Efficient Block-based Motion Segmentation Method using Motion Vector Consistency

Ronald H.Y. Chung, Francis Y.L. Chin, Kwan-Yee K. Wong, K.P. Chow, T. Luo and Henry S.K. Fung

Abstract

This paper presents a new and efficient block-based motion segmentation method based on a novel motion vectors consistency model. In essence, the method utilizes the motion vectors extracted from the video encoding process and groups the associated blocks by region growing technique to achieve motion segmentation. A new motion vector consistency criterion is also introduced to support this block-based region growing mechanism. The proposed method was evaluated by the "Hall Monitor" and "Table Tennis" sequences. Experimental results show that the proposed method can successfully segment the moving objects in the two video sequences, with each object being represented by reasonably good block-based boundary. In addition to that, it is also found that the motion segmentation method can operate together with a video codec to achieve simultaneous real-time video encoding and segmentation, demonstrating its potential in video surveillance applications and region-of-interest video encoding.

***Index Terms*— Motion Segmentation, Motion Vectors Consistency, Region Growing.**

1 Introduction

Motion segmentation is the process of decomposing a video scene into independently moving objects or regions. It has been regarded as an important and indispensable step in many vision-based applications. Recently, it is becoming more and more coupled to video coding. For instance, in object-based video coding, moving objects have to be segmented first, so that better video compression can be achieved by investing more bits in coding moving objects than coding the stationary background as suggested in the MPEG-4 standard [1]. On the other hand, many video surveillance systems nowadays need to perform intelligent tasks, in addition to video recording, such as tracking and intruder detection, etc. Moving objects must then be first extracted before any of these analysis tasks can proceed. Due to the increasing demand of object-based video coding and sophisticated video analysis in video surveillance applications, it is foreseeable that motion segmentation and video coding have to coexist in many applications.

There has been a number of motion segmentation methods proposed in the literature [2-4]. Many of them regard the extraction of motion information as a separate process without noting that some information derived in the video encoding process can be reused. For example, in most applications such as video conferencing and video surveillance, etc., motion information has already been extracted in the video encoding phase. As a result, it would be a more cost-effective approach if such motion information could be utilized for motion segmentation, rather than treating motion segmentation as another separate and independent process. In view of this, we are motivated to research into this aspect. As MPEG standards [1] [5-6] are widely used for video compression purpose, we started by investigating the motion information – motion vectors – obtained in MPEG compression to see whether they can help in motion segmentation.

Under the MPEG compression framework, each video frame is partitioned into non-overlapping regular blocks, with a typical size of 16x16 for MPEG-1 [5] and MPEG-2 [6], and a finer size of 8x8 for MPEG-4. For each block, temporal redundancies are first exploited in the motion estimation process to determine the motion vector of each block. Compression can therefore be achieved by encoding the motion vectors together with the error residues associated with the corresponding blocks. It is found in [7] that there is a relationship between the motion vectors of the blocks which belong to the same object. In this paper, we first derive a spatial motion vector consistency model from this observed motion vector relationship, and then formulate a block grouping criterion for grouping neighboring blocks with consistent motions. A new motion segmentation method is then proposed based on this block grouping mechanism. In essence, the proposed method is similar in spirit to region

Ronald H.Y. Chung is with the Department of Computer Science, University of Hong Kong, Pokfulam Road, Hong Kong (e-mail: hychung@cs.hku.hk)
Francis Y.L. Chin is with the Department of Computer Science, University of Hong Kong, Pokfulam Road, Hong Kong (e-mail: chin@cs.hku.hk).
Kwan-Yee K. Wong is with the Department of Computer Science, University of Hong Kong, Pokfulam Road, Hong Kong (e-mail: kykwong@cs.hku.hk).
K.P. Chow is with the Department of Computer Science, University of Hong Kong, Pokfulam Road, Hong Kong (e-mail: chow@cs.hku.hk).
T. Luo is with the Multivision Intelligent Surveillance Limited, Hong Kong (e-mail: tluo@cs.hku.hk).
Henry S.K. Fung is with the Department of Computer Science, University of Hong Kong, Pokfulam Road, Hong Kong (e-mail: skfung@cs.hku.hk).

growing segmentation except that the region growing is performed on a block-wise basis with the homogeneity measure being the motion vector consistency. We implemented the proposed motion segmentation method and evaluated its performance on the “Hall Monitor” and “Table Tennis” sequences. It is found that real-time motion segmentation can be achieved by using the motion vectors derived from an MPEG-4 codec and moving objects can be segmented with reasonably good block-based boundaries.

This paper is organized as follows. Section II presents the motion vector consistency model and the block grouping criterion, whereas Section III reveals the details of the proposed motion segmentation method. Experimental results are then given in Section IV, followed by the conclusion and discussion of future works in Section V.

2 Motion Vector Consistency Model and Block Grouping Criterion

2.1 Motion Vector Consistency Model

Let $B(m, n)$ denotes the block with block size W x W in the m-th column and n-th row of the current frame; and let $MV(m, n) = [MV_x(m, n), MV_y(m, n)]^t$ denotes the motion vector of $B(m, n)$. For any two blocks $B(m, n)$ and $B(m', n')$ in the current frame as depicted in Figure 1, it is shown in [7] that the following condition holds when both $B(m, n)$ and $B(m', n')$ lie on the same object:

$$(m'W - mW)(MV_x(m', n') - MV_x(m, n)) + (n'W - nW)(MV_y(m', n') - MV_y(m, n)) = 0, \quad (1)$$

which can be further simplified as follows:

$$[m' - m, n' - n][MV(m', n') - MV(m, n)] = 0. \quad (2)$$

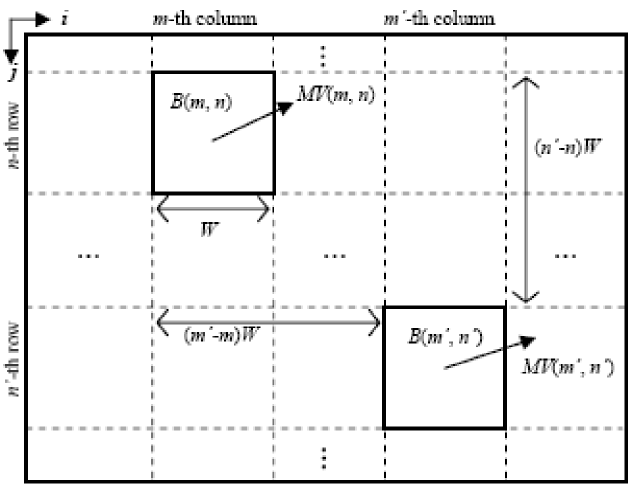


Figure 1. Two blocks $B(m, n)$ and $B(m', n')$ with motion vectors $MV(m, n)$ and $MV(m', n')$ in the current frame.

2.2 Block Grouping Criterion

Based on the aforementioned motion consistency model, a moving object can be theoretically identified as the group of blocks in which every block pairs satisfy Eq. (2). However, due to the imperfect precision of motion vectors and the fact that motion vectors can be trapped into non-optimal solutions, Eq. (2) may not always hold even if the two blocks really lie on the same object and hence it is not a practical block grouping criterion. To formulate a practical block grouping criterion, we first define the motion consistency index $d(B(m, n), B(m', n'))$ for the blocks $B(m, n)$, $B(m', n')$ as follows:

$$\begin{aligned} & d(B(m,n), B(m',n')) \\ & = |[m'-m, n'-n][MV(m',n') - MV(m,n)]|. \end{aligned} \quad (3)$$

With $d(\cdot,\cdot)$, the criterion for grouping a pair of blocks $B(m, n)$ and $B(m', n')$ is then revised as follows:

$$d((B(m, n), B(m', n')) < \varepsilon, \quad (4)$$

where ε is an introduced threshold that is video sequence dependent. Whenever (4) holds for any two blocks, these two blocks are likely to belong to the same object and hence they should be grouped together.

3 Proposed Motion Segmentation Method

As mentioned in Section II, a moving object can be theoretically identified as a group of blocks in which the motion vectors are consistent with each other. With the introduced motion consistency index in Eq. (3) and the block grouping criterion in Eq. (4), this segmentation problem can then be solved by region growing technique such as the one proposed in [8].

In essence, our proposed method is built upon the block-based region growing concept, under which each block is being treated as the elementary unit for region growing with Eq. (3) serving as the homogeneity measure. Before region growing can proceed, a seed block has to be determined first. In our method, a seed block $B(m_s, n_s)$ is a block which has consistent motion with its four direct neighbors $B(m_s - 1, n_s)$, $B(m_s, n_s - 1)$, $B(m_s + 1, n_s)$ and $B(m_s, n_s + 1)$. To ensure that the seed block is indeed a good starting point for region growing, we test all the block pairs between the seed block and these four neighboring blocks to guarantee strong motion consistencies. In short, the seed block should satisfy the following constraint:

$$\begin{aligned} & d(B(m_s, n_s), B(m_s + i, n_s + j)) < \varepsilon, \\ & \forall i, j \in \{-1,0,1\} \text{ and } |i| + |j| = 1. \end{aligned} \quad (5)$$

After a seed block is found, the seed block and its four direct neighbors are first grouped together to form a new region $R = \{B(m_s, n_s), B(m_s\text{-}1, n_s), B(m_s, n_s\text{-}1), B(m_s\text{+}1, n_s), B(m_s, n_s + 1)\}$. To grow a region R, each ungrouped surrounding block $B(m, n)$ that borders the region R is grouped into the region only if its motion consistency indices with every block in the region R satisfy Eq. (4) in the average sense. The grouping criterion for a block $B(m, n)$ and a region R can be best described by the following condition:

$$If \; \frac{\sum_{B(m_r,n_r)\in R} d(B(m,n), B(m_r,n_r))}{\|R\|} < \varepsilon,$$

$$then \; B(m,n) \; is \; grouped \; into \; R \,,$$

where $\|R\|$ denotes the number of blocks in the region R.

This region growing process repeats until no further block grouping can proceed. If no further block grouping is possible, another new seed block will then be identified as described before. In the case when neither block grouping is possible nor a new region can evolve, ε will be increased to the next smallest possible value that can trigger either one of the two events. This whole process will repeat until all the blocks are grouped into their respective regions. Upon completion of the motion segmentation, each region is represented by a corresponding group of blocks.

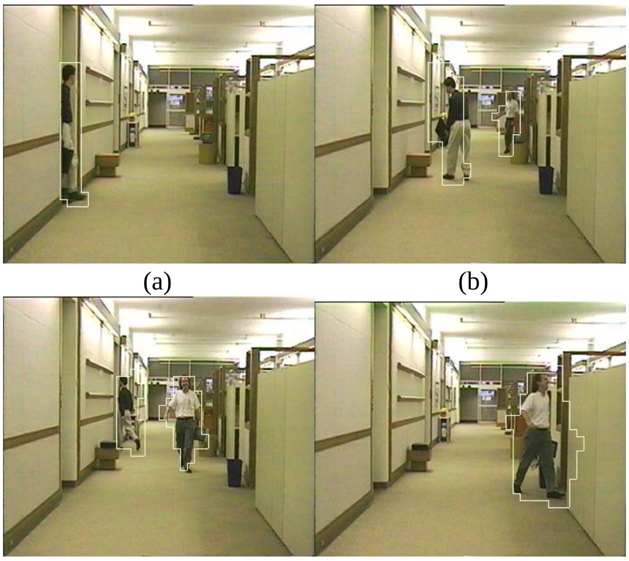
4 Experimental Results

We integrated the motion segmentation method described in Section III into the open source Xvid MPEG-4 codec [9], so that every motion vector obtained can be passed to our motion segmentation algorithm. We prefer to do it this way instead of extracting the motion vectors from the encoded MPEG-4 bitstream for two reasons. First, some of the blocks might be INTRA-coded, and therefore their motion vectors are no longer retained in the encoded bitstream for segmentation purpose. Second, it enables us to further investigate the possibility of region-of-interest coding, which is one of our future directions, if we can perform motion segmentation in the middle of the video encoding process. Two CIF video sequences, namely the “Hall Monitor” and the “Table Tennis” sequences, were used to evaluate the performance of the proposed motion segmentation method. The testing platform is a personal computer with a Pentium IV 2.26 GHz processor, on which each sequence was encoded into a simple profile MPEG-4 bitstream using the Xvid codec. The blocks can either be 16x16 or 8x8 in size (4-MV option), and their motion vectors obtained were then passed to our motion segmentation method. For each 16x16 block, it was treated as four 8x8 blocks with identical motion vectors so that a unified block size of 8x8 was used throughout the whole motion segmentation process.

When evaluated by the 300 frames “Hall Monitor” sequence, it was found that the proposed method could segment the two people successfully throughout the whole video sequence. Figure 2(a)-(d) shows the segmentation results in the 18-th, 88-th, 233-rd and 298-th frames, respectively. Our proposed motion segmentation method first segmented, in the 18-th frame shown in Figure 2(a), the man in black T-shirt on the left hand side who started to move out from a room in the 13-th frame. The reason for the delay is due to the fact that a new region requires the seed block to be motion consistent with its four direct neighbors. If a moving object cannot span over at least five blocks, no seed block can be found. Our proposed motion segmentation method therefore could not detect this man until the 18-th frame when over half portion of this man was visible. For the same reason, when the second man in white T-shirt on the right started to appear in the 74-th frame, our proposed method could not detect the second man until the 88-th frame as depicted in Figure 2(b). Nonetheless, once the two men were successfully segmented, our proposed method could keep them segmented before they disappeared. Figure 2(c) shows the segmentation result when the man on the left was about to disappear in the 233-rd frame whereas Figure 2(d) shows the result when the man on the right was about to disappear in the 298-th frame. This demonstrates that the proposed block-based motion segmentation method is quite successful in segmenting moving objects in video sequence. On the other hand, as depicted in Figure 2(a) – 2(d), the two segmented men can be reasonably well represented by the block-based boundaries.

Another 120 frames “Table Tennis” sequence was also used to evaluate the proposed method. Figure 3(a) – 3(d) shows the segmentation results of the sequence in the 17-th, 69-th, 80-th and 100-th frames, respectively. As depicted in Figure 3(a), both the hand and the ping-ping ball could be quite correctly segmented. However, it is worth mentioning that the proposed method could not correctly segment the player and the ball between the 22-nd and 67-th frames due to the global camera zoom-out motion and scene change. Figure 3(b)-3(d), however, shows that as long as there was no global camera motion, the proposed method correctly segmented the player from the sequence again. Another little problem that can be noticed from this sequence is that due to the block-based nature of the proposed method, a small moving object might not be segmented as depicted in Figure 3(c), where the small ping-pong ball was not segmented in the frame.

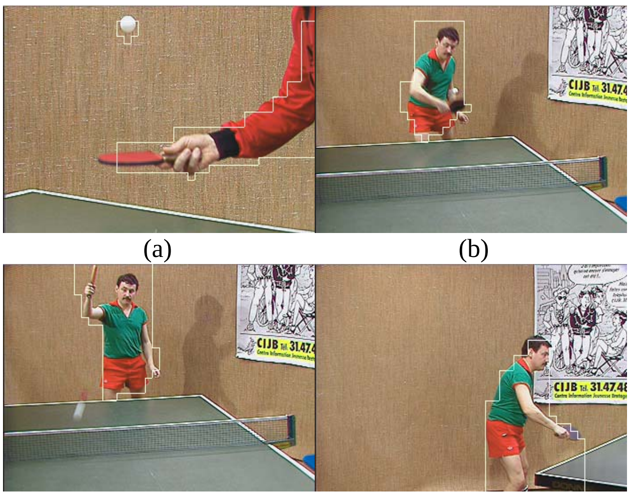
Despite the fact that the proposed method suffers from a few problems noticeable in the “Table Tennis” sequence, the experimental results show that the proposed method works well under two conditions: 1) when the motion vectors are properly compensated from global camera motion; and 2) when the moving objects are large compared with the block size. Besides, if the moving objects can be segmented, the block-based representation usually provides reasonably good approximations to the ideal object boundaries. Furthermore, it is also found that the proposed method, when integrated into a video codec, can achieve real-time performance in simultaneous video encoding and segmentation at a frame rate of 25 frames/sec. In particular, the proposed motion segmentation method takes less than 1 ms per frame, which means that it is suitable for applications that demand real-time video analysis.



(a) (b)

(c) (d)

Figure 2. Segmentation results in the "Hall Monitor" sequence (a) The first segmented man in the 18-th frame, (b) The second man segmented in the 88-th frame, (c) Segmentation results when the first man was about to disappear in the 233-rd frame, (d) Segmentation results when the second man was about to disappear in the 298-th frame.



(a) (b)

(c) (d)

Figure 3. Segmentation results in the "Table Tennis" sequence (a) The 17-th frame, (b) The 69-th frame, (c) The 80-th frame, (d) The 100-th frame.

5 Conclusions and Future Directions

In this paper, a motion vector consistency model has been proposed. Based on this model, a new and efficient block-based motion segmentation method is introduced. The segmentation results from the "Hall Monitor" and "Table Tennis" sequences show that the proposed method can utilize the motion vectors derived from a MPEG-4 codec to perform motion segmentation with each segmented object being approximated by a block-based boundary. It also offers a practical approach to integrate the video encoding and motion segmentation process, which indicates that the proposed segmentation method is suitable for real-time video surveillance applications. In addition to that, since moving regions can be identified during the encoding process, the proposed method also has the potential in supporting region-of-interest video encoding. Future directions will be focused on two areas: 1) To include global camera motion compensation in the motion consistency model to achieve a more robust motion segmentation; and 2) To work out a region-of-interest video encoding scheme based on the proposed motion segmentation method.

Acknowledgment

This research was jointly sponsored by Multivision Intelligence Surveillance Limited and the Innovation and Technology Commission of the Government of the Hong Kong Special Administrative Region, under the Grant UIM/118.

References

[1] ISO/IEC JTC1 IS 14496-2 (MPEG-4), "Information Technology – Generic Coding of Audio-Visual Objects (Final Draft of International Standard)", Oct. 1998.

[2] K.-W. Lee and J. Kim, "Moving object segmentation based on statistical motion model," *IEE Electronics Letters*, 1999, 35, (20), pp. 1719-1720.

[3] H. T. Nguyen, M. Worring and A. Dev, "Detection of moving objects in video using a robust motion similarity measure," *IEEE Trans. Image Process.*, 2000, 9, (1), pp. 137-141.

[4] J. Pons, J. Prades-Nebot, A. Albiol and J. Molina, "Fast motion detection in compressed domain for video surveillance," *IEE Electrons Letters*, 2002, 38, (9), pp. 409-411.

[5] ISO/IEC CD 11172-2 (MPEG-1 Video), "Information technology – Coding of moving pictures and associated audio for digital storage media at up to about 1.5 Mbits/s: Video," 1993.

[6] ISO/IEC CD 13818-2 – ITU-T H.262 (MPEG-2 Video), "Information technology – Generic coding of moving pictures and associated audio information: Video," 1995.

[7] H.Y. Chung, N.H.C. Yung and P.Y.S. Cheung, "Fast motion estimation with search center prediction," *Optical Engineering - The Journal of SPIE*, Vol. 40, No. 6, June 2001, pp.952-963.

[8] R. Adams and L. Bischof, "Seeded region growing," *IEEE Trans. PAMI*, vol. 16, no. 6, pp. 641-647, Jun. 1994.

[9] XviD codec, URL:http://www.xvid.org/