Watershed 変換を用いた動物体抽出に関する検討
A Study on Extraction of Moving Objects Using Watershed Transform

渡辺　成光 [*1]　　秦泉寺　久美 [*2]　　渡辺　裕 [*1]

Shigemitsu WATANABE [*1], Kumi JINZENJI [*2], Hiroshi WATANABE [*1]

[*1] 早稲田大学大学院　国際情報通信研究科

[*1] Graduate School of Global Info. and Telecom. Studies, Waseda University

[*2] NTT サイバースペース研究所

[*2] NTT Cyber Space Labs, NTT Corporation,

1. INTRODUCTION

MPEG-4 provides an efficient coding tool, "sprite coding." In sprite coding, a "sprite" i.e. a unified panoramic background image derived from a sequence having a camera motion, and a foreground object are utilized as VOP's. Hence, sprite coding offers high compression since number of frames can be represented by a single (panoramic) still image. An algorithm that automatically generates the background sprite and the foreground objects, called "Two-layer VOP generation scheme," has been proposed [1-2]. However, this method approximates the foreground object by using macroblocks. For this reason, unnecessary background regions are extracted as foreground, and the composed image may include visual discrepancy. Thus, a method to extract the foreground objects on a pixel base is necessary. We propose an algorithm that automatically extracts the moving objects on a pixel base using background difference, which is the difference between the original image and the background image, and watershed transform. Results given by our proposed method are more than twice as better than that of the conventional method.

2. OBJECT EXTRACTION ALGORITHM

2.1. Extraction of Foreground Candidate Image by Background Difference

First, we calculate the difference image between the original image and the background image. The background image is extracted from the background sprite that is generated by the conventional method. However, since the background image is extracted using the global motion, it does not always coincide with the original image if the image contains complicated background. It is rather shifted by a few pixels from the original image. For this reason, we use a window of certain size to scan the background image, and calculate the differences between the target pixel of the original image and the pixels of the background image within the window. Then, we define the minimum of the differences as the target pixel value. The resulting difference image becomes the foreground candidate image.

2.2. Image Segmentation by Watershed Transformation

In order to extract the foreground object accurately from the foreground candidate image, image segmentation such that the boundary of the segmented region coincides with the contour of the object must be needed. To satisfy this condition, we use the well-known image segmentation tool, "watershed transform" [3]. The concept of the watershed transform is to consider an image as a topographic surface, flood the surface from each minimum, and define watersheds to be the merging points of the water flooding from different minima. Thus, the input image of the watershed transform should be the gradient image of the image from which the contour of the object is extracted (we call this an initial image). We use the foreground candidate image generated by the background difference as the initial image. However, watershed transform tends to over-segment mainly due to noise. In order to avoid this problem, we generate a simplified mosaic image of the original image. The mosaic image can be generated in the following way. First, we calculate the watershed of the gradient image. Secondly, we fill the region between the watersheds with the gray value in the original image corresponding to the local minima of the gradient image. From this mosaic image, we generate a valued graph by defining boundaries between two regions of mosaic image as the gray tone difference between these regions. Then, deleting the boundaries less than some threshold gives the better watershed image.

2.3. Masking Process Using foreground Macroblock image

The conventional method generates a mask image of the foreground object that is approximated by macroblocks (Fig. 2b). We use this macroblock mask image to mask the watershed image obtained in Section 2.2. By this operation, the unnecessary edges obtained by watershed transformation can be deleted, leaving only the edges around the objects we want to extract.

2.4. Generation of the Object Mask Image by the Morphological Closing Process

In order to generate the final object image, we apply morphological closing process to the masked watershed image generated in section 2.3. Mathematical morphology is based on two operations: erosion and dilation. Closing operator is derived from these fundamental operations of erosion and dilation: it is simply a dilation followed by erosion. It has an effect of filling in particular background regions of the image such as gaps and holes. By applying this morphological closing operator to the watershed image where foreground is the contour of the object, we fill in the regions inside the contour, and hence, generate the final object mask image.

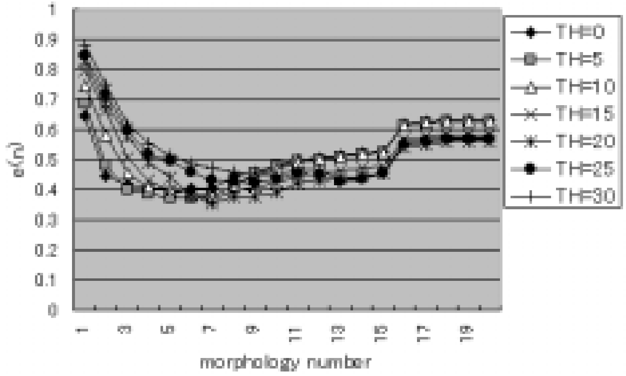


Fig. 1: Parameter Characteristic (Stefan, 100th frame)

3. DETERMINATION OF OPTIMUM PARAMETERS

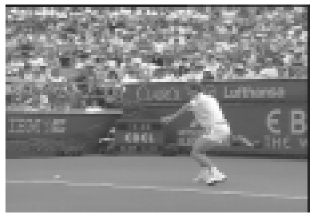
In our proposed method, we use two parameters: the threshold to delete the unnecessary boundaries generated by watershed transform, and the number of morphological operations to fill in the regions inside the contour of the object when generating the final object mask image. Both of these parameters affect the resulting images. In order to determine the parameters that give the optimum results, we investigated the correlation between the parameters and the value calculated by equation (1) which determines how much the resulting object image differs from the correct segmentation image.

$$e(n) = A_d(n) / A_s(n), \qquad (1)$$

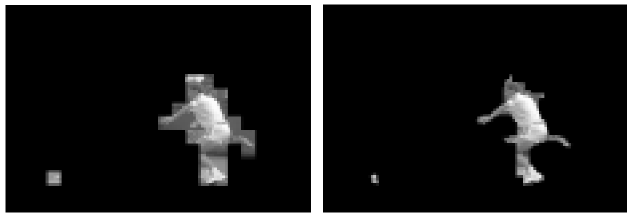
where n is the frame number, A_d is the area of regions that are different from the correct segmentation mask, and A_s is the area of the correct segmentation mask. We examined 20 randomly selected frames. In each of these frames, thresholds and the number of morphological operations are tested through 0 to 30, and 1 to 20, respectively. One example of the resulting graph is shown in Fig. 1. By investigating these graphs, we determined that the robust optimum results could be obtained when threshold is between 15-25, and the number of morphological operation is between 6-8. Therefore, we set the threshold to be 20, and the number of morphological operation to be 7.

4. RESULTS AND EVALUATION

The original image and the object image generated by the conventional method and by our proposed method are shown in Fig. 2. The method we are proposing has been successful in improving the object extraction accuracy. In order to evaluate the results numerically, we used the video sequence "Stefan" which has the correct object segmentation mask. We evaluated our results by equation (1). Comparison of the proposed method with the conventional macroblock based method is shown in Table 1. It shows that the results given by our proposed method are more than twice as better than that of the conventional method.



(a) Original Image



(b) Conventional Method (c) Proposed Method

Fig.2: Original and Object Images (Stefan, 100th frame)

Table 1: Comparison with the conventional method

Frame No	020	040	060	080	100
MB-based	0.672	0.642	1.384	1.190	0.932
Proposed	0.360	0.259	0.359	0.219	0.382

5. CONCLUSION

We proposed an algorithm to extract the moving objects automatically from the video sequence for sprite coding. The proposed method utilizes background difference to extract the foreground candidate image first. Secondly, it uses the watershed transform to extract the contour of the object in order to achieve higher accuracy. Then, it utilizes the macroblock image of foreground objects, which is generated by the conventional method, to delete unnecessary edges. Finally, by operating the morphological closing process, final object mask image is generated. The object image generated by our proposed method is more than twice as better as the conventional method.

5. REFERENCES

[1] H. Watanabe, and K. Jinzenji, "Sprite Coding in Object-based Video Coding Standard: MPEG-4," World Multiconference on SCI 2001 Proc. Vol. XIII, pp.420-425, July 2001.

[2] K. Jinzenji, H. Watanabe, S. Okada, and N. Kobayashi, "Very Low Bit-Rate Video Compression Using MPEG-4 Sprite Coding," IEICE Transaction, Vol.J84-D-II, No.5, pp.758-768, May 2001.

[3] S. Beucher, "The Watershed Transformation Applied to Image Segmentation", 10th Pfefferkorn Conf. on Signal and Image Processing in Microscopy and Microanalysis, 16-19 sept. 1991.

[4] S. Sakaida, M. Naemura, and Y. Kanatsugu, "Moving Object Extraction Using Background Difference and Region Growing with Spatio-Temporal Watersheds," IEICE Transaction, Vol.J84-D-II, No.12, pp.2541-2555, December 2001.