds + \lambda \int_0^{l(C)} \Psi(m'(s))ds + \int_0^{l(C)} \Theta(C(s), S(m(s)))ds + \gamma \int_0^{l(C)} f_{dist}(C(s))ds + \nu \int_0^{l(C)} (\beta_C(s) - \beta_S(m(s)))^2ds
\]

Due to the introduction of distance function, the edge weight of product graph should plus a new term \(E_{dist}\), which is defined as:

\[
E_{dist} = \sum_{m=1}^{j-1} \frac{1}{2} (|x - y| + |y - z|) \cdot \max \{f_{dist}(x), f_{dist}(y), f_{dist}(z)\}
\] (13)

Only the turning angle related three pixels are on the target object contour, this turning angle is valid. The integral of the distance function is also minimal. The object edge will have big weight in energy function. Even in low contrast image, the energy function will converge to real boundary of the object, thereby, our proposed algorithm can robustly segment the foreground object with rotation and structure change. To solve the Eqn.(12), we can use the Lawler’s minimum ratio cycles method in the product graph [22], [23]. Negative cycles can be found via the Moore-Bellman-Ford algorithm for distance calculations. For graphs without negative cycles the algorithm computes the distance from a given root node \(r\) to all other nodes in the graph.

V. EXPERIMENTAL RESULTS

We implement this proposed algorithm with VC++ and OpenCV library. In all experiments, we set \(K = 3\), \(\lambda = 0.2\), \(\nu = 0.8\), and \(\gamma = 0.005\). Here, we set \(\gamma\) as a very small value, if \(\gamma\) is too big, the target curve would have stronger influence which leads to the segmented result almost same as the shape prior template and artifact noise. Fig. 7 shows an experiment of a walking actor. From frame 1 to 7, the left hand disappears gradually, and from frame 8 to 30, the left hand appears again. From this result we can find our new algorithm can correctly segment and track the the moving hand with local rotation and structural change. In the second experiment, the actor performs a large scale action. Fig. 8 gives the segmentation and tracking results of the moving actor in frame 1, 10, 20, 30, 40, 50, 60 and 70, respectively. Compared with the bottom images in Fig. 2, the performing actor with structure changes can be correctly tracked over seventy frames. We should notice that the gray level of the actor’s wooly sweater is almost same as the background.
VI. CONCLUSIONS AND FUTURE WORKS

A. Conclusions

In this paper, aiming at the traditional shape prior based method can not segment the foreground object with local rotation and partial structure change, we propose a new prior shape based foreground segmentation method in low resolution and low contrast image to overcome the aforementioned drawbacks. Based on Schoenemann’s shape prior algorithm, some new constraints are added into the energy function to robustly track an object with shape structure change. The improvements of the energy function are as follows.

- A turning angle function is put into the energy function to rotate the local structure of the foreground object. When we compare the similarity of shape prior template and object contour, turning angle is the similarity measurement of corresponding points. Then, the new energy function can solve the problem of local part rotation.
- In order to track the object with structure change, we firstly extract the moving edge, and then the object edge. Next, we compute the distance of each pixel in narrow band. This distance function is combined in the energy function. Finally, we can use the new energy function with turning angle and distance function to effectively handle the object structure change.

B. Future Works

Though our new algorithm can process the object local rotation and object structure change, the computational timing cost is very prohibitive. The averaging processing time is over 60 sec for $640 \times 480$ images in our implementation, which is impractical in some real-time applications. The implementation of the proposed algorithm is only for the verification purpose. The code is not optimized and any speed up methods are not utilized. One of the future works is to accelerate the algorithm. One possible solution is using the GPU based parallel method to get a real-time implementation, like the work in [15].

VII. ACKNOWLEDGMENTS

This project is partially supported by Natural Science Foundation of China (NSFC:50805031,61063019) and Science & Technology Basic Research Projects of Shenzhen (No. JC200903120184A, JC201005260161A), Foundation of the State Key Lab of Digital Manufacturing Equipment & Technology (No. DMTKF2009013). Science & Technology Foundation of Shenzhen Nanshan(No.201002)
REFERENCES


