Improvement of convergence and stability in moving object extraction by the level set method

著者	Imamura Kousuke, Oota Tomohiro, Hashimoto Hideo
journal or publication title	ISPA 2009 - Proceedings of the 6th International Symposium on Image and Signal Processing and Analysis
page range	424-429
year	2009-01-01
URL	http://hdl.handle.net/2297/19926

Improvement of Convergence and Stability in Moving Object Extraction by the Level Set Method

Kousuke Imamura*, Tomohiro Oota**, Hideo Hashimoto†
Graduate School of Natural Science and Technology
Kanazawa University
Kakuma-machi, Kanazawa, Ishikawa 920-1192 Japan
{*imamura,†hasimoto}@ec.t.kanazawa-u.ac.jp, **oota@gin.ec.t.kanazawa-u.ac.jp

Abstract

Semantic object extraction from a video sequence is an indispensable technique in new content-based applications, such as in the international standards MPEG-4 and MPEG-7. In the present paper, we propose a technique that extracts the shape of moving objects from a video sequence with a stationary background by the level set method. In the proposed method, two concepts are incorporated into a novel speed function of the level set method in order to improve convergence and stability. The first concept is the object map, which represents the outline of object regions and the background. The speed function is changed using the object map in order to improve convergence. The second concept is the contour potential energy, which represents the energy in the direction of an object's contour. The efficiency of the proposed method for moving object extraction is demonstrated through computer simulations.

1. Introduction

Semantic object extraction from a video sequence is indispensable in new content-based applications, such as in the MPEG-4 and MPEG-7 international standards. The MPEG-4 multimedia coding standard provides an object-based representation of video data for high-efficiency coding, and the MPEG-7 standard of multimedia content data description handles the video object as meta-data for the description of content.

A number of object extraction techniques, such as chromakey, texture analysis, contour extraction, and contour tracking, have been proposed. However, a high-precision, general purpose technique for moving object extraction has not yet been established. The active contour model algorithm which is a type of contour extraction algorithm that minimizes an energy function, was proposed by Kass *et al.*[1]. This algorithm, which is also referred to as snakes, extracts smooth closed contours stably from an image. However, it may be difficult to adapt snakes to a change in the topology of the contour.

The level set method was proposed by S. Osher *et al.* as a topology free active contour model[2]. The level set method has attracted a great deal of attention, and techniques for improving the calculation cost have been proposed[3][4]. The level set method was applied to various applications, such as three-dimensional (3D) geometric modeling[6], simulation of the pattern formation of crystal[7], and surface editing[8]. Kurazume *et al.* applied the level set method to real-time detecting and tracking of moving objects for surveillance[9]. However, few studies have focused on the precision of moving object extraction from a video sequence by the level set method. Moreover, since the convergence and stability of the level set method depend on the speed function, it is important to define the speed function suitably for the individual application.

In the present paper, we propose a technique that extracts the shapes of moving objects from a video sequence with a stationary background by the level set method. In the proposed method, two concepts are incorporated into a novel speed function of the level set method in order to improve convergence and stability. The first concept is the object map, which represents the outline of object regions and the background, and the speed function is changed using object map to improve convergence. The object map is constructed based on the assumption of Gaussian noise distribution in the frame. The second concept is the contour potential energy, which is defined as the energy in the direction of the object's contour. The contour potential energy causes the contour obtained by the level set method to correspond to the Laplacian zero-crossing of the pixel intensity.

The remainder of the present paper is organized as follows. Section II introduces the level set method, a basic algorithm using the upwind scheme, and the properties of the speed function of the level set method. Detailed descriptions of the object map and the proposed speed function are presented in Section III and Section IV, respectively. Finally, conclusions are presented in Section VI.

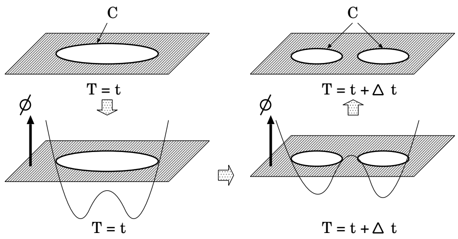


Figure 1. Concept Map of the Level Set Method.

2. Level Set Method

In this section, we describe the basic principles of the level set method. In the level set method, the implicit function ϕ is defined in a space that is one dimension higher than that in which contour C exists at time t. The contour is represented as the cells for which the implicit function has a value of zero. The implicit function ϕ is updated according to a renewal function. We obtain the contour at time $t + \Delta t$ as the cells for which $\phi = 0$ by iteration of the updated process until time $t + \Delta t$. Figure 1 shows the concept map of the level set method.

In the implementation of the level set method for an image, let $C(\mathbf{p}, t)$ be a contour at time t, which denotes the set of points for which $\phi = 0$, where $\mathbf{p}$ is (p_x, p_y).

Let $\mathbf{p}(t)$ be a point on the contour $C(\mathbf{p}, t)$ such that

$$\phi(\mathbf{p}(t), t) = 0. \tag{1}$$

By the chain rule, we have

$$\phi_t + \nabla\phi(\mathbf{p}(t), t)\mathbf{p}_t = 0. \tag{2}$$

The speed function $F(\kappa)$ in the direction $\mathbf{N}$ normal to the contour $C(\mathbf{p}, t)$ is given by

$$F(\kappa) = \mathbf{p}_t \cdot \mathbf{N}, \tag{3}$$

where the normal vector $\mathbf{N}$ on the curve is given by

$$\mathbf{N} = \frac{\nabla\phi}{|\nabla\phi|}, \tag{4}$$

and κ is the local curvature of ϕ. Namely, the evolution equation for ϕ is defined as

$$\begin{aligned} \phi_t + F(\kappa)|\nabla\phi| &= 0 && (5)\\ \phi(C_0(\mathbf{p}), 0) &= 0 \text{ (initial condition)}, && (6) \end{aligned}$$

where C_0 is the initial contour.

The implicit function ϕ is updated iteratively by Eq. (5), and we obtain a new contour satisfying the condition $\phi(x, y, t) = 0$. Thus, the level set method uses an implicit representation to express the contour's movement and is able to handle the topological change of the contour.

2.1. Basic Algorithm of the Level Set Method using the Upwind Scheme

The basic algorithm of the level set method using the upwind scheme[5] is summarized as follows.

1. Initialization

   First, an initial closed contour is set to be equal to the circumference of the image. The value of the implicit function ϕ on an image is approximated for each discrete point. The implicit function for the points on the contour are given as $\phi = 0$, and as the signed Euclidian distance from the contour for the other points.

2. Calculation of Evolution Speed

   The evolution speed $F_{i,j}$ in point (i, j) is calculated.

3. Renewal of the Implicit Function Value

   The value of implicit function is renewed as

$$\begin{aligned} \phi_{i,j} &\leftarrow \phi_{i,j} - F_{i,j}|\nabla\phi_{i,j}|\Delta t \\ &= \phi_{i,j} - \Delta t(max(F_{i,j}, 0)\nabla^+ \\ &\qquad + min(F_{i,j}, 0)\nabla^-), \qquad (7) \end{aligned}$$

   where

$$\begin{aligned} \nabla^+ &= (max(d^{-x}, -d^{+x}, 0)^2 \\ &\qquad + min(d^{-y}, -d^{+y}, 0)^2)^{\frac{1}{2}}, \qquad (8)\\ \nabla^- &= (max(d^{+x}, -d^{-x}, 0)^2 \\ &\qquad + min(d^{+y}, -d^{-y}, 0)^2)^{\frac{1}{2}}, \qquad (9) \end{aligned}$$

$$\begin{aligned} d^{+x} &= \frac{\phi_{i+1,j} - \phi_{i,j}}{h}, \quad d^{+y} = \frac{\phi_{i,j+1} - \phi_{i,j}}{h}, \\ d^{-x} &= \frac{\phi_{i,j} - \phi_{i-1,j}}{h}, \quad d^{-y} = \frac{\phi_{i,j} - \phi_{i,j-1}}{h}, \qquad (10) \end{aligned}$$

   and Δt is the integration interval.

4. Detection of Zero Level Set Cells

   The closed contour at the next time is detected as the cell with the implicit function $\phi = 0$ after the renewal process. For the case in which the implicit function is renewed by the upwind scheme, the integration error increases with the iteration of the renewal process. Therefore, re-initialization in which the signed Euclidian distance from the closed contour is newly set is required at regular intervals.

- Stop Condition

  The maximum number of iterations is decided as the threshold in advance, and the renewal is stopped when the number of iterations exceeds the threshold.

2.2. Properties of the Speed Function

The general form of the speed function of the level set method is

$$F_{i,j} = K_{i,j}(-a - b\kappa_{i,j}), \tag{11}$$

where a and b are positive constants, $\kappa_{i,j}$ is the local curvature of the implicit function ϕ, and the term $K_{i,j}$ is based on image features. The image-based term $K_{i,j}$ is given by

$$K_{i,j} = \frac{1}{1 + |M|^n}, \tag{12}$$

where n is a positive constant.

In Eq. (12), M has a strong influence on the convergence and stability of the extraction. When the evolution speed is fast and the contour moves to inward, the value of M is small, and when the evolution speed becomes nearly zero, the value of M is sufficiently large. Generally, $\nabla^2 G \otimes I_{i,j}$ is used as M for contour extraction from an image. $\nabla^2 G \otimes I_{i,j}$ represents the filtered intensity $I_{i,j}$ obtained by the Laplacian of Gaussian (LoG).

Figure 2 shows the extraction results obtained using $\nabla^2 G \otimes D_{i,j}$ as M for moving object extraction, where $D_{i,j}$ is the frame difference and the gaussian kernel size is 15×15 pixels. The test sequence included Gaussian noise of $N(0, 4)$. In Figure 2, the shapes of the extracted contours are inaccurate due to the influence of the smoothing operator. In addition, the extracted contours did not include the shapes of the feet because of the insufficient frame difference.

Next, Figure 3 shows the locus of the extracted contours in the case of using $D_{i,j}$ as M. The unsmooth locus of the contour is found because of the adverse noise effect. However, the extraction results of Figure 3 provide the relatively precise shapes of the moving objects.

Figure 2. Extraction from the Frame Difference with the LoG Filter.

Figure 3. Locus of Contour obtained by the Speed Function using the Frame Difference.

3. Object Map

In the proposed method, we introduce the object map, which represents the outline of moving object regions and the background, into the speed function of the level set method. The speed function is changed using the object map to improve the convergence.

We assume that a frame of a video sequence includes Gaussian noise with a normal distribution $N(0, \sigma^2)$. From the additivity of normal distributions, the frame difference includes Gaussian noise with $N(0, 2\sigma^2)$. An object map is generated based on this assumption.

3.1. Estimation of the Gaussian Noise Distribution

We roughly divide a frame into moving object regions and the background.

First, the frame difference image was partitioned into 16×16 pixel blocks, and the mean value m_i of the absolute frame difference for each block is calculated. A histogram is constructed for m_i. The threshold TH_m for detecting a block as part of a moving object is determined as the value around the upper tail of the histogram. We test the connectivity of the detected blocks as the moving object part ($m_i \geq TH_m$), and the spatial isolation block is deleted.

Next, a dilation operation with a 3×3 block window is applied to the region of the moving object block. The variance of frame difference in the background block is then calculated, and the obtained variance is used as the estimated variance of the Gaussian noise distribution in the frame difference.

3.2. Pixel Detection in the Noise-only Region

The observed pixel (x, y) is located in either the noise-only region or the other regions by

$$\mathrm{Map}(x, y) = \begin{cases} 0 & \text{if } S^2 \leq 2\sigma^2(1+\alpha)\text{: noise-only region} \\ 1 & \text{if } S^2 > 2\sigma^2(1+\alpha)\text{: other regions} \end{cases} \tag{13}$$

where S^2 is defined as

$$S^2 = \frac{1}{N} \sum_{(i,j)\in\Omega} \left(D_{i,j} - \frac{1}{N} \sum_{(i,j)\in\Omega} D_{i,j} \right)^2 , \tag{14}$$

and S^2 denotes the variance of frame difference in a $N = (2n+1) \times (2n+1)$ square block window centered at the observed pixel and α is determined by the significance level as follows:

$$P_e = \Pr\left[S^2 > 2\sigma^2(1+\alpha)|\text{noise-only region}\right], \tag{15}$$

because $N\frac{S^2}{\sigma^2}$ denotes a χ^2 distribution with $N-1$ degrees of freedom.

3.3. Estimation of Moving Object Regions

The obtained noise-only region (Map = 0) in the frame difference represents the background, and we assume that the pixels in the other regions (Map = 1) belong to moving object regions. However, we regard the small moving object regions as unsemantic objects. The morphological closing operation with 5×5 pixel window is first applied to the moving object regions in Map in order to remove the unsemantic objects. Next, for the case in which the obtained moving object region is smaller than 0.1% of the entire image, the region changes comes to belong to the background.

3.4. Construction of the Object Map

Generally, the contour extracted from the frame difference is not precise because the frame difference includes uncovered-background. Hence, the object map is constructed by Eq. (16) so as to exclude the uncovered-background as follows:

$$\begin{aligned} \text{ObjectMap} = {} & \text{Map}\left[I(t+1) - I(t)\right] \\ & \wedge\ \text{Map}\left[I(t) - I(t-1)\right], \end{aligned} \tag{16}$$

where $\wedge$ denotes logical conjunction and Map$[I(t+1) - I(t)]$ is the Map constructed from the frame difference between frame $I(t+1)$ and frame $I(t)$.

4. Speed Function

4.1. Contour Potential Energy

The contour potential energy is introduced to the speed function of the proposed method, which is defined as the energy in the direction of the object's contour. The contour potential energy causes the contour by the level set method to correspond to the Laplacian zero-crossing of the pixel intensity.

The contour potential energy is calculated as follows:

1. A Gaussian filter is applied to the frame.

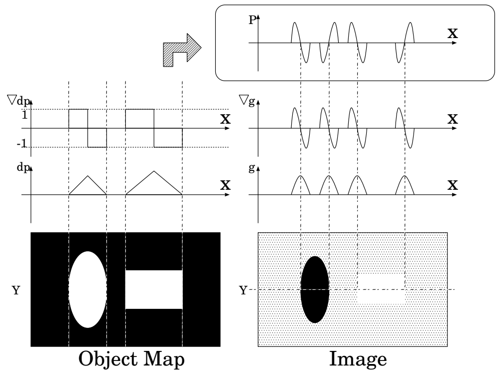


Figure 4. Concept Map of the Contour Potential Energy.

2. A Sobel filter is applied to the obtained smoothed frame. This process provides the gradient of the smoothed frame g.
3. The potential vector d_p at a point in the moving object regions (ObjectMap = 1) is calculated as the distance from the nearest background region (ObjectMap = 0), and d_p of the background region is zero vector.
4. The contour potential energy is given by

$$P = \nabla g \cdot \nabla d_p \tag{17}$$

Figure 4 shows a concept map of contour potential energy.

4.2. Definition of the Novel Speed Function

The novel speed function of the proposed method using the object map and the contour potential energy P is defined as

$$F_{i,j} = \begin{cases} K_{I,i,j}(-a - b\kappa_{i,j}) & \text{if ObjectMap}(i,j) = 0 \\ K_{I,i,j}(-a' - b'\kappa_{i,j}) & \text{if ObjectMap}(i,j) = 1 \\ & \quad \& \ K_{I,i,j} \geq 0 \\ K_{I,i,j}(-a' + b'\kappa_{i,j}) & \text{if ObjectMap}(i,j) = 1 \\ & \quad \& \ K_{I,i,j} < 0, \end{cases} \tag{18}$$

where

$$K_{I,i,j} = \begin{cases} 1 & \text{if ObjectMap}(i,j) = 0 \\ K_d + (1 - K_d)K_v & \text{if ObjectMap}(i,j) = 1, \end{cases} \tag{19}$$

and

$$K_d = \frac{1}{1 + \left(\frac{|D_{i,j}|}{\sigma_D}\right)^n}, \tag{20}$$

$$K_v = max\left(-1.0, min\left(1.0, H_{i,j,\alpha}\right)\right), \tag{21}$$

$$\begin{aligned} H_{i,j,\alpha} = {} & -0.5 + 0.5\,\frac{(P_{i,j} + \alpha)}{\alpha} \\ & + \frac{1}{\pi}\sin\left(\frac{\pi(P_{i,j} + \alpha)}{\alpha}\right). \end{aligned} \tag{22}$$

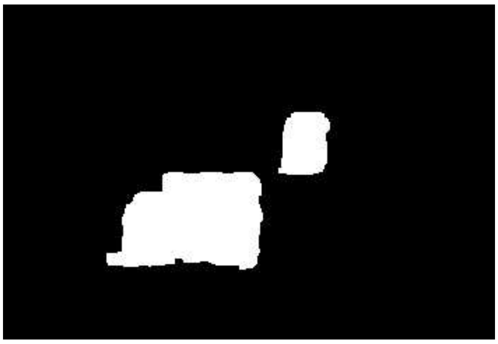


Figure 5. Object Map (Intersection).

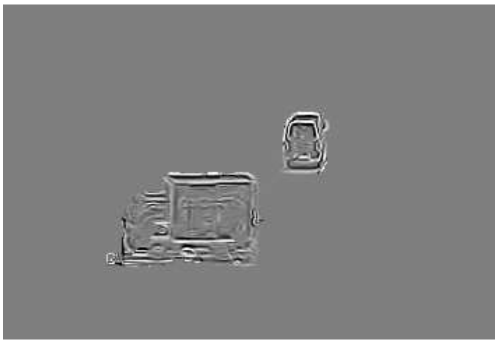
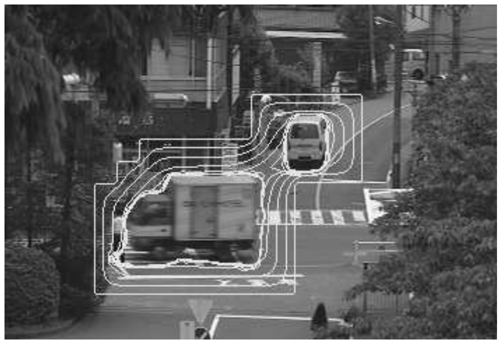


Figure 6. Image Representation of Contour Potential Energy (Intersection).



(a) Intersection

(b) Japanese Room

Figure 7. Locus of the Contour obtained using the Proposed Method.

and a, a', b, b', n and α are positive constants.

The image-based term of the proposed speed function is constructed from the term K_d based on the frame difference and the term K_v based on contour potential energy P, and the function is adaptively changed using the object map.

In other words, the speed function causes the contour to move based on the local curvature term κ in background of the object map, so that the convergence is immune to the influence of noise. In addition, the contour is made to correspond to the zero-crossing point of the Laplacian of the pixel intensity by the action of the term of the contour potential energy in the neighboring boundary of the moving objects. Consequently, the proposed speed function provides precision extraction of the shape of the moving object.

5. Simulation and Results

The proposed moving object extraction was examined by computer simulations. "Intersection" and "Japanese Room" (SIF, grayscale) were used as test sequences. The parameters for the speed function were set as $a = 0.5$, $a' = 0.1$, $b = 2$, $b' = 0.7$, $n = 1$, and $\alpha = 30$.

First, we verified the object map and the contour potential energy. Figure 5 shows the object map of the "Intersection" sequence. From Figure 5, the outlines of the moving object regions were obtained as the object map. Figure 6 shows the image representation of the contour potential energy with an offset of 127. The contour potential energy in the proposed speed function gives the energy in the direction of the zero-crossing points in the image.

Next, we verified the improvement of the stability by the proposed method. Figure 7 shows the locus of the contour ($\phi = 0$) obtained by the proposed method. From Figure 7, the locus of the contour was smooth, so the proposed method improved the stability by avoiding the influence of noise in the background of the object map.

Finally, we evaluated the precision of the moving object extraction by the proposed method. Figure 8 shows the extraction results obtained by the level set method using the LoG filtered frame difference for the purpose of comparison, and Figure 9 shows the extraction results obtained by the proposed method. From the comparison of the results, the proposed method extracts the contour more accurately than by the level set method using the current speed function.

6. Conclusions

In the present paper, we proposed a technique that extracts the shapes of moving objects from a video sequence with a stationary background by the level set method. A speed function with an object map and a contour potential energy was proposed in order to improve the convergence and stability of moving object extraction. The simulation results showed that the proposed method improved the convergence and stability of moving object extraction by the

(a) Intersection

(b) Japanese Room

Figure 8. Extraction Results obtained by the Previous Level Set Method.

(a) Intersection

(b) Japanese Room

Figure 9. Extraction Results obtained using the Proposed Method.

level set method.

In the future, we intend to improve the method used to estimate the Gaussian noise distribution σ^2. In addition, we intend to examine the possibility of adapting the proposed method to the extraction of moving objects from a moving background.

References

[1] M. Kass, A. Witikin, and D. Terzopoulos, "SNAKES: Active Contour Models," *International Journal of Computer Vision*, Vol.1, No.4, pp.321-331, 1988.

[2] S. Osher and J. A. Sethian, "Fronts Propagating with Curvature-dependent Speed: Algorithms Based on Hamilton-Jacobi Formulations," *Journal of Computational Physics*, Vol.79 No.1, pp.12-49, 1988.

[3] Adalsteinsson, D. and J. A. Sethian, "A Fast Level Set Method for Propagating Interfaces," *Journal of Computational Physics*, Vol.118, pp.269-277, 1995.

[4] J. A. Sethian, "Level Set Methods and Fast Marching Methods," *Volume 3 of Cambridge Monographs on Applied and Computational Mathematics*, Cambridge University Press, Cambridge, Second Edition, 1999.

[5] J. A. Sethian, "Level Set Methods" Cambridge University Press, Cambridge, 1996.

[6] C. Baillard, P. Hellier, and C. Barillot, "Cooperation between Level Set Techniques and Dense 3D Registration for the Segmentation of Brain Structure," *International Conference on Pattern Recognition*, Vol.1, pp.991-994, 2000.

[7] Kei, K. and Kokichi, S., "Approximation of Multiplicatively Weighted Crystal Growth Voronoi Diagram and Its Application," *The Transactions of the Institute of Electronics, Information and Communication Engineers. A*, Vol.J83-A, No.12, pp.1495-1504, 2000.

[8] K. Museth, D. Breen, R. T. Whitaker, and A. H. Barr, "Level Set Surface Editing Operators," *ACM Transactions on Graphics*, Vol.21, No.3, ACM SIGGRAPH '02, pp.330-338, 2002.

[9] R. Kurazume, S. Yui, T. Tsuji, Y. Iwashita, K. Hara, and B. Hasegawa, "Fast Level Set Method and Realtime Tracking of Moving Objects in a Sequence of Images," *Transactions of Information Processing Society of Japan*, Vol.44, No.8, pp.2244-2254, 2003.

[10] K. Haris, G. Tziritas, and S. Orphanoudakis, "Smoothing 2-D or 3-D Images Using Local Classification," *Proc. EUSIPCO*, Edinburgh, U.K., Sept. 1994.