FAST IMPLEMENTATION OF VARIATIONAL, CONTOUR-BASED OBJECT TRACKING

Petr Dokladal (1), Eva Dokladalova (2)

(1) School of Mines of Paris
35, rue Saint Honoré, 77300 Fontainebleau, France
e-mail: petr.dokladal@ensmp.fr

(2) DRT/DTSI/SARC/LCEI
CEA Saclay, 91191 Gif sur Yvette, France
e-mail: eva.dokladalova@cea.fr

ABSTRACT

We present a fast contour-based object tracking using an active contour which clings towards the maximum of contrast alongside the object boundary. The contour motion, modeled by a time-differential equation, is driven by a gradient descent flow with a variable integration time step. The best-fit time step to reach the extrema of the functional is estimated in the least squared error sense. The stability of the solution is reached in a few iterations. The performance is shown on a driver's head tracking application.

Keywords: Object Tracking, Active Contours, Level Set

1. INTRODUCTION

The object extraction and tracking can be done by using two approaches: i) region-based techniques and ii) contour-based techniques. Compared to the contour-based techniques, the region-based methods make use of statistical information from inside the regions. Their robustnesse is usually better w.r.t. errors in data, but their computational complexity is also one magnitude greater.

In this paper we concentrate on contour-based object tracking. It will cling to the maxima of the contrast alongside the object contour. We revisit the technique presented in [4] where the curve representing the contour is driven to the new position by an attraction force pointing from both sides towards the maximum of the contrast. There, where the contrast is not strong enough, the contour doesn't move to prevent leakage.

2. BASIC NOTIONS

2.1 Active Contours and Snakes

An important breakthrough in variational contour-oriented object extraction domain represented the introduction of the snakes by Kass *et al.* in [6] in 1987 and by Terzopoulos *et al.* [13] one year later. Another model was presented by Cohen [3]. The geodesic active contours were studied later by Caselles *et al.* in [2] and by Kichissammany in [7]. Another model was proposed later in [9].

We consider a commonly adopted context of a parametrically represented, closed, planar curve $C : [a,b] \subset \mathbb{R} \to \mathbb{R}^2$. If $C(a) = C(b)$ for $a = b$ the curve is closed. If for $\forall p,q \in]a,b[$, $p \neq q$ it holds $C(p) \neq C(q)$ the curve is simple (i.e. without intersections). We assume the generally adopted scheme for a moving curve $C(t,s)$, with s being the euclidean arc length parametrizing the curve and t representing the time. The travelling curve obeys

$$\frac{}{t}C(t,s) = v\vec{N} \tag{1}$$

$\vec{N}$ represents the normal vector to C and v the (scalar) travelling speed.

Equation 1 is often used to drive snakes, where the solution is sought by minimizing (or maximizing) some energy $\mathscr{E}$.

$$C = \arg\min_{\mathscr{C}} \mathscr{E} \tag{2}$$

The commonly used gradient descent technique solves Eq. 2 by employing $v = - \quad \mathscr{E}$ in Eq. 1. It usually finds in $\mathscr{E}$ the local minimum which is marked by the initial condition C_0.

Consider an image $I : \mathbb{Z}^2 \to \mathbb{Z}^+$ and $\quad I$ some gradient on I. We search for a simple and closed curve $C \subset \mathbb{R}^2$ in a family of such curves $\mathscr{C}$. We want C to cling alongside the maximum of contrast in I, i.e. we search for

$$C = \arg\max_{\mathscr{C}} \oint_C | \quad I| ds \tag{3}$$

Here, the solution C is a closed curve and the $\mathscr{C}$ the family of all closed curves. The energy $\mathscr{E}$ to maximize is the contrast $| \quad I|$ alongside C. The extrema of $\quad I$ are found where $\quad {}^2 I = 0$. If the solution to Eq. 3 can generally not be found analytically, the fastest numerical solution can be obtained by means of a curve travelling according to Eq. 1 by

$$\frac{}{t}C(t,s) = \quad I\vec{N} \tag{4}$$

Recall that the laplacian $\quad$ often denotes $\quad {}^2$. The normal speed $\quad I.\vec{N}$ drives the curve C towards the maximum of $\quad I$ where it stops. This approach is similar to the edge detection by using the zero-crossing of the Laplacian in the sense that the functional (3) searches for local minimum where the laplacian is zero.

Note that $\quad I$ in Eq. 3 must be $\mathbb{R}^2 \to \mathbb{R}$ (calculated on $\mathbb{Z}^2$ e.g. by bilinear interpolation). Maximum of contrast may represent the spatial, or the temporal gradient, if one wants to detect moving objects.

2.2 The Level Set Context

The level set was proposed by Osher and Sethian in 1988 in [12] as a simple method to model or analyze the motion of a travelling interface. The implicit representation of the interface as a constant-level set of another function was studied by Caselles *et al.* [1] and later by Malladi *et al.* [10] and [8]. For applications and other references see e.g. [11].

The interface C travelling in $\mathbb{R}^2$ is represented implicitly in $\mathbb{Z}^2$ by a signed distance function $u : \mathbb{Z}^2 \to \mathbb{R}$ calculated to

C. C therefore coincides with $u = 0$. The parametric curve motion Eq. 1 loses the parameter s

$$u_t = -\vec{\mathscr{F}}. \quad u \tag{5}$$

u replaces $\vec{N}$ since u points in the normal direction to C and $| \quad u| = 1$ if u is a distance function. The function u_t denotes the derivative of the distance in time. Consequently, $v = \vec{\mathscr{F}}. \quad u$ is the travelling speed. The term $\vec{\mathscr{F}}. \quad u$ is in Eq. 5 negative since we consider $\mathscr{F}$ as an attracting (and not pushing) force.

3. CONTOUR TRACKING CONTEXT

The attraction force $\mathscr{F}$ in Eq. 5 is based on some gradient calculated in the image sequence. According to what is known about the objects to extract one chooses the appropriate gradient. For example, if one wants to extract moving objects in a stale background, one may choose the temporal gradient extracting the motion. The motion will be detected at least on the contours of the objects. Also, if one can make a hypothesis on the hue then the appropriate gradient is taken on the hue. The hue can successfuly be used to track the skin, for example.

Suppose that one can make a hypothesis on the maximum velocity of the objects, and consequently the maximum frame-to-frame displacement R of the objects given the configuration of the scene, position of the camera, number of frames per second, etc. Then the closed curve C tracking the object will be sought in some compact neighborhood around the (known) object's contour C_0 (obtained in the previous frame). In this sense, $\mathscr{C}$ becomes in Eq. 3 the family of curves laying in the R-neigborhood of C_0. The R-neighborhood denotes the dilation of C_0 by a ball with radius R.

Let I be an image containing non-textured, i.e. smooth objects to track. The gradient $\quad I$ is (almost) zero in the smooth zones and contains high values (diracs) on the objets boundaries. Tracking an object within a given R-neighborhood can be achieved by using the attraction force $\vec{\mathscr{F}} = \quad g$, where $g = K * \quad I$, wkere K is a 2-D triangular window: $\mathbb{Z}^2 \to \mathbb{R}^+$, such that

$$K(x,y) = \begin{cases} 1 - \quad (x^2+y^2)^{1/2} & \text{if } (x^2+y^2)^{1/2} < 1/ \\ 0 & \text{otherwise} \end{cases} \tag{6}$$

To extend the zone $\mathscr{C}$ to R pixels at each side, use $\quad = 1/R$. Convoluting $\quad I$ with K extends the attraction field of each impulse to the zone $\mathscr{C}$, where the solution will be sought. $\vec{F} = \quad g$ points towards the maxima of $\quad I$. Secondly, convoluting an originally non-continuous function $\quad I$ with a continuous function K makes g a continuous function, permitting to employ a fast gradient descent technique presented below.

In this context one looks for the *strong* solution C of the functional Eq. 3 in the domain $\mathscr{C}$ corresponding to the R-neighborhood of the initial condition C_0. Looking for the strong solution is made easier by the fact that g is smoothed with respect to the original I.

4. GRADIENT DESCENT WITH VARIABLE STEP

The Eq. 4 represents a monotonically advancing curve C which stops wherever v becomes zero. Implemented in a discrete form, one has to integrate with a small dt to prevent from creating shocks.

We use Eq. 4 to have C converge fast towards a functional solution as in Eq. 2. We want C:

1. to remain continuous and
2. not to quit the attraction basin in $\quad \mathscr{E}$ in Eq. 2

Note: In the context of functional solution we don't need C to advance monotonically.

4.1 Estimation of the time step

To ensure 1), it can be shown that v must be continuous if C has to remain continuous. The condition to ensure the continuity of v over the support of an image f is

$$dt < \min_{x \in supp(f)} \frac{dx}{\max_{x_i \in N_4(x)} |\mathscr{F}(x) - \mathscr{F}(x_i)|}$$

where $N_4(x)$ is the 4-neighborhood of x and dx the space discretization step. Practically, we do not need to check the continuity over the whole supp(f) but only over the domain where C evolves.

To ensure 2), we analyse the attraction force $\mathscr{F}$. First, we estimate in the least squared error sense (LSE) the best-fit time step to reach the functional (Eq. 3) minimum. Since the LSE value is obtained globally and not locally, it doesn't guarantee that the contour will not locally get outside the attraction basin. Secondly, the value is therefore compared with the minimum time step, necessary to reach another attraction basin.

- The attraction force equals $\vec{\mathscr{F}} = \quad (K * \quad I)$. The contour travels with speed $v = \vec{\mathscr{F}}. \quad u$, which basically means that even if $|\vec{\mathscr{F}}| \neq 0$ its effect on the contour is zero if $\vec{\mathscr{F}}. \quad u$=0. (Which prevents leakage on object boundaries corrupted with noise, since $\mathscr{F}$ becomes perpendicular to C inside noisy breaches.)
- From Eq. 5 we know that the contour is attracted towards (repelled from) zones where v is positive (negative, resp.). We can identify these zones $X \in \mathbb{Z}^2$ by taking $X = \{x \mid \text{sign}(u(x))\text{=sign}(v(x))\}$.
- Recall that $C \subset \mathbb{R}^2$. Let CS denote $\mathbb{Z}^2$-grid points adjacent to C, $CS = \{x \mid |u(x)| \leq 1\}$. Note that $X \cap CS \neq \emptyset$ (except when the stability is reached; in which case $X = \emptyset$ itself). X contains several connected components: $\{X_i\} = \text{cc}(X)$. Here, a connected component in a binary set is the equivalence class for path-connectedness of sequentially adjacent points. To separate the attraction basins, we need to use the 4-connection. The operator cc denotes the extraction of connected components from a binary set. In the zone X we identify the current attraction basin $B = X_i$ such that $X_i \cap CS \neq \emptyset$.
- In the basin B, we identify the local minima of v: $M = \{x \mid v(x) \approx 0,\ x \in B\}$. The stability of C is predicted to be reached in these points.
- We estimate the best time step t to get into the minimum by solving an overdeterminated system for the LSE value of t for all $x \in M$:

$$v(\lfloor x - u(x) \quad u(x) \rceil)\, t = u(x)$$

$\lfloor . \rceil$ denotes rounding to the nearest integer and $(x_j - u(x_j) \quad u(x_j))$ in $v(.)$ means the descent from x on $\quad u$ towards $u = 0$ to use the speed v on the zero-level set of u.

Condition to stay in the local attraction basin:

- The condition to respect, in order to remain in the current attraction basin B, is to impose t inferior to the minimum necessary time step to reach other connected components of X. The other attraction basins are $\{X_i \mid X_i \cap CS = \emptyset\}$. In this case, all points in each X_i are taken into the computation of t_{other}, not only the minima.

For all $x_j \in X_i$, over all X_i, we solve a set of equations

$$\{t_j\} = (u(x_j) - 1) \,/\, v(\lfloor x_j - u(x_j) \quad u(x_j) \rceil)$$

and take $t_{other} = \min\{t_j\}, \forall t_j > 0$.

We consider $u(x_j) - 1$ to compute the time *before* C reaches another basin.

4.2 Integration

- Integrate Eq. 5 with the smalest value obtained in the section 4.1 above: $t = \min\{t,\, t_{other}, dt\}$:

$$u^{n+1} = u^n - vt$$

Note: Obviously, this scheme implements euler integration of Eq. 5, and does not pretend to be absolutely stable. The time step can not be arbitrarily large. We have observed that reasonably fast integration, without developing oscillations, allows the contour to advance in one iteration by five to ten pixels.

Fig. 1 gives an example (taken from the sequence below and donwsampled for illustration). Fig. 1 (a) gives the initial condition C_0 obtained from the previous frame. C_0 is placed in the force field $\mathscr{F}$, attracting C towards the local maximum of contrast. $\mathscr{F}$ is computed from the gradient $g = \quad h$ in an HSV-coded image $I = I(h,\ s,\ v)$ by taking $\mathscr{F} = \quad (K * g)$ with K four pixels wide, i.e. $R = 2$ (cf. Eq. 6). Fig. 1 (b) gives the zone $X = \{X_1,\ \ldots\ X_6\}$ where C is attracted. The current attraction basin B equals X_1. The curve C converges towards the minimum of B (set M at Fig. 1 (c)) and finally Fig. 1 (d) gives the contour position C^1 after one iteration.

4.3 Object Tracking Scheme

The algorithm is based on the habitual integration scheme used to model the motion of the contour. The upper index n denotes the iteration number.

```
$n = 0$
Initialize $u^n$
-----------------------------
repeat
    Estimate the best integration step $t$
    Integrate: $u^{n+1} = u^n - vt$
    Reconstruct $u^{n+1}$
    $n = n + 1$
    until stability $v \approx 0$
```

Initialization: The initialization of u^0 is done by reconstruction of u from the result u of the previous frame (or by interpolation for the first frame in the sequence).

Best time step estimation: The integration step t to compute the curve motion in the current frame is estimated by using the speed v obtained from the gradient on the current frame and the distance u^n in the current integration step. The integration is repeated until the contour reaches the stability. The stability is reached as soon as the speed of the contour C becomes zero $v(x) \approx 0$, $\forall x$, $|u(x)| \leq 1$ and $u^n = u$.

Integration: During the integration to compute the motion of u, we use the simple idea (already studied in [5]) that all levels must travel with the same speed as the set $u = 0$, otherwise u deforms.

The speed $v(x)$ of all points $x \in \mathbb{Z}^2$, $u(x) \neq 0$ is obtained by $v(x) = v(x_0)$, where $x_0 = x - \quad u(x)u(x)$.

Since $x_0 \in \mathbb{R}^2$ (and not in $\mathbb{Z}^2$) Gomez and Faugeras [5] propose obtaining the value $v(x_0)$ by bilinear interpolation. For our application we use $x_0 \in \mathbb{Z}^2$, $x_0 = \lfloor x - \quad u(x)u(x) \rceil$ where $\lfloor . \rceil$ denotes rounding to the nearest integer to obtain the $\mathbb{Z}^2$-grid point closest to the intersection of $u = 0$ and the extension of $\quad u(x)$. Obtaining $x_0 \in \mathbb{Z}^2$ can be done fast in one vector-like operation for all x where u is defined and leads to a significant limitation of memory accesses.

Reconstruction: The Gomez technique can be used to have C travel fast (several pixels in one step), but doesn't prevent from having to reconstruct periodically u. Indeed, even if C travels with a normal speed (the tangential component being always assumed zero), it doesn't mean that C doesn't bend and that $\quad u^{n+1}$ remains parallel to $\quad u^n$. The function u actually loses its properties ($| \quad u| = 1$) whenever any two close points on C travel with different speed and its normal $\vec{N}$ turns. Indeed, the necessary condition to preserve $| \quad u| = 1$ is $v = const.$ which is equivalent only to morphological erosion/dilation. Nonetheless, the Gomez technique efficiently prevents from creating discontinuities.

5. APPLICATION RESULTS

Tracking of the driver's head (see Fig. 2) was done on the boundary of thresholded, filtered skin hue of the input HSV-coded image $I = I(h,s,v)$. The skin area is obtained by

$$h_{skin} = \begin{cases} 1 \text{ for } h \in [-77^\circ,\, 84^\circ] \\ 0 \text{ otherwise} \end{cases}$$

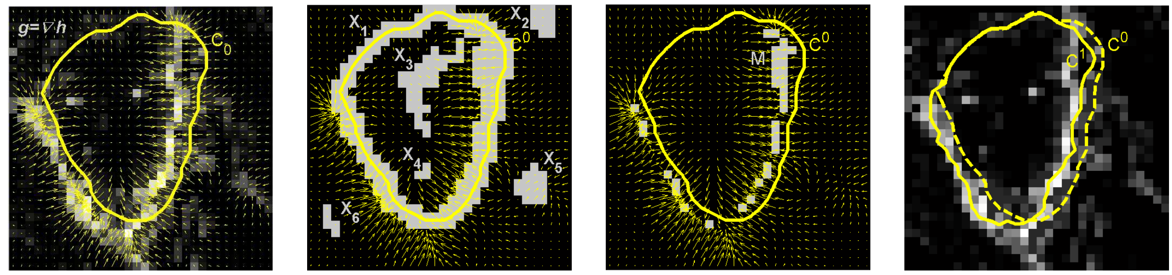
and then $H = \arg\max \mathrm{cc}(h_{skin})$ to obtain the largest connected component, which is further filtered by $\quad_B(.)$ denoting morphological filtering by opening and closing by B. We use B being five point diameter ball. The attraction force $\mathscr{F} = \quad (K * \quad_B(H))$, with K of five pixels radius at the base. $\quad_B(H)$ contains a Dirac impulse at the boundary of the largest, filtered, skin component and zero elsewhere. $\quad_B(H)$ itself is unstable, contains oscillations and is not visually nice. Nonetheless, it can successfuly be used to drive the tracking.

For the tracking algorithm one iteration per frame was sufficient. This could be done thanks to the frame rate of the camera (15 frames par second), limiting the inter-frame motion of the head to less then five pixels. (We have observed that ten pixels per iteration is probably the maximum motion which can achieved with this iteration in one iteration. Tracking higher motion speed in one iteration would lead to oscillations.)

Compared to the tracking method presented in [4], where the gradient descent was done with a fixed-step euler integration in about 20 iterations per frame, with 4 intermediate reconstructions of u, we obtain almost 20 times increase of the speed.

6. CONCLUSIONS

This paper proposes a variational-based contour object tracking. The method uses an active contour which clings towards the maximum contrast in the image. The motion of the contour is controled by a time-differential equation. Its solution



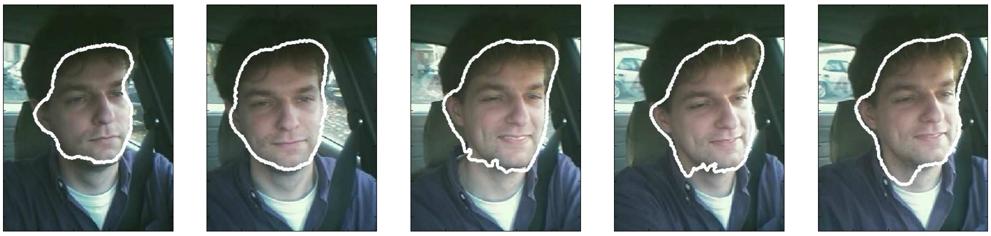
(a) Initial position C_0, gradient $g = \quad h$, and force $\mathscr{F}$ (arrows)

(b) Attraction zone $X = \{X_i\}$

(c) Minima M in the attraction basin: $M \subset B = X_1$.

(d) Contour position: initial C_0, and after one iteration C^1

Figure 1: One iteration of the contour motion scheme



(a) Frame 008

(b) Frame 047

(c) Frame 087

(d) Frame 124

(e) Frame 130

Figure 2: Tracking of the drivers head. Randomnly chosen frames from a movie.

is found by using the euler integration scheme. The convergency is accellerated by employing in the integration a variable time step. The best-fit time is estimated in the sense of least squared error to speed up the convergence towards the predicted solution. The stability is reached in a limited number of iterations.

This technique is applied to the extraction of car driver's head (originally published in [4]) from a sequence acquired by a camera embedded in the vehicle. Using the best-fit integration step prediction the tracking of the head could be done in one iteration per frame which represents a speed up by almost twenty times over the traditional euler integration.

REFERENCES

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

[2] V. Caselles, R. Kimmel, and G. Sapiro. Geodesic active contours. In *ICCV*, pages 694–699, 1995.

[3] L. Cohen. On active contour models and balloons. In *Computer Vision, Graphics, and Image Processing*, volume 53, pages 211–218, 1991.

[4] P. Dokládal, R. Enficiaud, and E. Dejnožková. Contour-based object tracking with gradient-based contour attraction field. In *Proceedings ICASSP*, May 2004.

[5] J. Gomez and O. Faugeras. Reconciling distance functions and level sets. Technical report No: 3666, INRIA, avril 1992.

[6] M. Kass, A. Witkin, and D. Terzopoulos. Snakes: Active contour models. *International Journal of Computer Vision*, 1(4):321–331, 1987.

[7] S. Kichissammany, A. Kumar, P. J. Olver, A. Tannenbaum, and A. Yezzi. Gradient flows and geometric active contours. In *Fifth International Conference on Computer Vision*, pages 810–815, Cambridge, 1995.

[8] R. Malladi, R. Kimmel, D. Adalsteinsson, G. Sapiro, V. Caselles, and J. A. Sethian. A geometric approach to segmentation and analysis of 3d medical images. In *Proc. Mathematical Methods in Biomedical Image Analysis Workshop*, June 21-22 1996.

[9] R. Malladi and J. A. Sethian. Image processing : Flows under min/max curvature and mean curvature. *Graphical models and image processing*, 58(2):127–141, 1996.

[10] R. Malladi, J. A. Sethian, and B. C. Vemuri. Shape modeling with front propagation: A level set approach. *IEEE TPAMI*, 17(2):158:175, February 1995.

[11] S. Osher and N. Paragios, editors. *Geometric level set methods in imaging, vision and graphics*. Springer Verlag, New York, 2003.

[12] S. Osher and J.A. Sethian. Fronts propagating with curvature-dependent speed: Algorithms based on Hamilton-Jacobi formulations. *Journal of Computational Physics*, 79:12–49, 1988.

[13] D. Terzopoulos, A. Witkin, and M. Kass. Constraints on Deformable Models: Recovering 3D Shape and Nonrigid Motions. *Artificial Intelligence*, 36:91–123, 1988.