VIDEO OBJECT SEGMENTATION BASED ON ACCUMULATIVE FRAME DIFFERENCE

Bing Leng   *Qionghai Dai*

Broadband Network & Digital Media Lab of Dept. Automation in Tsinghua University, 100084, Beijing, P.R.China

ABSTRACT

This paper addresses the problem of extracting video objects from head-shoulder video sequences. A method based on accumulative frame difference is proposed. First, a preliminary motion analysis is performed to each block of the frame and the blocks with fast moving edges are detected. Then, for each block, we accumulate frame difference with a different amount of frames, based on its motion attributes. After thresholding and post processing, the objects are obtained. Experimental results demonstrate that the proposed method can eliminate the expanded changed region and thus achieves a significantly improved segmentation result.

Index Terms—Video object segmentation, accumulative frame difference, motion analysis.

1. INTRODUCTION

Extracting moving objects from video sequences is of widespread interests and a fundamental step in a large number of applications, such as object-based coding and content-based applications. Plentiful research work has been devoted to this subject. However there has not been a universal method so far, which can achieve satisfying results in all the various situations.

As a certain category of video sequence, head-shoulder videos have their particularities that make some of the methods unable to give convincing results. First, the objects usually cover a large percentage of the area in a frame and the movement region is restricted. So it is difficult to reconstruct the background picture, and therefore unable to perform the algorithm based on background registration[1,2]. Secondly, the objects are non-rigid and different body parts might have remarkably different motion. As a sequence, motion estimation and object tracking will be difficult or very complicated.

Change detection is widely used in video processing and analyzing [3,4,5,6], and it is effective and efficient in the segmentation of head-shoulder video sequences. Traditional change detection algorithms are based on the evaluation of the frame difference of two successive frames. However, these methods suffer from the fact that noise pixels in the video might be mistaken as the object. Also, those low motion parts of the object might be determined as background. Methods with accumulative frame difference (AFD)[6] can avoid such shortcomings. Nonetheless, by accumulating the frame difference, some uncovered background parts are also included into the object region.

In this paper, we propose a new method based on accumulative frame difference to improve segmentation quality. Our method includes the following innovations: first of all, frame difference accumulation is not carried out on the whole frame, but on square blocks, in order to utilize more accurate localized information. Secondly, we perform a motion analysis on each block so that we can calculate an adaptive number of frames to accumulate frame difference. The rest of this paper is organized as follows: section 2 gives a short recall of the accumulative frame difference method. In section 3 our approach is presented. The experimental results are given in section 4 and conclusion is made in section 5.

2. ACCUMULATIVE FRAME DIFFERENCE

Change detection methods separate the temporally changed and unchanged regions of two successive images based on the evaluation of the frame difference(FD):

$$FD_k(x,y) = \left| I_k(x,y) - I_{k-1}(x,y) \right| \qquad (1)$$

By thresholding FD, we can get a change mask. But this mask might include some noise pixels and exclude the regions with low motion or low texture, as shown in Fig.1(b). So further information is needed to make a more accurate determination.

One approach is to use the temporal information of a longer period by accumulating frame difference of the past N frames together:

$$AFD_k(x,y) = \sum_{i=k}^{k-N+1} FD_i(x,y) \qquad (2)$$

(a) (b) (c)

Fig. 1. Effect of FD and AFD (a)original frame (b) FD (c)AFD of 32 frames

As shown in Fig.1(c), in the AFD picture, the random background noise is limited. Also the low motion parts of the object, such as the daughter's head, can be separate from the background.

This work is supported by Distinguished Young Scholars of NSFC No.60525111 and key project of NSFC No.60432030.

However, such accumulation will also expand the actual changed region when the object is moving fast. The regions that the object passes by in the past several frames will be accumulated to the current frame. Such as in Fig.1(c), the fast movement of mother's head leaves a much larger changed region in the AFD picture. To obtain an accurate segmentation result, these expanded changed regions require further processing.

3. PROPOSED APPROACH

Our approach can be described in four major steps: 1). Frame motion analysis: a preliminary analysis on the motion attributes of each block in the current frame. 2). Frame difference accumulation: we calculate the number of frames to accumulate for each block and accumulate the frame difference. 3). Thresholding: we calculate the threshold value then threshold the accumulative frame difference. 4).post processing: morphological operations are performed to eliminate small noise regions and region growing is done to obtain more accurate object boundary.

3.1. Frame motion analysis

First, frame difference is obtained by Eq.(1) in section 2 and the current frame difference picture is divided into square blocks (8×8 pixels in our implement). Then we calculate the sum of FD within a block as the criterion of motion analysis:

$$SBD_k(i,j) = \sum_{(x,y)\in B(i,j)} FD_k(x,y) \qquad (3)$$

where B(i,j) is the block with the coordinate (i,j).

Fig2 illustrates SBD of four blocks in the sequence *mother and daughter*. The mother moves her head rapidly from the right of the picture to the left as shown in Fig2(a) and (b). The SBD curves of those four blocks are shown in Fig2(c). We can identify different types of motion in the block from the shape of its SBD curve:

A: the SBD curve changes slowly. It corresponds to the blocks that are inner parts of a slowly moving object.

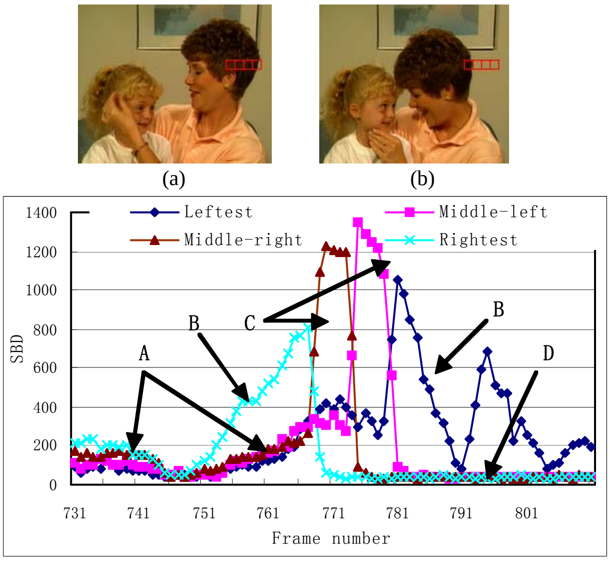
B: the curve reaches a very high value, but changes relatively slowly. It corresponds to the blocks containing accelerating or decelerating moving edges.

C: the curve increases and decreases very fast and on the peak there are several points with approximately the same value. This kind of curve corresponds to the blocks where fast moving edges move in and out, with uniform velocity.

D: the curve remains at a low value. It corresponds to the blocks with still areas.

After an object leaves an area, the frame difference in this area before the object leaves becomes useless and should not be accumulated because it belongs to a different situation. So among the above four motion types, it is significant to identify type B and C which contain fast moving edges. According to the shape characters of curve type B and C, if SBD of one block is big enough:

$$SBD_k(i,j) > \lambda S_{n.\max} \qquad (4)$$



(a) (b)

(c)

Fig.2 Blocks with fast moving edges

or SBD changes rapidly:

$$\left|SBD_k(i,j) - SBD_{k-1}(i,j)\right| > \delta S_{n.\max} \qquad (5)$$

we can identify that this block contains fast moving edges. In (4)(5), $\lambda = 10, \delta = 2$, $S_{n.\max}$ is the temporal maximum of the SBD of the reference background block. By default, the blocks in the top-left and top-right corners of the image are chosen to be reference background blocks, for objects rarely appear in this area. When the SBD of these two blocks exceed a certain limit, which means the object might move in this block, we can choose another block whose SBD has been below this limit for a certain period.

3.2. Frame difference accumulation

The method based on accumulative frame difference requires a proper number of frames to accumulate (NFA). Low motion or texture regions need a larger number to be distinguished from the background. For example, the block on mother's left shoulder shown in Fig.3(a) has little texture and there is no intense movement in this region. In Fig.3(b), we compare the SBD of the shoulder block and reference background block. Obviously they cannot be separated from each other. The situation gets better when we accumulate 8 frames shown in Fig.3(c). The difference between these two blocks is distinct if 32 frames are accumulated in Fig.3(d).

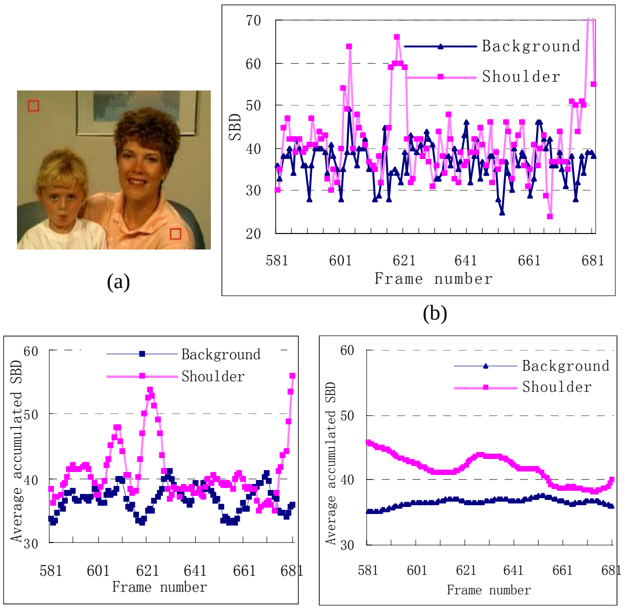
On the other hand, for the regions with high motion we need to decrease the NFA, in order to eliminate the expanded changed region described in section 2. So we calculate each block's NFA based on its motion attributes. The rules are:

A. If $SBD_k < 2S_{n.\max}$, $NFA_k = NFA_{k-1} + \alpha N_{\max}$, where $\alpha = 1/8$, $N_{\max}$ is the maximum possible value of NFA.

B. If $SBD_k - SBD_{k-1} > \delta S_{n.\max}$, $NFA_k = NFA_{k-1} \bullet S_{n.\max} / [SBD_k - SBD_{k-1}]$.

C. If $SBD_k - SBD_{k-1} < -\delta S_{n.\max}$,

$$NFA_k = NFA_{k-1} \bullet [SBD_{k-1} - SBD_k] / S_{n.\max} .$$



(a) (b) (c) (d)

Fig.3 SBD with different accumulated frame numbers (a)two blocks (b)BSD (c)average SBD with 8 frames accumulated (d) average SBD with 32 frames accumulated

The next step is to accumulate the frame difference. For each pixel (x,y), its frame difference FD(x,y) of prior several frames is added up. The max number of frame difference added is the NFA of the block B(i,j) which this pixel belongs to. But if within these frames, there is one in which B(i,j) contains fast moving edges, the frame difference value of this frame and all the preceding ones should not be added.

3.3. Thresholding

The threshold value is determined by the reference background blocks. We use the average AFD of all the pixels in those two blocks as threshold:

$$T_k^N = \frac{\sum_{(x,y)\in B_{bg}} \sum_{i=k}^{k-N} FD_i(x,y)}{\left|B_{bg}\right|} \quad (6)$$

where N is the number of accumulated frames, B_{bg} is background block. As different blocks have different NFA, the threshold for this block might also be different. After comparing each pixel's AFD with the threshold of its region, an initial object mask is generated.

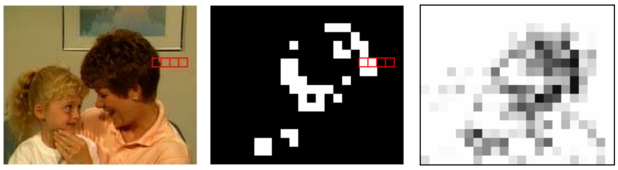
3.4. Post processing

The initial object mask still contains small noise regions, so the first step of post processing is to eliminate these regions. We use morphological open operation to eliminate the background noise regions and close operation to remove the noise regions inside the object. In order to remove lager noise regions, we use the classic connected components algorithm[7] to obtain isolated regions and remove the regions with areas smaller than a certain limit from the object mask.

The next step is to adjust the boundary of the object mask. First open and close operations are applied to the object mask to smooth the boundary. In section 3.2 we decrease the NFA to eliminate the expanded changed region, but there might still be small uncovered regions surrounding the object boundary. So we use region grow method to remove such regions. We use the boundary pixels of the initial object mask as seeds and grow towards the inside of the mask. And if it reaches an edge or the difference between the seed and the current pixel exceeds a threshold, the grow process stops. After this, the post processing is finished and the final segmentation result is obtained.

4. EXPERIMENTAL RESULTS

We applied our approach on the head-shoulder sequences (QCIF format, 10Hz frame rate): *akyio*, *claire*, *mother and daughter*, *salesman*, *silent*, *suzie*, and the performance is very well. Here we select the videos with noise and fast, intense movement to demonstrate the advantage of our method.



(a) (b) (c)

Fig. 4. Result of motion analysis of sequence *mother and daughter* (a)frame 780 (b)fast moving edge detection (c)NFA result

First, we present the result of motion analysis in our approach. As shown in Fig4(a), mother's head moves towards the daughter fast. Fig.4(b) shows the result of the detection of blocks containing fast moving edges(the white ones). We can see the moving edges of mother's head and hand are detected successfully. Fig.4(c) shows the NFA of all the blocks in this frame, where white represents the maximum possible NFA and black is the minimum value. In this picture, we can see the blocks belonging to mother's head get smaller NFA, and the blocks with low motion, such as the background, mother and daughter's shoulders get bigger NFA.

Fig.5 Comparison of final result of sequence *mother and daughter* , frame 725, 780, 850. First line: proposed approach. Second line: reference approach with a constant NFA 32.

The final segmentation results are presented in Fig.5~8 with comparison of the algorithm in [6], which is based on a simple accumulative frame difference method. Fig.5 illustrates a sequence at the moment when mother's head moves suddenly towards the daughter and moves back. As we can see, with a constant NFA for the whole frame, former movement is accumulated and expands the object's boundary in the segmentation result. The proposed approach can eliminate such expansion very well and significantly improves the segmentation result.

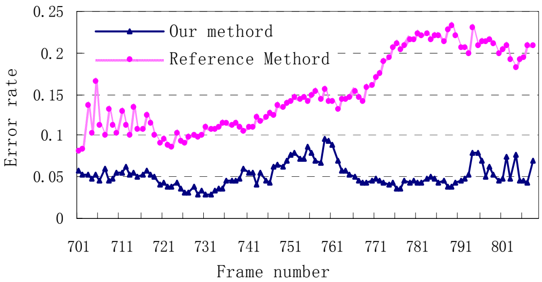


Fig. 6 Error rate of sequence *mother and daughter*.

An objective evaluation is also carried out on this sequence. The error rate of the object mask is adopted to present the improvement of our approach, shown in Fig.6. The error rate is the ratio of error pixel count to the object size. Mother's head starts to move after frame 760 and the error rate of the reference method start to increase greatly. In contrast, our method still can maintain a low error rate. And when the object does not have intense movement, our approach can also achieve a better segmentation quality.

Fig.7 and Fig. 8 show another two comparison of segmentation results, which also indicate that our approach can achieve accurate segmentation in cases of intense movement.

5. CONCLUSION

In this paper, we proposed a method based on accumulative frame difference to extract video objects from head-shoulder video sequences. We perform a preliminary motion analysis of the current frame first and calculate an adaptive accumulated frame number for each block of the frame based on the local motion attributes. Then frame difference is accumulated. After thresholding the accumulated frame difference, an initial object is generated. Then post processing removes the small noise region and adjusts the object's boundary. Experimental results demonstrate that the proposed method eliminates the over accumulated frame difference and significantly improves the segmentation result.

6. REFERENCES

[1] A. Neri, S. Colonnese, G. Russo, and P. Talone, "Automatic moving object and background separation," Signal Processing, vol. 66, no. 2, pp. 219–232, 1998.


Fig. 7 Comparison of final result of sequence *silent*, frame 100, 200, 300. First line: proposed approach. Second line: reference approach with a constant NFA 32.


Fig. 8 Comparison of final result of sequence *suzie* , frame 40, 50, 60. First line: proposed approach. Second line: reference approach with a constant NFA 32.

[2] Shao-Yi Chien, Shyh-Yih Ma, Liang-Gee Chen, "Efficient moving object segmentation algorithm using background registration technique," Circuits and Systems for Video Technology, IEEE Transactions on, Volume 12, Issue 7, July 2002 Page(s):577 - 586

[3] R. J. Radke, S. Andra, 0. A-Kofahi, and B. Roysam, "Image change detection algorithms: a systematic survey," IEEE Transactions On Image Processing, vol. 14, no. 3, pp. 294-307, 2005.

[4] D. Zhang and G. Lu, "Segmentation of moving objects in image sequence: a review," Circuits Systems Signal Processing, vol. 20, no. 2, pp. 143-183, 2001.

[5] Til Aach, André Kaup, Rudolf Mester, "Statistical model-based change detection in moving video", Signal Processing. Vol. 31, no. 2, pp. 165-180. 1993

[6] Mao-Ying Liu, Qiong-Hai Dai, Xiao-Dong Liu and Gui-Hua Er, "Automatic extraction of initial moving object based on advanced feature &video analysis," Proc. Visual Communication and Image Processing VCIP-2005, pp.160-167, Beijing, China, July 2005

[7] R. M. Haralick and L. G. Shapiro, Computer and Robot Vision. Reading, MA: Addison-Wesley, 1992, pp. 28–48.