A PREDICTIVE CONTOUR INERTIA SNAKE MODEL FOR GENERAL VIDEO TRACKING

Hao Jiang and Mark S. Drew
School of Computing Science, Simon Fraser University,
Vancouver, B.C., Canada V5A 1S6
{hjiangb,mark}@cs.sfu.ca

ABSTRACT

In this paper, we present a modified snake model for the problem of general video object tracking. We introduce a new external force into the snake equation based on the predictive contour such that the active contour is attracted to a shape similar to the one in the previous video frame. New methods of contour prediction and contour smoothing are presented. The proposed methods can deal with the problem of an object's stopping movement temporarily and can also avoid the problem of the snake tracking into the object interior. Global affine motion estimation is applied to eliminate the effect of camera motion and hence the method can be applied in a general video environment. Experimental results show that the proposed method exhibits increased robustness over a traditional snake algorithm and works well for general video object tracking.

2. INTRODUCTION

The snake model [1] forms a general scheme for deformable object tracking. However, in practical video environments a traditional snake is found to be easily confused by complex backgrounds. One scheme to enhance the robustness of snakes is to predict the initial contour of the future frame based on motion information. If the contour prediction is accurate enough, the initial contour is adjacent to the true boundary of the object so that the snake can successfully track the object. Unfortunately, problems occur when the object stops moving and then moves on. For motion-based tracking, the snake will eventually collapse into one point, using the usual contour initialization scheme. Problems also occur when the features of interest exist not only at the boundary of the object but also inside the object. A prediction error often causes the snake to become trapped inside the object. This kind of error cannot be self-corrected and causes the failure of the tracking system. Therefore, using only contour prediction to re-initialize the snake cannot guarantee robustness of the tracking system. In [2], an affine motion constraint snake is put forward for increasing robustness, with object deformation restricted to an affine model. A similar scheme is presented by Blake [3]. Other schemes try to include dynamic information for the object. In [4], the incremental contour estimation is fit into a Bayesian estimation framework. Based on the probability representation, Kalman filtering can be incorporated into the snake tracking system. But the Gaussian assumption of a Kalman snake cannot be maintained in highly cluttered environments. Blake [5,6] proposed a CONDENSATION (conditional probability propagation) method to solve this problem. As for a Kalman snake, this method needs to learn the dynamics of the object, based on training sets, before tracking can be applied.

We propose a different snake model for tracking a single motion object in a general video environment that does perform robustly in these problem cases. In our model, another energy term is included, based on the predictive contour. This term *attracts the active contour to converge to a shape similar to the one in the previous video frame*. As well, instead of simply using the predictive contour to re-initialize the snake, we construct new initial contour by a uniform expansion along the normal of the previous contour. This scheme prevents the contour from erroneously tracking the features inside the true boundary of the object. At the same time, the new predictive contour inertia energy term makes the snake ignore distracting elements. As well, if the object stops moving temporarily, the snake will evolve according to the inertia term in the predictive contour and converge to a shape that corresponds to the motion prediction result. We adopt an affine motion model for global motion estimation and camera motion compensation with the result that our scheme can work in a general video environment. The contribution of the proposed method is that it presents a generation model without limiting the object's deformation or dynamics pattern so that it can fit general video tracking problems. A learning phase is also not needed in the proposed scheme.

The paper is arranged as follows. In §3, we present the modified snake method for general video tracking, with the modified snake equation presented in §3.1. We discuss contour prediction and smoothing in §3.2, and in §3.3 set out affine global motion estimation and snake external force generation. Experimental results are given in §4 and conclusions in §5.

3. MODIFIED SNAKE

3. 1 Snake Equation with Predictive Contour Inertia

In this section, we present an equation for enhancing the robustness of active contours in the tracking problem. We formulate a new snake equation:

$$\min \int_s \frac{\alpha}{2} |\nabla X(s)|^2 + \frac{\beta}{2} |\nabla^2 X(s)|^2 + P(X(s))ds + \frac{\lambda}{2} E(X(s),C(s)) \qquad (1)$$

where $X(s): R \to R^2$, $s \in [0,1]$, is the active contour for the current frame, $C(s): R \to R^2$ is the predictive contour from the previous frame, $E(X(s),C(s))$ is a new predictive contour inertia energy term which measures the difference between $X(s)$ and $C(s)$. As for a traditional snake, the internal energy and smoothness of the active contour is represented by the first two terms, expressing tension and rigidity. $P(X(s))$ is the external energy based on the features of interest such as edges or object motion. For the new prediction-correction term E(X(s),C(s)) we posit the function

$$E(X(s),C(s)) \equiv \int_s [x(s) - C(s)]^2 ds \qquad (2)$$

in order to restrict the snake from changing too much from frame to frame.

With Eq.(2), the variational derivative of Eq.(1) yields an Euler-Lagrange equation

$$\alpha X_{ss} - \beta X_{ssss} - \nabla P(X) + \lambda(C - X) = 0 \qquad (3)$$

The steepest descent solution of (3) is generated by

$$\frac{\partial X}{\partial t} = \alpha X_{ss} - \beta X_{ssss} - \nabla P(X) + \lambda(C - X) \qquad (4)$$

The generalization of the above equation derives from substituting $F(X)$ for $-\nabla P(X)$, yielding the generalized modified snake equation

$$\frac{\partial X}{\partial t} = \alpha X_{ss} - \beta X_{ssss} + F(X) + \lambda(C - X) \qquad (5)$$

We employ finite differences to discretize Eq.(5): let X_i^n denote the ith point in the active contour at discrete artificial relaxation time n,

$$X_i^n = (x_i^n, y_i^n) = (x(i\Delta s, n\Delta t), y(i\Delta s, n\Delta t)) \qquad (6)$$

where Δs and Δt are distance intervals for the active contour and time, respectively. Then Eq.(5) becomes

$$\begin{aligned} \frac{X_i^{n+1} - X_i^n}{\Delta t} = {} & \frac{\alpha}{\Delta s^2}[X_{i+1}^{n+1} + X_{i-1}^{n+1} - 2X_i^{n+1}] - \\ & \frac{\beta}{\Delta s^4}\{[X_{i+2}^{n+1} + X_i^{n+1} - 2X_{i+1}^{n+1}] - \\ & 2[X_{i+1}^{n+1} + X_{i-1}^{n+1} - 2X_i^{n+1}] + \\ & [X_i^{n+1} + X_{i-2}^{n+1} - 2X_{i-1}^{n+1}]\} + F(X_i^n) + \lambda(C_i^n - X_i^n) \end{aligned} \qquad (7)$$

In matrix form, Eq.(7) reads

$$X^{n+1} = (I - A)^{-1}[\Delta t F(X_i^n) + \Delta t \lambda (C_i^n - X_i^n) + X_i^n] \qquad (8)$$

where $X^n = (X_1^n, X_2^n ... X_N^n)$ and A is the N by N coefficient matrix. The parameters $\Delta t\alpha / \Delta s^2$, $\Delta t\beta / \Delta s^4$, $\lambda\Delta t$ and Δt are to be determined. If $F(x)$ is normalized to [0,1], typical parameters settings are 0.4, 0.4, 0.25 and 0.6, respectively.

3.2 Contour Prediction and Smoothing

We predict the future contour position and shape by the method of block-wise motion estimation followed by a smoothing process. For every point (x,y) in the contour, a block size of d is constructed centered on the pixel. By block matching in a searching window of size w, the best matching block center is selected as the estimated position of the predictive contour.

$$\begin{aligned} &(u,v) = X(s) \\ &(\Delta\hat{x}, \Delta\hat{y}) = \min_{(\Delta x, \Delta y) \in W} \sum_{u,v} |I(u,v) - I(u + \Delta x, v + \Delta y)| \\ &C(s) = X(s) + (\Delta\hat{x}, \Delta\hat{y}) \end{aligned} \qquad (9)$$

where $X(s)$ is the current contour and $C(s)$ is the predictive contour, $W = [-w/2, w/2] \times [-w/2, w/2]$. Motion estimation sometimes fails to estimate the correct contour position: this occurs if some part of the previous contour does not fall on the boundary of an object. This situation is very common for snake tracking of objects with concave boundaries. To eliminate the error prediction points, a global affine motion model can be used. This approach can be viewed as a smoothing process for the motion estimation result. We propose a different approach for smoothing the prediction contour. The smoothing process is a self-evolving curve without the external force,

$$\frac{\partial C}{\partial t} = \alpha_0 X_{ss} - \beta_0 X_{ssss} \qquad (10)$$

A stopping condition (fixed number of iterations) has to be specified so that the curve will not distort too much while smoothing the singular points. The parameters of α_0 and β_0 are determined by experiment—α_0 should be smaller than β_0 to prevent over-shrinking the curve. The second-order term β_0 acts as a Gaussian filter to smooth the contour.

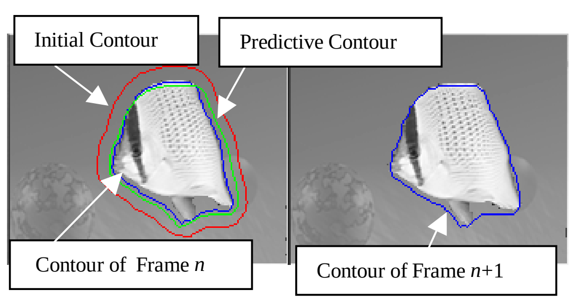


Figure 1. Scheme of modified snake. (a): blue- previous contour; red- initial contour; green- predictive contour. (b): new contour.

Before the search, the contour is initialized as a uniform expansion of the previous frame active contour:

$$X^i = X + c\vec{n} \qquad (11)$$

where $\vec{n}$ is the exterior normal vector of the contour, and c is a constant. In our scheme c = 10. Fig.1 illustrates the scheme for construction of predictive contours and initial contours.

3.3 Global Motion Estimation and External Force Generation

In the modified snake-tracking scheme, we use motion as the object feature. We extract the moving object from the scene and track it along video frames. To remove camera motion, we estimate global motion in the whole picture. The affine motion equation is

$$\nabla I^T (B(x,y)\vec{p}) + I_t = 0 \qquad (12)$$

where $\vec{p}$ is the affine motion vector and B is the affine motion matrix. The Least Squares solution [7] of the Eq.(12) is

$$\vec{p} = (\sum B^T \nabla I \nabla I^T B)^{-1} (-\sum I_t B^T \nabla I) \qquad (13)$$

We calculate global motion based only on the region outside of the motion object. The object region is defined as the region inside the initial contour. *The object region is not included during global motion estimation*. We estimate global motion in a low resolution version of the image and scale the transition values. Based on the estimated affine model, video frames are warped and frame differences are calculated. We then use a simple thresholding scheme to detect the motion object (More complex models including optical flow could be incorporated into the algorithm.) Motion detection produces an image segmentation map: Fig.2(c) shows an example. As expected, a thresholded direct frame difference image as in Fig.2(d) fails to detect the moving object—global motion compensation is a key step for object extraction. Based on the segmentation map, the external force field of the snake model is constructed based on the Gradient Vector Flow scheme [8]: the external force F=[u,v] is defined via

$$\min \iint_{x,y} |\nabla I|^2 |F - \nabla I| + \mu(u^2{}_x + u^2{}_y + v^2{}_x + v^2{}_y)dxdy \qquad (14)$$

with μ a weighting coefficient. The steepest descent solution is

$$u_t(x,y,t) = -(I_x{}^2 + I_y{}^2)(u - I_x) + \mu(u_{xx} + u_{yy})$$
$$v_t(x,y,t) = -(I_x{}^2 + I_y{}^2)(v - I_y) + \mu(v_{xx} + v_{yy}) \qquad (15)$$

In our numerical implementation,

$$I_x \approx [I(m+1,n) - I(m-1,n)]/2$$
$$I_y \approx [I(m,n+1) - I(m,n-1)]/2$$
$$u_{xx} \approx u(m+1,n) + u(m-1,n) - 2u(m,n) \qquad (16)$$
$$u_{yy} \approx u(m,n+1) + u(m,n-1) - 2u(m,n)$$

where we assume unit grid size. The discretization of v_{xx} and v_{yy} follows the same scheme as u. Fig.2 shows the schema of the proposed method. Fig.2(j,k) show that the modified snake ignores distracting elements and tracks the true object.

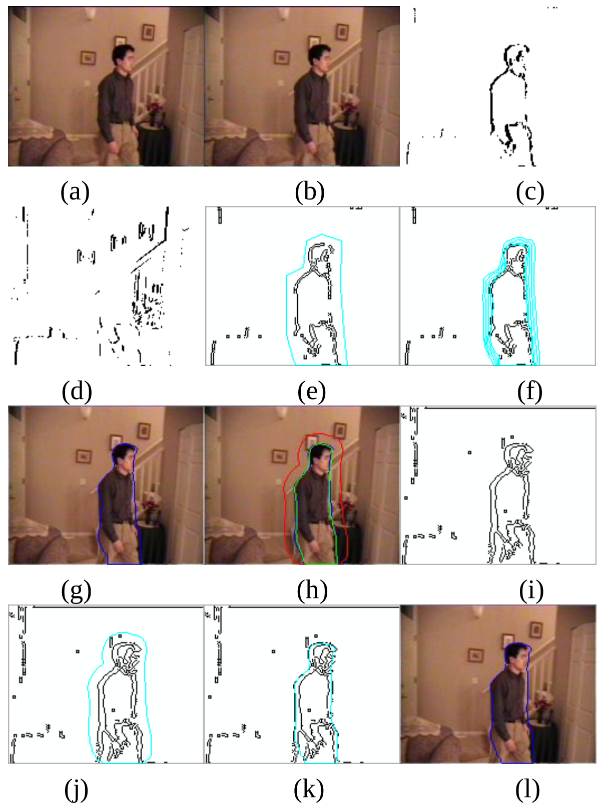
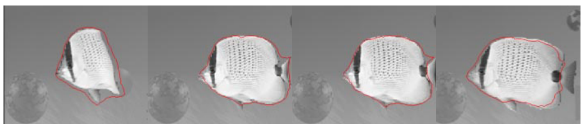


Figure 2. Modified snake tracking. (a) Frame 1. (b) Frame 2. (c) Motion detection with global motion compensation. (d) Motion detection without global motion compensation. (e) Initial contour drawn by hand. (f) Snake attracted to the object. (g) Tracking result for Frame 1. (h) Initial contour and predictive contour. (i) Edge detection of motion map for Frame 2. (j) Initial contour. (k) Snake converges to the boundary of object. (l) Tracking result for Frame 2.

4. EXPERIMENTAL RESULTS

We compared the proposed snake tracking method with the usual method of using the predictive contour as the initial one in the future tracking frame, using sequence Bream. We use a simple change detection algorithm with a square 3 by 3 window to extract object features. The threshold is set to be 30 for a coarse segmentation of the fish. Motion pixels are those with consecutive frame differences after global affine motion compensation above the given threshold. The contour for the first frame is drawn by hand near the boundary of the object. Contour prediction is based on full-search block matching motion estimation, with block size 7 by 7 and search range [-5,5] by [-5,5].

Fig.3 shows the result for the traditional algorithm which uses the predictive contour to initialize snake tracking. The starting frame is 122. In frame 129, an error occurs for contour prediction—the contour crosses the object; since there is motion detected inside the fish, these pixels have an external force pulling the contour inside. As shown by experiment, this error cannot correct itself. This kind of problem also occurs in frame 156 and causes the contour enter the object further. Experimental results for the proposed method are shown in Fig.4. The starting frame is 1, with a total of 180 frames. Because of the new term we introduce to limit contour changes, the extra force makes tracking much more robust. Similar results for video sequence Foreman are shown in Fig.4.



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

Figure 3. Tracking result of traditional snake. Starting from frame 122. (a) Frame 122. (b) Frame 129. (c) Frame 130. (d) Frame 156.

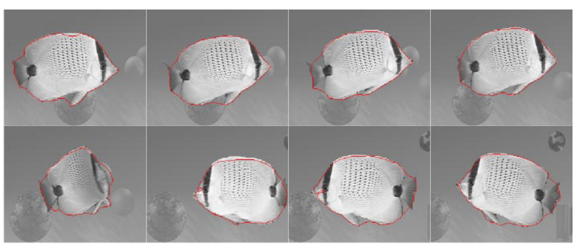


Figure 4. Modified snake tracking result for Bream.

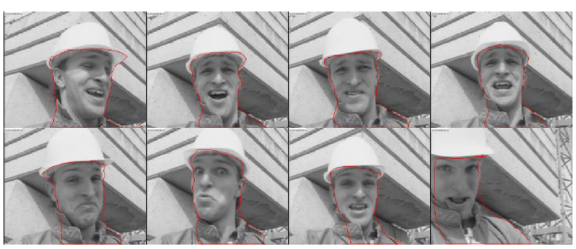
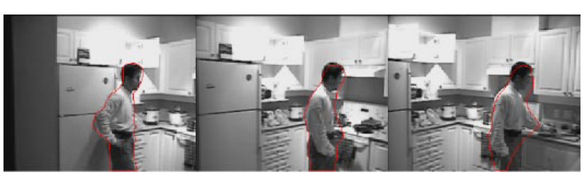


Figure 5. Modified snake tracking for Foreman.



(a) (b) (c)

Figure 6. Tracking result of modified snake for a panning image sequence. (a) Frame 1 (b) Frame 30 (c) Frame 50.

In the tracking experiments for Bream and Foreman, the object's motion is sometimes less than the threshold specified. Experiments show that the proposed method can still track the object for this kind of temporary slow motion or still state case.

Another experiment was carried out to test the proposed method for tracking a walking man with camera panning[1], with tracking result shown in Fig.10. The proposed tracker works well for this type of general video as well.

5. CONCLUSIONS

In this paper, we present a modified snake model for the problem of general video tracking. In the proposed scheme, we introduce a new force into the snake equation based on the predictive contour. Global affine estimation is applied to eliminate the effect of camera motion; therefore the system can be applied in a general video environment. We also present schemes for contour prediction and smoothing. We present a generation model without limiting the object's deformation or dynamics pattern and thus it can be applied to general video tracking problems. As well, a learning phase is not needed in the proposed scheme. In future work, other features such as color will be incorporated to further increase robustness. The complexity of the proposed algorithm is mainly due to the motion estimation and external GVF force generation steps. Real-time implementation is another future research topic.

6. REFERENCES

[1] M. Kass, A.Wikin, and D. Terzopaulos, "Snakes: active contour models", Int. J. Comput. Vision, 1:321-331, 1988.

[2] B. Bascle, P. Bouthemy, R. Deriche and F. Meyer, "Tracking complex primitives in an image sequence", ICPR'94, A:426-431.

[3] A. Blake, R. Curwen and A. Zisserman. "Affine-invariant contour tracking with automatic control of spatial-temporal scale". ICCV'93, 66-75.

[4] D. Terzopoulos and R. Hallinan, "Tracking with Kalman snakes", *Active Vision*, The MIT Press, 1992, pp.4-20.

[5] M. Isard and A. Blake, "Contour tracking by stochastic propagation of conditional density", ECCV'96, 1:343--356.

[6] M. Isard and A. Blake, "CONDENSATION – Conditional density propagation for visual tracking", Int. J. Comp. Vis., 29, 5-28, 1998.

[7] D.J. Heeger, "Notes on motion estimation", Oct. 30, 1996. http://white.stanford.edu/~heeger.

[8] C. Xu and J.L. Prince, "Snake, shapes, and Gradient Vector Flow", IEEE Trans. Image Proc., 7:359-369, 1998.

[1] Video is available in www.sfu.edu/~mark/Icip01