A New Video Object Segmentation Algorithm using the Morphological Technique

Jen-Shiun Chiang, Chih-Yueh Chang, Chih-Hsien Hsia, and Wei-Hsuan Chang
Multimedia IC Design Lab., Department of Electrical Engineering
Tamkang UniversityEmail: chiang@ee.tku.edu.tw; chyuchang@ee.tku.edu.tw; chhsia@ee.tku.edu.tw; whchang@ee.tku.edu.tw

Abstract

In this paper, we propose a new video object segmentation algorithm using the morphological technique. Several video object segmentation algorithms use mathematical morphology to generate the object masks, however the operations of the mathematical morphology have two drawbacks: (1) *high computation complexity and* (2) *the quality of the object masks depends on the chosen morphological structuring element. There are many techniques to speed up the morphological operations by hardware implementation, but without discussions about reducing the influence of the choice of the structuring elements. By adding a pre-processing mechanism, the proposed algorithm effectively reduce the influences of the chosen structuring elements based on continuity of shape features and times of morphological operations. Experimental results show that our algorithm can improve the speed of filling operations of the object masks and accuracy of segmentation.*

Keywords: Edge detection, mathematical morphology, video object segmentation

1. Introduction

As the techniques of storage media and internet are developed rapidly in recent years, the demands for video sequences are not only quality but also interaction; the conventional video coding standard, such as MPEG-1, MPEG-2, and H.263 cannot satisfy such demands. MPEG-4 [1] video coding standard is the first to support randomly accessing video objects by the concept of video-object-plane (VOP). It can support high interaction and more flexible video coding. Therefore, to segment the shapes of the video objects is very important, and many video object segmentation algorithms have been presented [2]-[6]. They can be summarily classified into two types: (1) temporal analysis and (2) temporal-spatial analysis. One typical kind of temporal analysis is to find the moving information of the video objects between two continuous frames on the temporal domain [7]. Although these methods have lower computation complexity, the quality of segmentation is not good enough. The objects needed to be segmented can be easily influenced by their brightness or shadow, and therefore some kind of typical methods combining spatial and temporal domain was proposed [14]. This approach uses the algorithms of edge detection to find the shapes of the foreground objects and then detect the edge of the moving regions by using the analysis on the temporal axis, and then the filling technique is used to generate the masks of the foreground objects. Although this method has higher robustness and better quality for segmentation, the similarity between the original video objects and masks depends on the chosen morphological structuring element and times of processing during the process of filling masks by using morphological operations. If a complex structuring element is chosen, it will raise the computation complexity and reduce the efficiency. In order to overcome the mentioned shortcomings, we propose a new algorithm by adding a pre-processing mechanism to improve the object segmentation.

The rest of this paper is organized as follows. Section 2 briefly describes the concepts of mathematical morphology. Section 3 illustrates the proposed algorithm. The experimental results and comparisons are described in section 4. Finally a briefly conclusion is given in Section 5.

2. Mathematical Morphology

Mathematical morphology is stemmed from set theory. It is a powerful technique in many fields, especially for image processing [11]-[12]. The aims of the mathematical morphology are to analyze the shapes by using the information of the objects. The principle of the mathematical morphology can extract the related structure by probing the input video frame with a set called structuring elements. There are several operators in the mathematical morphology, such as dilation, erosion, opening, and closing. The goal is to extract the related structures of the image. Let us consider the 2-D gray-tone signals with the input signal f and structuring element B as illustrated in Fig. 1. The dilation operation is denoted by "$\oplus$" and is defined as:

$$f \oplus B = \max_{b \in B} f(x + b) \tag{1}$$

The dilation operation is graphically illustrated in Fig. 2. The erosion operation is denoted by "e " and is defined as:

$$f \text{ e } B = \max_{b \in B} f(x - b) \tag{2}$$

The erosion operation is graphically illustrated in Fig. 3. The opening operation is denoted by "o" and is defined as:

$$f \circ B = (f \ominus B) \oplus B \tag{3}$$

The closing operation is denoted by " $\bullet$ " and is defined as:

$$f \bullet B = (f \oplus B) \ominus B \tag{4}$$

3. Overview of the Proposed Algorithm

Hsiao *et al.* [8] and Bosworth *et al.* [11] proposed the mathematical morphology techniques to segment video object effectively. However, their approaches take a lot of computation efforts to achieve the demand of efficiency. To overcome the problems of high computation complexity and the choice of structuring element, we propose a new object segmentation algorithm by adding a mask pre-filling (MPF) mechanism before the operations of the mathematical morphology. It can effectively reduce the influences of the chosen structuring elements based on the continuity of the shape features and times of the morphological operations.

Our proposed algorithm is based on the fixed background and camera. The proposed new video object segmentation includes the following steps: (1) edge detection, (2) frame difference, (3) mask pre-processing, (4) morphological filling processing, and (5) extracting video objects. The details of each step of the proposed algorithm are described in the following subsections and the block diagram of the algorithm is illustrated in Fig. 4.

3.1 Edge Detection

Several edge detection methods have been proposed, such as Sobel and Canny edge detection algorithms [7]; these methods can find good detecting results of edges, but the operation is too complicated. Here we use the morphology gradient filter to find the edges of the foreground objects in the video frame. By using the morphology gradient filter, we can find edges without complex operation, and it can also cancel the shadow effect. The operations can be described by the following equation:

$$E_G = (f \oplus B) - (f \ominus B), \tag{7}$$

where E_G is the results after passing the morphology gradient filter. Figure 5(a) is a video frame and Fig. 5(b) shows the detected edges of Fig. 5(a) by the proposed approach.

3.2 Frame Difference

We first calculate the difference between the current frame and the previous frame to find the moving information. Then, the region is divided into foreground and background parts by using a threshold value. The detail is shown in the following equations:

$$FD_{edge}(x, y, t) = | I_G(x, y, t) - I_G(x, y, t-1) | \tag{8}$$

$$F(x, y, t) = \begin{cases} 255 & \text{,if FGD} \geq \text{Th} \\ 0 & \text{,if FGD} < \text{Th} \end{cases} \tag{9}$$

3.3 Mask Pre-Processing

According to the continuity of the object feature, we can extend the information of the foreground by the relationship between the current pixel and its upper left 4 neighbors. The processing steps of the pre-processing are summarized as follows:

Step 1: If the value of the current pixel is 255 and the value of the upper-left neighbor pixel is also 255, then the value of the lower-right neighbor pixel is 255. The procedure is illustrated in Fig. 6(a).

Step2: If the value of the current pixel is 255 and the value of the upper neighbor pixel is also 255, then the value of the lower neighbor pixel is 255. This procedure is illustrated in Fig. 6(b).

Step3: If the value of the current pixel is 255 and the value of the upper-right neighbor pixel is also 255, then the value of the lower-left neighbor pixel is 255. This is illustrated in Fig. 6(c).

Step4: If the value of the current pixel is 255 and the value of the right neighbor pixel is also 255, then the value of the left neighbor pixel is 255. This procedure is illustrated in Fig. 6(d).

The processing result after mask pre-processing is shown in Fig.7.

3.4 Morphological Filling Processing

After pre-processing, we use the opening-operator and closing-operator of the mathematical morphology by a 3×3 square structuring element to fill the mask. The equations of the opening-operator and closing-operator are as follows:

$$f \circ B = (f \ominus B) \oplus B \tag{10}$$

$$f \bullet B = (f \oplus B) \ominus B \tag{11}$$

Figure 8(a) shows the result after the morphological filling processing. By using the final object mask of Fig. 8(a), the video object can be extracted as shown in Fig. 8(b).

4. EXPERIMENTAL RESULTS

We use the *Weather* video sequence [16] for simulation. The experimental results of the proposed algorithm indicate that the new approach not only can reduce the times of the operations but also keep good accuracy. Figure 9 is the segmentation results of the *Weather* video sequence and *Akiyo* video sequence. Figures 10 and 11 show the accuracy of the results. Our approach has the accuracy more than 96% of the *Weather* video sequence on average, and more than 90% accuracy of the *Akiyo* video sequence without complex operations.

Table 1 indicates the decision times between the methods with pre-filling processing and without pre-filling processing under the condition of using a 3×3 square structuring element.

The performance of the proposed algorithm is much better than the conventional approach.

5. CONCLUSIONS

For the real-time and interaction demands, the efficiency of video object segmentation is a very important issue for the content-based applications. In this work, an efficient video object segmentation based on mask pre-filling (MPF) algorithm is proposed. We try to develop a new idea of video object segmentation using MPF mechanism. The MPF mechanism can reduce the influence of the structuring element and speed up the processing operation without complex computation. The experimental results indicate that the proposed approach can improve the efficiency and keep good segmentation quality as well.

6. REFERENCES

[1] T. Sikora, "The MPEG-4 video standard verification model," *IEEE Transactions on Circuits and Systems for Video Technology*, vol.1, pp. 19-31, February 1997.

[2] R. V. Babu, K. R. Ramakrishnan, and S. H. Srinivasan, "Video object segmentation: a compressed domain approach," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 14, no. 4, pp. 462-474, April 2004.

[3] L. Yang, H. Zhang, and Y. Li, "Automatic segmentation of moving objects in video sequences", *Journal of Computer-Aided Design and Computer Graphics*, vol. 16, no.3, pp. 597-612, July 2004.

[4] S.-Y.Chien, Y.-W. Huang, and L.-G. Chen, "Predictive watershed: A fast watershed algorithm for video segmentation," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 13, no. 5, pp. 453-461, May 2003.

[5] Y. Tsaig and A. Averbuch, "Automatic segmentation of moving objects in video sequences: a region labeling approach," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 12, no. 7, pp. 597-612, July 2002.

[6] C. Kim and J.-N. Hwang, "Fast and robust moving object segmentation in video sequences," *International Conference on Image Processing*, vol. 2, pp. 131-134, 1999.

[7] M. Tekalp, *Digital Video Processing*, Prentice Hall PTR, 1995.

[8] Y.-T. Hsiao, C.-L. Chuang, S.-H. Yen, and H. -J. Lin, "A mathematical morphological approach to thin edge detection in dark region," *IEEE International Symposium on Signal Processing and Information Technology*, pp. 310-313, December 2004.

[9] J. Serra, Image Analysis and Mathematical Morphology, New York: Academic. 1982.

[10] P. Soille, Morphological Image Analysis: Principles and Applications, Springer: Berlin Heidelberg, 1999.

[11] J. H. Bosworth and S. T. Acton, "Morphological image segmentation by local monotonicity," *IEEE Conference on Signals, Systems, and Computers*, vol. 1, pp. 53-57, October 1999.

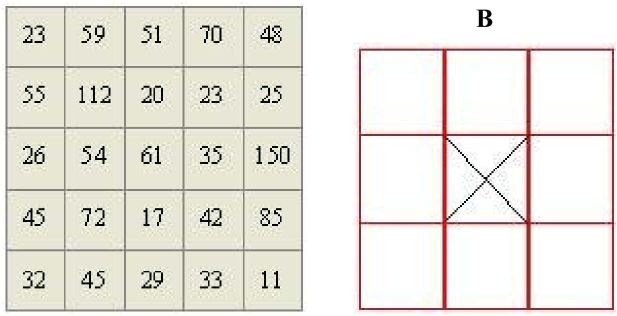
[12] S.-C. Cheng, "Region-growing approach to colour segmentation using 3-D clustering and relaxation labeling," *IET proceedings- Visual Image Signal Processing*, vol. 150, no. 4, pp. 270-276, August 2003.

[13] Petrou and P. Bosdogianni, *Image Processing*: *The Fundamentals*, John Wiley & Sons, 1999.

[14] C. Kim and J.-N. Hwang, "Fast and automatic video object segmentation and tracking for content-based applications," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 12, no.2, pp. 122-129, February 2002.

[15] Neri, S. Colonnese, G. Russo, and P. Talone, "Automatic moving object and background separation," *Signal Processing, Special Issue on Video Sequence Segmentation for Content-based Processing and Manipulation, Elsevier*, vol. 66, no. 2, pp. 129-132, April 1998.

[16] Test Sequences: http://media.xiph.org/video/derf.



(a) (b)

Fig. 1. (a) Image array *f*. (b) 3×3 square structuring element *B*.

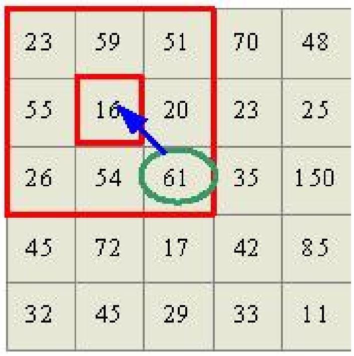


Figure 2. Operations of morphology dilation $f \oplus B$.

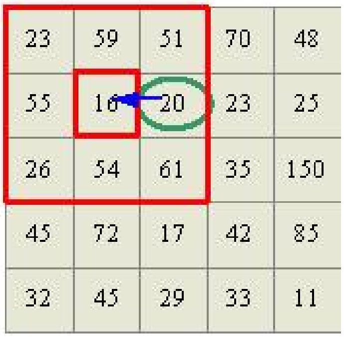


Figure 3. Operations of morphology dilation $f \oplus B$.

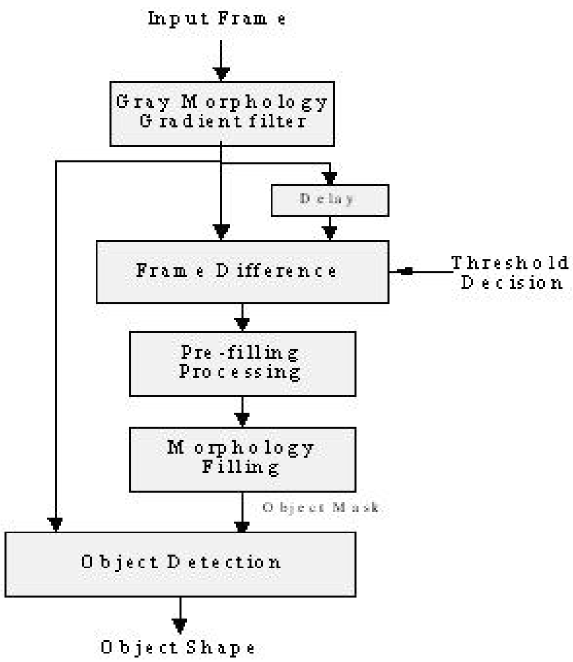


Figure 4. The proposed algorithm.

(a) (b)

Figure 5. (a) Original frame, (b) the detected edges by using the morphology gradient filter.

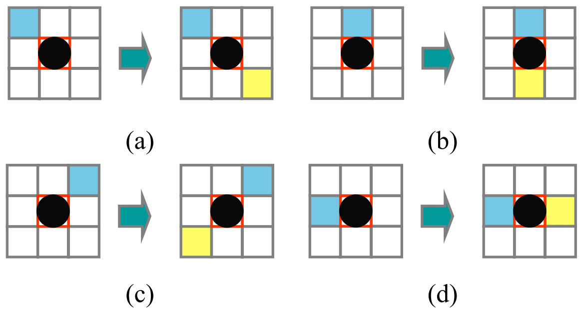


Figure 6. The extending rule of the pre-processing step: (a) Calculated by the upper left neighbor, (b) Calculated by upper neighbor, (c) Calculated by the upper right neighbor, (d) Calculated by the right neighbor.

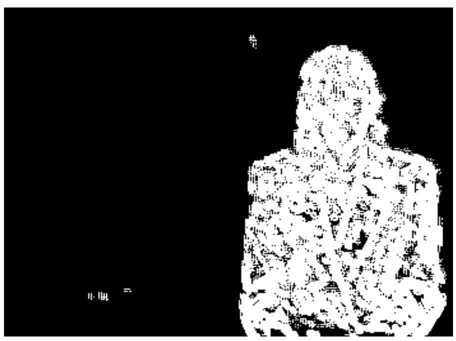


Figure 7. The results after the process of mask pre-filling (MPF).

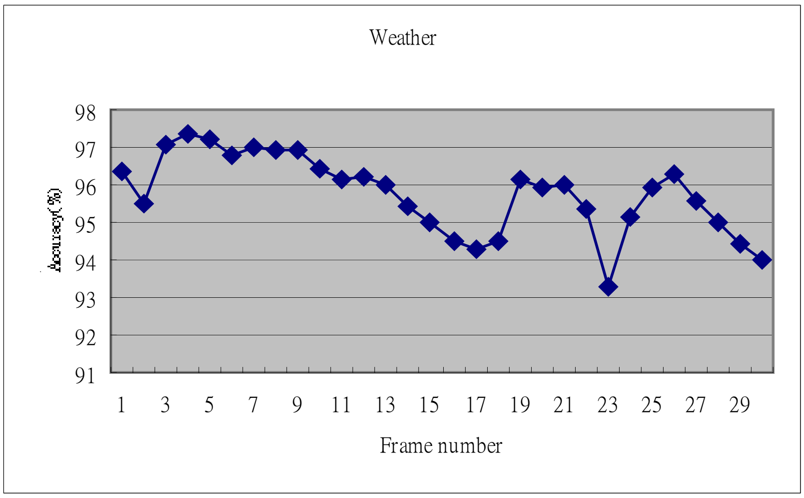
(a) (b)

Figure 8. (a) The final object mask. (b) The result of the object extraction.

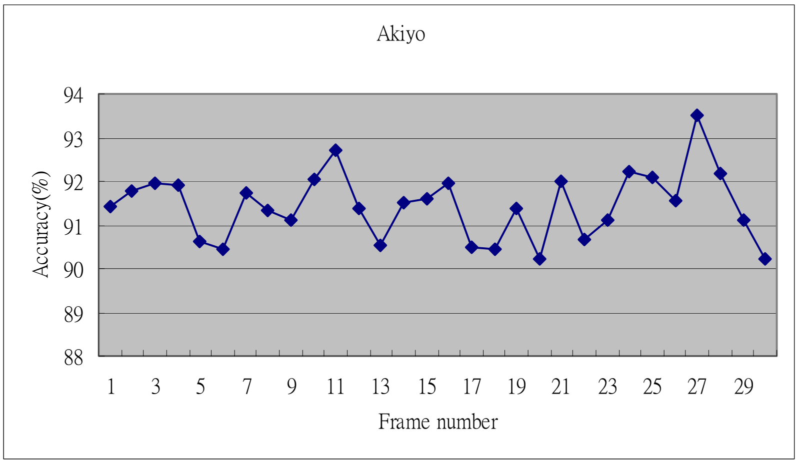
(a) (b)

(c) (d)

Figure 9. The results of Weather video sequence: (a) Frame#4, (b) Frame#16,
The results of Akiyo video sequence: (c) Frame#6, (d) Frame#12.



(a)



(b)

Figure 10. (a) The curve of spatail accuracy for *Weather*, (b) The spatail accuracy for *Akiyo*.

TABLE 1. DECISION TIMES

Case	Condition	
	Without Pre-Filling	With Pre-Filling
Opening	3 times	1 time
Closing	3 times	1times
Total times of decision for each pixel	72 times	41 times