Watershed algorithm for moving object extraction considering energy minimization by snakes

WATERSHED ALGORITHM FOR MOVING OBJECT EXTRACTION CONSIDERING ENERGY MINIMIZATION BY SNAKES

Kousuke Imamura, Masaki Hiraoka, Hideo Hashimoto
Kanazawa University, Division of Electrical Engineering and Computer Science
Kakuma-machi, Kanazawa, Ishikawa 920-1192 Japan

Abstract

MPEG-4, which is a video coding standard, supports object-based functionalities for high efficiency coding. MPEG-7, a multimedia content description interface, handles the object data in, for example, retrieval and/or editing systems. Therefore, extraction of semantic video objects is an indispensable tool that benefits these newly developed schemes. In the present paper, we propose a technique that extracts the shape of moving objects by combining snakes and watershed algorithm. The proposed method comprises two steps. In the first step, snakes extract contours of moving objects as a result of the minimization of an energy function. In the second step, the conditional watershed algorithm extracts contours from a topographical surface including a new function term. This function term is introduced to improve the estimated contours considering boundaries of moving objects obtained by snakes. The efficiency of the proposed approach in moving object extraction is demonstrated through computer simulations.

1 Introduction

MPEG-4, which is a video coding standard, supports object-based functionalities for high efficiency coding. MPEG-7, a multimedia content description interface, handles the object data in systems such as retrieval and/or editing systems. Therefore, extraction of semantic video objects is an indispensable tool that benefits these newly developed schemes. Since these standards do not prescribe the technique for object extraction, a number of object extraction techniques, such as chromakey, texture analysis, contour extraction, and contour tracking, have been proposed.

Snakes (active contour models), which are a type of contour extraction algorithm by minimizing an energy function, were proposed by Kass *et al.* [1]. Snakes stably extract smooth closed contours from an image. Hence, this scheme has been used for region extraction and image recognition. A number of attempts have been made to improve the models with respect to the reduction of computational complexity and adaptability to more than one object, for example [2, 3]. In snakes, it may be difficult to set the initial contour and the suitable energy functional for object extraction. In addition, the closed contour is often defined as a set of discrete points for the reduction of noise influence and computational complexity, but the closed contour is not able to accurately represent the true curve. Vieren *et al.* [4] applied snakes to interframe difference images for the contour extraction of moving objects. The problem with this approach is that although it can provide a rough contour, it may not include accurate boundaries of moving objects.

On the other hand, watershed algorithm has been proposed as a technique for region segmentation [5]. Watershed algorithm is a type of region-growing algorithm and treats the input image as a topographic surface. The boundary of segments obtained by watershed algorithm is in accordance with the edge of the object, so we can obtain accurate shape information. However, the influence of noise and the lighting condition lead to over-segmentation. Therefore, a number of preprocessing tasks are required for eliminating the unnecessary edges. Moreover, in the case of moving object extraction, it is difficult to judge whether each region belongs to an object.

New efficiency approaches, which combine snakes and watershed algorithm, were proposed in image segmentation. In [6], the watershed is represented as the energy minimum point. In [7], over-segmentation in watershed algorithm are restrained by using the energy criterion of snakes.

In the present paper, we propose an alternative technique that extracts the shape of moving objects by combining snakes and watershed algorithm. First, snakes extract contours of the moving objects from the interframe difference image as the result of minimization of an energy function. Second, the conditional watershed algorithm extracts edge information from a topographic surface including a new function term. We introduce a new function that incorporates the result of energy minimization by snakes into watershed algorithm. The conditional watershed algorithm extracts one closed contour from each local region.

2 Snakes and watershed algorithm

2.1 Snakes

Snake is represented parametrically by a vector $\mathbf{v}(s) = (x(s),\ y(s))\quad (0 \leq s \leq 1)$ and the shape of the object is extracted by changing the contour through the iterative minimization of the energy. The energy functional of the contour is defined as

$$E_{snakes} = \int_0^1 \{E_{int}(\mathbf{v}(s)) + E_{image}(\mathbf{v}(s)) + E_{con}(\mathbf{v}(s))\} ds, \quad (1)$$

where $E_{int}(\mathbf{v}(s))$ represents the internal energy of the contour due to bending, $E_{image}(\mathbf{v}(s))$ is the image force, and $E_{con}(\mathbf{v}(s))$ is the external constraint.

Snakes proposed by Kass *et al.* are sensitive to noise and minimization of the functional requires computational cost. In order to prevent this problem, Williams *et al.* proposed snakes based on a discrete model for improvement of the noise tolerance and computational complexity. The discrete contour of snakes is represented by control points $\mathbf{v}_i = (x_i, y_i)(i = 1, 2, \cdots, n)$, which are defined in a clockwise manner ($\mathbf{v}_{n+1} = \mathbf{v}_1$). The contour energy in this approach is minimized by a greedy algorithm. In the greedy algorithm, the energy is calculated in the neighborhood of each control point $\mathbf{v}_i$, and the control point $\mathbf{v}_i$ is moved to the minimum energy position. This process is iterated until convergence is attained, and we obtain the final contour.

2.2 Watershed algorithm

Watershed algorithm is a region-growing algorithm and treats the input image as a topographic surface. The luminance gradient is assumed to be the altitude of the topographic surface. The surface is slowly immersed from the minima at the lowest altitude. Dams are erected at locations where the waters coming from two different minima regions merge. The dam corresponds to the border of each region.

3 Moving object extraction algorithm

We describe the proposed moving object extraction algorithm using snakes and watershed algorithm.

3.1 Setting of initial contours

In the case of applying the splitting snakes proposed by Araki *et al.* [3], it is not necessary to prepare initial contours corresponding to the number of objects in advance, and the one initial contour is set on the outer frame of the image. However, the setting of the initial contour on the outer frame involves the problems of the computational costs for convergence and sensitivity to the local minima. In the present paper, we set the initial contours around regions that include moving objects.

Considering the histogram of amount of frame difference for each block, the initial contour setting is performed as follows:

1. The frame difference image is partitioned into 16×16 pixel blocks, and the mean value m for each block is calculated.
2. The occurrence f_m of mean value m is counted in its higher order. And, the threshold TH_m is defined as m at that value the increasing ratio from f_m to f_{m-1} exceeding 10%.
3. If $m_i \geq TH_m$, the block is detected as a part of a moving object.
4. The block detected as the moving object part is tested for its connectivity in a 7×7 block window. For the case in which less than three blocks are connected, this block is deleted through error detection.
5. Dilation operation with a 3×3 block window is applied to the region of the object blocks.
6. The initial control points are set at every eight pixels in the clockwise direction on the outer circumference of the extended region.

3.2 Moving object extraction of snakes

The initial contour converges on the neighborhood of the moving object boundary by energy functional minimization. In the present paper, the energy functions of the snakes for a frame difference image are defined as

$$E_{spline}(\mathbf{v}_i) = \frac{1}{2}\sum_{i=1}^{n}(w_{sp1}|\mathbf{v}_i - \mathbf{v}_{i-1}|^2 + w_{sp2}|\mathbf{v}_{i+1} - 2\mathbf{v}_i + \mathbf{v}_{i-1}|^2), \quad (2)$$

$$E_{area}(\mathbf{v}_i) = \frac{1}{2}\sum_{i=1}^{n} w_{area}[x_i(y_{i+1} - y_i) - (x_{i+1} - x_i)y_i], \quad (3)$$

$$E_{diff}(\mathbf{v}_i) = -\sum_{i=1}^{n} w_{diff}|D(\mathbf{v}_i)|^2, \quad (4)$$

where w_{sp1}, w_{sp2}, w_{area}, and $w_{diff} \geq 0$ are used to balance the relative influence of the terms. The first term of

E_{spline} represents the elasticity of the contour, and the second term represents the stiffness. E_{area} denotes the area energy of the region closed by the contour. These two energies depend on the shape of the contour. In addition, we use the difference energy E_{diff}, which is obtained from frame difference image D. The difference energy causes the contour to converge to the high value of frame difference due to its negative enforcement on the whole.

The contour model is renewed in order to minimize the energy using a greedy algorithm. In the renewal, if the distance between the adjacent control points is more than 10 pixels, then the new control point is inserted midway between these points. In addition, if the distance is less than two pixels, then one of the pixels is deleted. For the case in which the total number of contour models is less than 20, the contour model is deleted as an insignificant object. The renewal process is iterated until the number of moving control points decreases to less than 5% of the initial number.

3.3 Topographic map for watershed algorithm

Unnecessary information, such as that caused by noise and/or local texture, should be removed for region segmentation by watershed algorithm. Thus, we carry out preprocessing in order to obtain the luminance gradient image. This preprocessing is not performed on the entire image, but rather on limited regions, because of the computational costs involved.

In the proposed method, the preprocessing is performed on the inside of the initial contour of snakes because this area includes the target region of watershed algorithm and may have a variable size depending on the object. We describe the procedure used to make the local luminance gradient image on which watershed algorithm is performed. First, a morphological filter [8] smoothes the image while maintaining the edge features. Next, the filtered image is transformed to the luminance gradient image by the multiscale morphological gradient [9]. The morphological reconstruction [10] is applied to the luminance gradient image for the prevention of over-segmentation.

Watershed algorithm of the proposed method employs a function term that is added to the luminance gradient as a topographic map. This term corresponds to distance evaluation between the energy minimum line by snakes and the estimation point. And, the distance evaluation function $d(x)$ is defined as

$$d(x) = e^{-\frac{x^2}{2\sigma^2}} \tag{5}$$

where x is the distance from the contour obtained by snakes, and σ is a positive constant.

As a result, the topographic map T at a point (i, j) is represented as:

$$T(i,j) = \alpha \cdot g(i,j) + (1-\alpha) \cdot g_{max} \cdot d(i,j) \tag{6}$$

where g denotes the luminance gradient, and g_{max} is the highest gradient value. α denotes weighting between the luminance gradient and the distance evaluation, and α is a positive constant in $[0, 1]$.

3.4 Object shape decision by the conditional watershed algorithm

We assume that the obtained energy minimized contour circumferences include the boundary of the moving object, so watershed algorithm extracts this boundary from only the topographic map of the contour circumference. For this purpose, we define watershed areas of width L from the contours obtained by the snakes, and the value of the topographic map in the outer watershed area is changed to zero. However, the plural edges may be extracted from the area by the ordinary watershed algorithm. Therefore, for the case in which watershed area has plural local maxima, the additional condition whereby the maximum among them is regarded as the contour of the moving object is added.

4 Simulation and results

The proposed moving object extraction was examined by computer simulation. “Hall Monitor”,“Bream” and “Japanese Room” (CIF, grayscale) were used as test sequences.

4.1 Setting of the initial contours

We first verify the initial contour setting in snakes. The threshold TH_m for moving object detection is used to judge whether the block is included in the moving object. For the case in which the image includes a high degree of noise, we may need to revise the threshold TH_m.

Figure 1 shows the initial contour of Hall Monitor at $TH_m = 5$. From Figure 1, the initial contour is appropriately placed around the moving objects.

4.2 Energy minimization by snakes

We verified the contour extraction by snakes to the frame difference image. Figure 2 shows the convergent result from the initial contour in Figure 1. The number of iterations until convergence was 59. The number of initial control points was 72, and the number of final control points was same. The weights w_{sp1}, w_{sp2}, w_{diff}, and w_{area} were set to 20.0, 5.0, 1.0, and 24.0, respectively. From Figure 2, the contour of the walker was extracted. However, part of the walker's

Figure 1: Setting of initial contour (Hall Monitor).

Figure 2: Contour extraction by snakes.

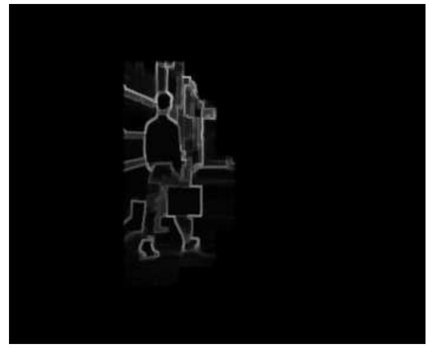


Figure 3: Local luminance gradient image.

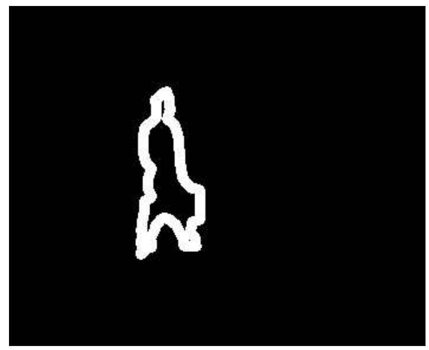


Figure 4: Watershed area (L=9).

leg was not extracted properly because its movement was not as great.

4.3 Topographic map in watershed area

Next, we made a topographic map for the conditional watershed algorithm. Figure 3 shows the image obtained by morphological reconstruction after multiscale morphological gradient estimation and morphological filtering. From Figure 3, a luminance gradient image enhancing the contour with little influence of noise was obtained.

Watershed algorithm extracts the contour from the watershed area around the contour obtained by snakes. Figure 4 shows the watershed area with the expanding width of $L = 9$.

4.4 Contour decision by watershed algorithm

Finally, we verified that the contour was obtained by the proposed method.

Figures 5 through 8 show the effectiveness of the new topographic function T with weight α in Eq. 6. From these figures, as α increases, the extracted contour gradually becomes close to the conditional watershed contours (α = 1.0). The extracted contours with α less than 0.8 has good smoothness and the lost walker's leg can be partly recovered. In addition, the contour of the walker's head is extracted without wrong notches.

Figures 9 and 10 show the results of contour extraction for other test sequences, Bream and Japanese Room, respectively. Comparing these results, the proposed method extracts the contour more accurately than by snakes and more smoothly than by watershed. In particular, the right hand of the lady in the Japanese Room is improved.

5 Conclusion

In the present paper, we proposed a technique for motion object extraction combining snakes and watershed algorithm. The simulation results show that the proposed method provides accurate moving object extraction. As a result, we have confirmed the possibility of the novel moving object extraction method combining snakes and watershed algorithm. We will examine the possibility of adapting the proposed method to the extraction of moving objects from a moving background.

Figure 5: Results of contour extraction ($\alpha = 0.2$).

Figure 6: Results of contour extraction ($\alpha = 0.4$).

Figure 7: Results of contour extraction ($\alpha = 0.8$).

Figure 8: Results of contour extraction ($\alpha = 1.0$).

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] D.J. Williams and M. Shah, "A Fast Algorithm for Active Contours," *Proc. of 3rd ICCV*, pp. 592-595, 1990.

[3] S. Araki, N. Yokoya, H. Iwasa and H. Takemura, "Splitting Active Contour Models Based on Crossing Detection for Extracting Multiple Objects," *IEICE Trans. on Information and Systems*, vol. 179-D-II, no. 10, pp. 1704-1711, 1996

[4] C. Vieren, F. Cabestaing and J.-G. Postaire, "Catching moving objects with snakes for motion tracking," *Pattern Recognition Letters 16*, Elsevier, pp. 679-685, 1995.

[5] L. Vincent and P. Soille, "Watersheds in Digital Spaces: An Efficient Algorithm Based on Immersion Simulations," *IEEE Trans. Pattern Analysys and Machine Intelligence*, vol. 13, no. 6, pp. 583-598, 1991.

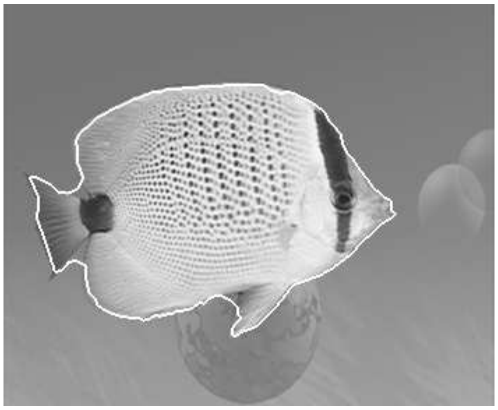
[6] J. Park and J.M. Keller, "Snakes on the Watershed," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 23, no. 10, pp. 1201-1205, 2001.

[7] H.T. Nguyen, M. Worring and R.van den Boomgaard, "Watersnakes: energy-driven watershed segmentation," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 25, no. 3, pp. 330-342, 2003.

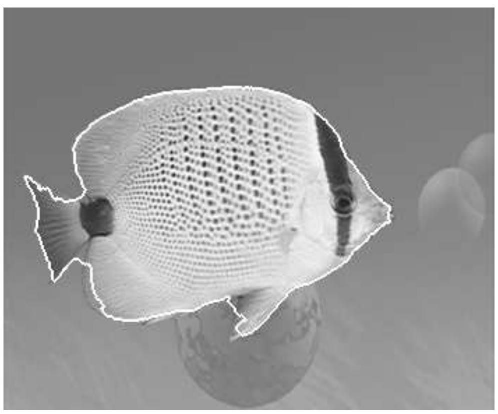
[8] D. Cortez et al., "Image Segmentation Towards New Image Representation Methods," *Signal Processing: Image Communication*, vol. 6, pp. 485-498, 1995.

[9] D. Wang, "A Multiscale Gradient Algorithm for Image Segmentation Using Waterhseds," *Pattern Recognition*, vol. 30, no. 12, pp. 2043-2052, 1997.

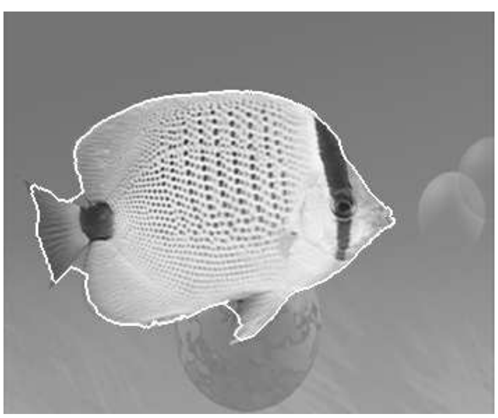
[10] L. Vincent, "Morphological Grayscale Reconstruction in Image Analysis: Applications and Efficient Algorithm," *IEEE Trans. Image Processing*, vol. 2, no. 2, pp. 177-201, 1993.



(a) Snakes (α = 0.0),



(b) Watershed (α = 1.0),



(c) Proposed Method (α = 0.8)

Figure 9: Results of contour extraction (Bream).

(a) Snakes (α = 0.0),

(b) Watershed (α = 1.0),

(c) Proposed Method (α = 0.8)

Figure 10: Results of contour extraction (Japanese Room).