OBJECT EXTRACTION COMBINING IMAGE PARTITION WITH MOTION DETECTION

Wenming Yang[1], Wang Lu[2], Naitong Zhang[1]

1 Communication Research Center, Harbin Institute of Technology, China
2 Department of Automation,Tsinghua University, China

ABSTRACT

We present a video object extraction algorithm combining image partition with motion detection. It consists of three stages: spatial image partition(SIP), temporal motion detection (TMD) and spatiotemporal projection (STP). Firstly, we partition the current video frame into a series of image regions by a modified watershed transform with double open-closing reconstruction and non-linear pixel classification, and the corresponding region boundaries are achieved. In the next stage, global motion estimation and compensation (GME&GMC) are performed, then motion detection based on Gaussianity test is applied to pixels on region boundaries, not to all pixels belonging to regions. In final stage, motion mask is projected to the current frame to segmenting video objects. Preliminary simulation results demonstrate the performance of the proposed algorithm.

Index Terms— object extraction, image partition, motion detection, region boundary, non-linear pixel classification

1. INTRODUCTION

Video object extraction has been an area of intensive research in the field of computer vision. With the advent of emerging multimedia standards like MPEG-4 it has become even more critical to develop a system that performs video applications in a robust as well as computationally efficient manner. Applications of object segmentation range from video surveillance, video compression, video retrieval video summarization, video indexing, etc[1],[2]. Some of these have been manifested in the new multimedia standards including MPEG-4 and MPEG-7[3].

A variety of techniques have been employed for segmenting a semantically meaningful object out of video sequences. The most common approaches that have been proposed fall into the following categories: Motion field based, data clustering based, temporal tracking based and change detection based segmenting.

Approaches based on motion field can be used but it is noise sensitive and computation is expensive. Data clustering methods are iterative and require exhaustive computation[4]. The third approach, temporal tracking, is to combine a technique for single image segmentation with a temporal tracking procedure[5]. Unfortunately, single image segmentation is itself a very difficult problem, so it is not easy to attain accurate video object defination. In addition, partial occlusion is a main problem confronted by temporal tracking algorithm. In this paper, a robust video object extraction algorithm is developed with the emphasis placed on image partition and motion dectection.

2. ALGORITHM OVERVIEW

The block diagram of the proposed video segmentation algorithm is shown in Fig.1. SIP aims at achieving a series of image region with texture homogeneity and boundary information of current video frame. After SIP, TMD was performed on the current video frame and previous video frame to attain motion mask. It contains global motion estimation and compensation(GME&GMC). Furtherly, motion mask is projected to image regions of current video frame to extract video object. Novel and pragmatic idea is introduced in SIP and TMD, respectively.

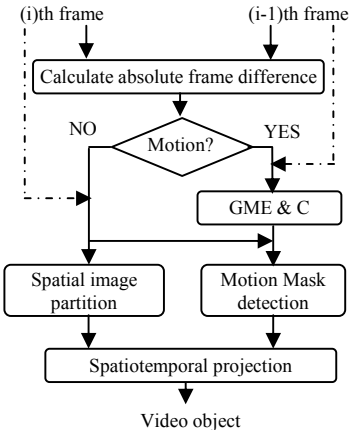


Fig.1. Block diagram of the proposed algorithm

3. SPATIAL IMAGE PARTITION

Watershed transform is a well established tool for the image partition[6]. However, over-segmentation, a major problem with the watershed transform, always remains unsolved. Region merging seems to be a preferable way for dealing with over-segmentation[7]. However, region merging is computation expensive and a good merging measure is quite difficult to obtain. We consider that noise elimination before

performing watershed is the root key to over-segmentation in that noise sensitivity is primary cause of over-segmentation in watershed transform. Hence we introduce an improved watershed transform scheme with a series of pre-processing procedures to eliminate noise in original video frame.

The block diagram of the proposed Watershed transform scheme is shown in Fig.2. It contains three main parts: double open-closing reconstruction, non-linear pixel classification, Vincent watershed labeling.

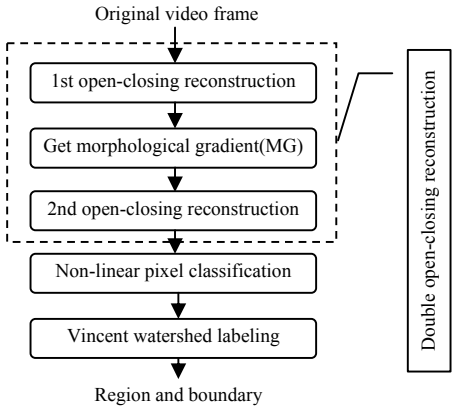


Fig.2. Block diagram of the proposed watershed scheme

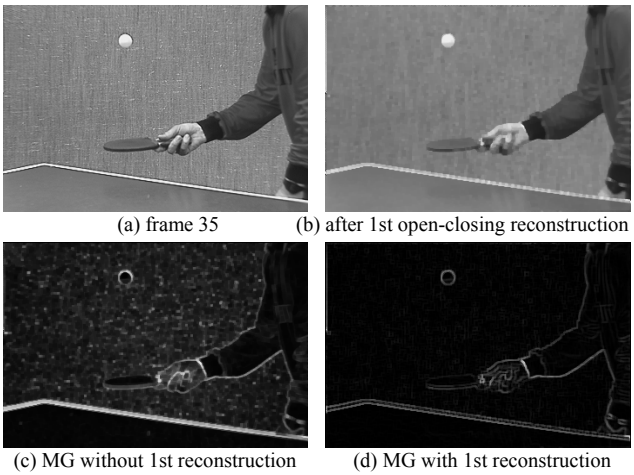
3.1. Double Open-closing Reconstruction

Morphological open-closing reconstruction is an effective pre-processing way to remove bright and shade features[8], which can reduce influence of noise on following watershed labeling. Our method named double open-closing reconstruction is different from others in that open-closing reconstruction is not only used to derive smooth MG but used in original video frame.

In general, noise and signal in a gray level image are independent from each other, and they have different distribution with regard to both size(or area) and gray level. Certainly, pixels belonging to noise are relatively dispersed in 2-D space and gray while those belonging to signal are opposite. Strictly, it can be concluded as follows.

1) *Disparity in size:* Size of signal is larger than noise's in limited local regions of image.
2) *Disparity in gray level:* Their gray levels are different in limited local regions of image.

Considering two disparities and properties of open-closing reconstruction, we can infer that extremal(maximum or minimum) operation in open-closing reconstruction should eliminate noise from disparity in gray level; meanwhile, adjustment of size of structure element should remove noise with smaller size while signal with larger size should be reserved. So we insert 1st open-closing reconstruction before calculating MG. The result of applying 1st open-closing reconstruction on tennis sequence is shown in Fig.3.



(a) frame 35 (b) after 1st open-closing reconstruction

(c) MG without 1st reconstruction (d) MG with 1st reconstruction

Fig.3. Double open-closing reconstruction for tennis

We see that noise is eliminated after 1st open-closing reconstruction, which is quite helpful for reducing amount of region extremal that contributes to over-segmentation of watershed transform.

3.2.Non-linear pixel classification

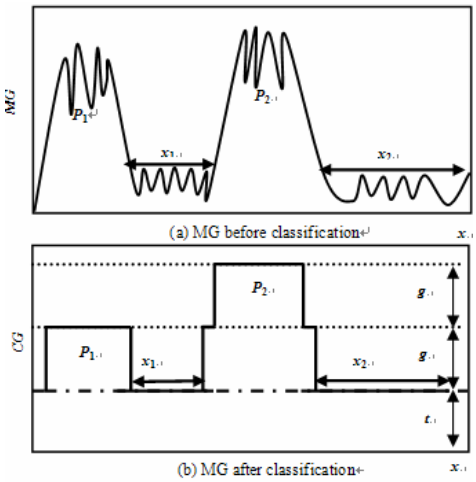
In the reconstructed MG image, pixels belonging to the same region maybe different in MG, especially, when thin and close textures exist in original image in that they can result in high gradient level. To debase influence of textures, we introduce a non-linear pixel classification defined as follows. *INT* () is the round function; *t*, *g* are two parameters between 0 and 255.

$$CG = \begin{cases} INT((MG - t)/g) \cdot g + t & if \quad MG > t \\ t, & otherwise \end{cases} \qquad (1)$$

Using (1), we are able to restrain noise in reconstructed MG. Fig.4. illustrates an example of non-linear pixel classification. The x-axis indicates spatial neighboring relationship of pixels. The smaller the Euclid distance of two pixel in image is, the closer they are in x-axis. It can be found that *t* is relative to eliminate noise with low gradient; and g is for overcoming noise caused by thin and close textures with high gradient inside plat regions. Further, we can infer that influence of close textures on region partition will be overcome only if parameters *t* and *g* are chosen properly. It is a typical instance as shown in Fig.4. for good parameter choice.

3.3.Vincent watershed labeling

Vincent watershed labeling algorithm, the well-known algorithm, is deemed as the fastest watershed transform[9], which is employed for region labelling. After all pixels are labeled, the image is partitioned into many homogeneous closed regions with different labels.

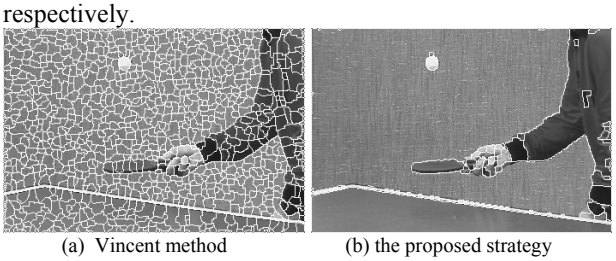


(a) MG before classification

(b) MG after classification

Fig.4. CG after non-linear pixel classification

To attain regions and region boundaries mask, we perform image scan once more, place watersheds into regions, and mark the boundaries of every region. We labeled regions as $R(i)$, and marked boundary of $R(i)$ as $B(i)$. Fig.5. shows the results of applying Vincent watershed method and the proposed scheme on tennis sequence, respectively.



(a) Vincent method (b) the proposed strategy

Fig.5. Compared results for Vincent and our method

4.TEMPORAL MOTION DETECTION

In this stage, the first step is to calculate average absolute difference of two consecutive video frame, which is used to determine whether GME&GMC should be introduced. If the average absolute difference is greater than threshold A given, GME&GMC will be performed. The affine motion model in GME is defined by

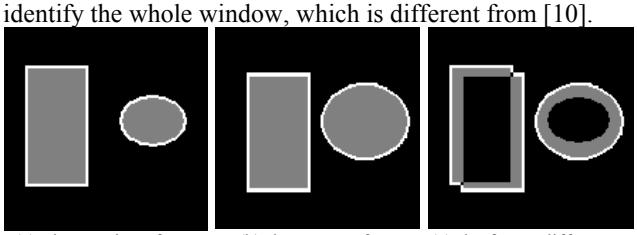
$$\begin{bmatrix} u' \\ v' \\ 1 \end{bmatrix} = \begin{bmatrix} k_0 & k_1 & k_2 \\ k_3 & k_4 & k_5 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} \qquad (2)$$

Where k_j, j = 0,1…5 are the model parameters. The transformation displaces a point(u, v, i+1) in the reference frame to a new location(u', v', i) in the previous frame. A least squares and Levenberg-Marquardt based algorithm is used to extract the motion model parameters. Bilinear interpolation is employed in GMC.

Once pixels' gray inside of region is homogenous, absolute frame difference will fail. In Fig.6, there are one rectangular region and one circle region in video sequence, which are both homogeneous in gray level. In consecutive frames, the rectangular region has a motion of translation while the circle one has a deformation as shown in Fig.6.(a) and Fig.6.(b). On one hand, we find that a lot of pixels with zero gray level appears inside motion region, and the cause is region homogeneity in gray level. So we denote zero difference inside motion region as pseudo-zero frame difference. On the other hand, high gray level remains particularly in region boundaries in frame difference. Two features are not difficult to come by for pixels on region boundary in moving procedure as follows.

1) *Place's Change:* some pixels on region boundary are bound to leave their original places.
2) *Overlapping probability:* pixels on boundary have a smaller overlapping probability than those inside region.

Depending on the analysis above, we proposed a motion detection algorithm based on pixels on region boundary, with conventional algorithm based on all pixels in region avoided. For the same window consisting of $P\times Q$ pixels to be computed, our Gaussianity test is used to identify whether or not centered pixel is moving, but not to identify the whole window, which is different from [10].



(a) the previous frame (b) the current frame (c) the frame difference

Fig.6. Frame difference of region translation&deformation

Let $D(x,y,i)$ be frame difference between two consecutive image in video sequences, measure of Gaussianity test $H(i)$ is defined as follows.

$$H(i)=I_4(i)+I_3(i)-3I_1(i)(I_2(i)-I_1^2(i))-3I_2^2(i)-I_1^3(i)-2I_1^4(i) \qquad (3)$$

where $$I_k(i)=1/(P\times Q)\sum_{x=1}^{P}\sum_{y=1}^{Q}D^k(x,y,i) \qquad (4)$$

Let $M(x,y,i)$ denote motion mask obtained by gaussianity test model, it can be formulated as the following equation:

$$M(x,y,i)=\begin{cases} 0 & |H(i)|\le G \\ 255 & |H(i)|> G \end{cases} \qquad (5)$$

In (5), G is a threshold close to zero.

5. SPATIOTEMPORAL PROJECTION

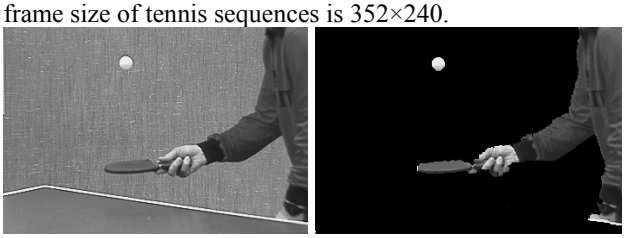
According to the steps above, we have obtained mask M containing moving pixels on region boundaries B. Subsequently, we can get the amount of moving pixels of every $B(i)$, let the amount be $C(i)$, and denote amount of pixels of $B(i)$ as $S(i)$. Using (6), we can attain motion regions mask.

$$R(i)=\begin{cases}255 & (C(i)/S(i)\geq T_s \;\&\; (x,y)\in B(i)\\ 0 & (C(i)/S(i)\geq T_s \;\&\; (x,y)\in B(i)\end{cases} \qquad (6)$$

Where T_s is a threshold between 0 and 1. 255 represents that $R(i)$ is moving. Therefore, video object can be achieved by projecting moving $R(i)$ into image regions of current video frame.

6. SIMULATED EXPERIMENTAL RESULTS

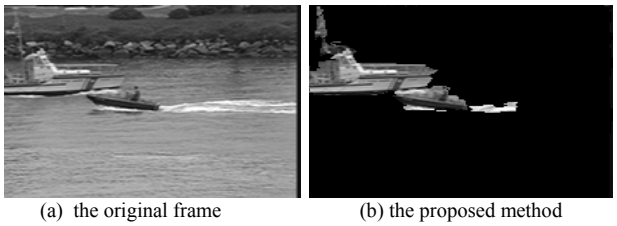
Standard MPEG-4 testing sequences have been used to test our algorithm. Fig.7. illustrates the final result of segmentation for one video frame of tennis sequence. The frame size of tennis sequences is 352×240.



(a) the original frame (b) the proposed method

Fig.7. the segmentation results for 35 frame of tennis

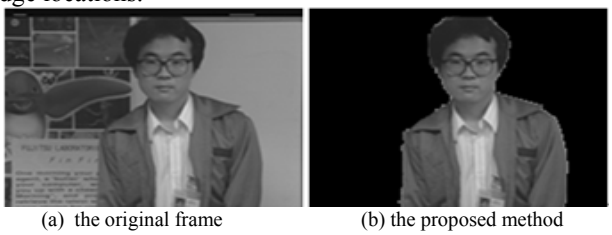
The results of testing sequence "coastguard" are shown in Fig.8. The frame size of coastguard sequences is 176×144. We see that video objects, two ships and human, are extracted and motion background is eliminated despite a few parts of foreground and background being misclassified.



(a) the original frame (b) the proposed method

Fig.8. the compared results for 75 frame of coastguard

Fig.9. illustrates the final result of segmentation for one video frame of bowing sequence. Visually, the proposed algorithm gives a reliable and intact video object with exact edge locations.



(a) the original frame (b) the proposed method

Fig.9. the segmentation results for one frame of bowing

7.CONCLUSION

In our work, we emphasized the image partition as the initial step. Based on the segmented homogeneous regions and boundaries, motion parts can be detected by Gaussianity test on boundary pixels. Projection is used to segment video objects. Our SIP algorithm gives a reliable result with a smaller number of segments without employment of region merging, which illustrates over-segmentation in watershed is restrained. In addition, intact motion parts are obtained, which proves our motion detection method based on region boundary is capable of overcoming influence of pseudo-zero frame difference. The segmented video objects show that our approach is promising.

We found that, in our experiments, it is possible for our approach to misclassify part of background as video object, especially when the part is surrounded by video objects. To overcome this problem, more complicated techniques and further study are required.

8.REFERENCES

[1] C. Kim and J. N. Hwang, "Fast and automatic video object segmentation and tracking for content-based application," *IEEE Trans. Circuit Syst. Video Technol.*, vol.12, no.2, pp.122-129, Feb. 2002.

[2] A. Prati, I. Mikic, M. Trivedi, and R. Cucchiara, "Detecting moving shadows, algorithms, and evaluation," *IEEE Trans. Pattern Anal. Machine Intell.*, vol.25, no.7, pp.918-923, Jul. 2003.

[3] I. Kompatsiarisb and M. G. Strintzis, "Spatiotemporal segmentation and tracking of objects for visualization of videoconference image sequences," *IEEE Trans. Circuit Syst. Video Technol.*, vol.10, no.8, pp.1388-1402, Dec. 2000.

[4] B. G. Kim and D. J. Park, "Novel noncontrast-based edge descriptor for image segmentation," *IEEE Trans. Circuit Syst. Video Technol.*, vol.16, no.9, pp.1086-1095, Sep. 2006.

[5] K. Hariharakrishnan and D. Schonfeld, "Fast object tracking using adaptive block matching," *IEEE Trans. Multimedia*, vol.7, no.5, pp.853-859, Oct. 2005.

[6] L. Vincent and P. Sollie, "Watershed in digital spaces: an efficient algorithm based on immersion simulations," *IEEE Trans. Pattern Anal. Machine Intell.*, vol.13, no.6, pp.583-598, Jun. 1991.

[7] K. Haris, SN. Efstratiadis, etal, "Hybrid image segmentation using watersheds and fast region merging," *IEEE Trans. Image Proc.*, vol.7, no.12, pp.1684-1699, Dec. 1998.

[8] S. Mukhopadhyay and B. Chanda, "Multiscale morphological segmentation of gray-scale images," *IEEE Trans. Image Proc.*, vol.12, no.5, pp.533-549, May. 2003.

[9] S. Y. Chien, Y. W. Huang and L. G. Chen, "Predictive watershed: a fast watershed algorithm for video segmentation," *IEEE Trans. Circuit Syst. Video Technol.*, vol.13, no.5, pp.453-461, May. 2003.

[10] S. Y. Chien, Y. W. Huang and B. Y. Hsieh, etal, "Fast video segmentation algorithm with shadow cancellation, global motion compensation, and adaptive threshold techniques," IEEE Trans. Multimedia, vol.6, no.5, pp.732-748, Oct. 2005.