

# Video Object Extraction Integrating Temporal-Spatial Information

Shiping Zhu<sup>a</sup>, Jie Gao<sup>b</sup>

Department of Measurement Control and Information Technology, School of Instrumentation  
Science and Optoelectronics Engineering, Beihang University

Beijing 100191, China.

<sup>a</sup>spzhu@163.com, <sup>b</sup>happyjiejie@yahoo.cn

**Keywords:** video object segmentation;  $t$ -distribution significance test; Sobel; temporal-spatial filter

**Abstract.** A novel video object segmentation method combines change detection and edge detection is proposed here. The process of the algorithm can be divided into three parts: motion detection, spatial segmentation and temporal-spatial filter, which integrates the spatial and temporal information of the video sequence. The motion detection step makes use of the  $t$ -distribution significance test; then, the initial motion detection mask of every frame in the video sequence can be integrated to form the movement mask. For spatial segmentation the Sobel edge detection operator is used to get the boundary of the video objects in current frame; temporal-spatial filter then integrates the temporal and spatial information, extracts the precise boundary of moving object and also reduces the residual noise. Finally, the segmentation of video objects can be optimized by filling and morphology operation.

## Introduction

The standard MPEG-4 focuses on improving the video coding efficiency and content-based interactivity. The later provides new functions to the application of multimedia. Video object segmentation is a process that involves partitioning a video frame into a set of meaningful objects or regions i.e. VOP (video object plane). Thus the scene is composed from independently encoded video objects. Video object segmentation is a hot spot of the research [1]. Recently, video object segmentation has been widely used. Its applications are extended to video retrieval, video index, visual communication, remote sensing, virtual reality, etc.

Algorithms of video object segmentation can be mainly classified into two categories: temporal segmentation algorithm and spatial segmentation algorithm. Temporal algorithms [2-3] utilize the temporal properties of the video sequence. It detects the temporal changes to get the moving objects. Only considering the temporal information is impossible to extract the precise contour of the object. Spatial segmentation algorithms [4] utilize the color, light, texture and edge information to obtain the video object. Compared to the former algorithm, spatial segmentation algorithms can get the accurate contour information. Taking the edge detection algorithm as an example, although the accurate edge of the object can be obtained, the edge of the background can't be neglected too. This brings much interference for the segmentation. No matter which algorithm we choose, the occlusion problem and irregular movement can cause the drop of the segmentation precision.

In order to overcome the occlusion problem and irregular movement, in this paper, we propose a novel video object segmentation algorithm, which makes full use of temporal and spatial information, and obtains a satisfactory segmentation result.

## Principle of the Proposed Algorithm

The proposed algorithm utilizes three symmetrical frames at a distance of  $k$  for change detection. Firstly, get the frame difference image. Taking  $t$ -distribution significance test to detect the change between frames; then accumulate frame differences of a fixed period of time; Sobel edge detection operator is used to detect the accurate boundary of all the video objects in current frame; Temporal-

spatial filter combines the temporal and spatial information to extract the complete and precise boundary of moving objects. Finally, the segmentation of video objects can be obtained by filling and doing morphology operations. The flowchart of the proposed algorithm is shown as in Fig. 1.

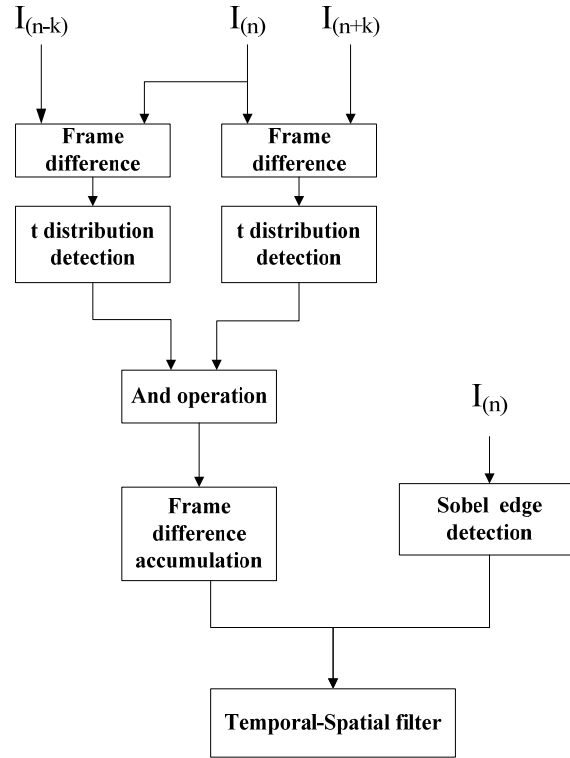


Fig. 1 Flowchart of the proposed algorithm

### *t*-distribution Significance Test

Through the analysis of the image, set the unchanged pixel in the frame difference image to be 0. Do *t*-distribution significance test for each of the pixel to decide whether it belongs to the video object or not.

For each pixel  $(x, y)_n$  that to be tested, centering at this pixel and applying *t*-distribution significance test is better than just tests a single pixel. When constructing the statistics of test, the neighborhood window is centered at  $(x, y)_n$  and the size of the window is  $(2n+1) \times (2n+1)$ . We can use a  $3 \times 3$  or  $5 \times 5$  neighborhood window. We construct statistics in this window and applying *t*-distribution significance test.

Since the noises distribute in every single frame, modeling the noise as Gaussian noise. Mark the noise in  $F_{(n)}$  as  $n_n(x, y)$  and mark the variance as  $\delta_c^2$ . The image of the frame difference is:

$$I_d(x, y) = | \bar{I}_n(x, y) - \bar{I}_{(n-k)}(x, y) + n_n(x, y) - n_{(n-k)}(x, y) |. \quad (1)$$

In which  $\bar{I}_n(x, y)$  is the actual value.  $d_{(x,y)} = n_n(x, y) - n_{(n-k)}(x, y)$  is the noise in the frame difference image.  $n_n(x, y)$  and  $n_{(n-k)}(x, y)$  have the same probability density, and they are mutual independence, so  $d_{(x,y)}$  still belongs to Gaussian noise which has a zero mean and a variance  $\delta^2 = 2 \times \delta_c^2$ .

If all the nonzero frame differences are caused by the noise, the mean  $\mu$  should all be 0. Then Hypothesis Test is taken, make the null Hypothesis as:  $H_0: \mu = 0$ . According to the pixel in the neighborhood window, constructing statistics for *t* distribution detection:

$$t = \frac{f_d(n)}{s / \sqrt{p}}. \quad (2)$$

$$\text{In which, } f_d(n) = \frac{1}{p} \sum_{-n}^n \sum_{-n}^n |I_d(x+l, y+m)|, \quad s = \sqrt{\frac{1}{p-1} \sum_{-n}^n \sum_{-n}^n (I_d(x+l, y+m) - f_d(n))^2}.$$

According to the theory of significance, the threshold is decided by the distribution of significance level  $\alpha$  and the  $t$  distribution:

$$|t| \geq t_{\frac{\alpha}{2}}(p-1). \quad (3)$$

The choice of significance level  $\alpha$  is relative to the camera noise, it is usually set to be  $10^{-2}$ ,  $10^{-6}$  and so on. If equation (3) is true, the center pixel belongs to  $t(n)$ , which means the pixel belongs to the moving region, the initial moving region template can be described as:

$$T_{n,n-k}(x, y) = \begin{cases} 255, & (x, y)_n \in t(n) \\ 0, & \text{others} \end{cases}. \quad (4)$$

After the integrating operation between symmetrical moving region mask, the unobvious part of the video object can be included, thus the complete moving region mask is:

$$T_n(x, y) = \begin{cases} 255, & T_{n, n-k}(x, y) = 1 \\ & \cap T_{n, n-k}(x, y) = 1. \\ 0, & \text{others} \end{cases}. \quad (5)$$

### Frame Difference Accumulation during a Time Period

For the irregular movement of the video object, we can't get the whole video object precisely. In order to solve the irregular movement problem, the proposed method which accumulates the frame differences during a period of time, it can solve the loss of the contour information [5]. This method utilizes the relevance between frames and the continuity of the motion then calculates the frequency of each pixel. The pixel with the highest frequency can be included in the effective mask.

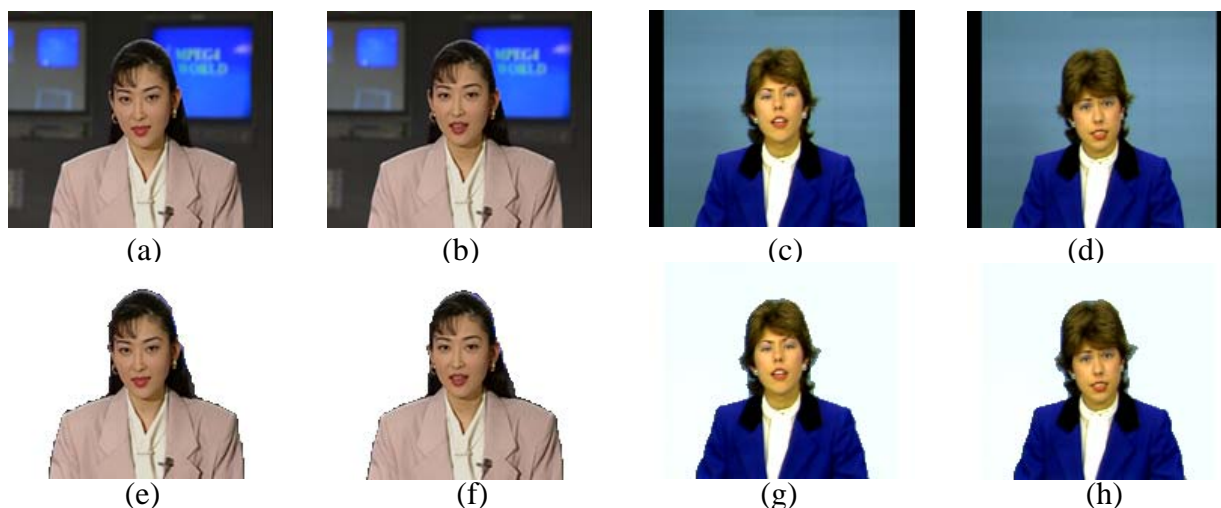
If the count of  $(x, y)$  marked as the moving point in the time period is larger than  $\tau$ , the corresponding pixel belongs to the effective template.  $\tau$  is a thresh hold value. For each pixel  $(x, y)$ . If the corresponding value in effective mask is 0, we'll pass the accumulating step, if the corresponding value in effective mask is 255, then we'll accumulate the frame differences.

### The Generation of Video Object

After the last step, although the effect is improved significantly, there are holes in the video object still. We propose memory mask (MM), which means filling and doing morphology processing to get the whole mask. In order to get a more precise result, we do Sobel edge detection [4] for each frame and get the boundary  $E_s$ , then we blend the MM and  $E_s$  together and threshold the result  $B(x, y)$ . If the corresponding value of  $B(x, y)$  is 255, this pixel will finally be marked as foreground. Then combine the initial video sequence  $O(x, y)$  to finish the segmentation.

### The Result of the Algorithm

In order to show the advantages of the proposed algorithm, we choose two QCIF video sequences (Akiyo and Claire) to do the experiment. The upper row shows the original video sequence and the lower row shows the corresponding segmentation results after using our algorithm. The experimental results are shown as in Fig. 2.



(a) Akiyo 11<sup>#</sup>; (b) Akiyo 22<sup>#</sup>; (c) Claire 12<sup>#</sup>; (d) Claire 39<sup>#</sup>; (e) VOP of Akiyo 11<sup>#</sup>; (f) VOP of Akiyo 22<sup>#</sup>; (g) VOP of Claire 12<sup>#</sup>; (h) VOP of Claire 39<sup>#</sup>.

Fig. 2 VOPs of the video sequences extracted by the proposed algorithm.

## Summary

Based on the traditional spatial-temporal algorithm, a new algorithm utilizes change detection and edge detection is proposed here. It integrates temporal-spatial information and can correspond to the fast changing environment. Experimental results indicate the correctness of the proposed algorithm.

## Acknowledgment

The project is sponsored by the National Natural Science Foundation of China (NSFC) under grants No. 61075011 and No. 60675018, also the Scientific Research Foundation for the Returned Overseas Chinese Scholars from the State Education Ministry of China.

## References

- [1] Bugeau A, Perez P.: Detection and Segmentation of Moving Objects in Highly Dynamic Scenes, IEEE Conference on Computer Vision and Pattern Recognition, (2007), p. 1-8.
- [2] Zhang Xiao-bo, Liu Wen-yao: A New Video Segmentation Method Based on Modified Watershed Combined with Temporal Information, Chinese Journal of Sensors and Actuators, Vol. 20 (2007), p. 2248-2251.
- [3] Shao-Yi Chien, Shyh-Yih Ma, Liang-Gee Chen: Efficient Moving Object Segmentation Algorithm Using Background Registration Technique, IEEE Transactions on Image Processing, Vol. 12 (2002), p. 577-586.
- [4] Gao Wen-shuo, Yang Lei, Zhang Xiao-guang, Liu Hui-zhong: An Improved Sobel Edge Detection, IEEE International Conference on Computer Science and Information Technology, Vol. 5 (2010), p. 67-71.
- [5] Zhu Shi-ping, Lin Yun-yu, Zhang Qing-rong: A Modified Spatial-temporal Video Object Segmentation Algorithm Based on Region Compensation in Fixed Time Interval, Journal of Optoelectronics Laser, Vol. 20 (2009), p. 1376-1380.