MOVING-OBJECT SEGMENTATION WITH ADAPTIVE SPRITE FOR DCT-BASED VIDEO CODER

Salkmann Ji [1] *and HyunWook Park* [2]

[1] Digital Media Research Laboratory, LG Electronics, Korea
simonji@lge.com
[2] Department of Electrical Engineering, KAIST, Korea
hwpark@athena.kaist.ac.kr

ABSTRACT

We have developed a moving-object segmentation method for standard-compliant DCT-based video coder. To be compliant with current block-based video coding standards, the proposed algorithm uses the DCT coefficients and motion vector of a block, respectively, as spatial and temporal features of segmentation. Since block-based motion vector may not be sufficient to represent true motion of object, the proposed algorithm overcomes this problem by using statistical approach. Moreover, the proposed segmentation algorithm operates in parallel with the video coding. Therefore, the proposed segmentation algorithm can be efficiently integrated in parallel with the DCT-based video coder for the improvement of coding efficiency as well as segmentation accuracy.

1. INTRODUCTION

Video segmentation is to partition a video sequence into several regions according to a given criterion. In MPEG-4, video coder requires object segmentation as pre-processing, so that the compression efficiency can be improved. In addition, static sprite in MPEG-4 can improve the compression efficieny, where a static sprite is a large image describing panoramic background [1][2].

Depending on a user interaction, segmentation techniques can be categorized into automatic and semi-automatic segmentation [3][4]. However, in real-time applications such as video conferencing and remote surveillance, the user interation is limited. Thus, automatic segmentation is preferred to semi-automatic in real-time applications.

Most video coding standards such as MPEG-4 and H.263 adopt spatial-domain motion-estimation (SD-ME) video coder, where DCT and inverse DCT are located within the loop of temporal prediction and motion estimation is performed in spatial domain. However, if DCT is moved out of the loop, IDCT is not necessary in the loop. This alternative coder is called the transform-domain motion-estimation (TD-ME) coder. In this case, the motion estimation is performed in transform domain. If motion estimation is performed in DCT-domain, DCT coefficients of all the candidate blocks of the reference frame should be obtained for motion estimation and compensationa. Chang [5] had developed the DCT-domain motion compensation algorithm for manipulating and compositing of motion-compensated DCT-based image. Also, Song [6] developed the DCT-domain motion estimation algorithm.

In this paper, we have developed automatic and block-based moving-object segmentation. Since the proposed segmentation uses the DCT coefficients and motion vector of a block as spatio-temporal features for segmentation, it is integrated with the TD-ME coder. In addition, the proposed algorithm is used for a sprite-based coding without any input information such as segmentation mask. The proposed algorithm not only segments moving objects but also improves the coding efficiency of the coder.

2. PROPOSED ALGORITHM

2.1. Overview

Figure 1 shows the structure of the proposed segmentation in the TD-ME coder. In the TD-ME coder, motion vector and DCT coefficients are produced for the video coding (compression), and the proposed segmentation algorithm uses the motion vector and DCT coefficients as spatio-temporal features for segmentation. Thus, the proposed segmentation algorithm operates in parallel with the video coding.

0-7803-6725-1/01/$10.00 ©2001 IEEE

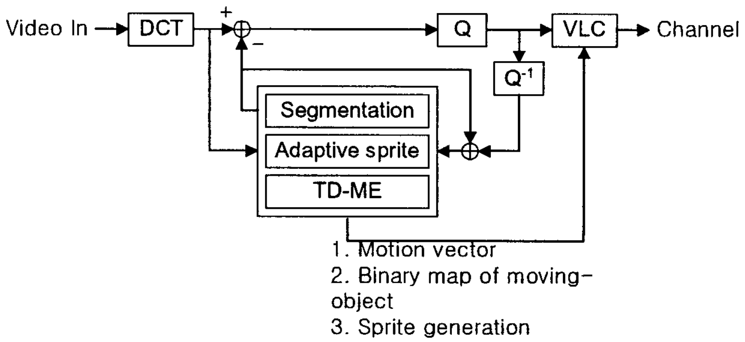

Figure 1. Structure of the proposed segmentation in the TD-ME coder

Figure 2 shows the block diagram of the proposed moving-object segmentation algorithm, which uses adaptive sprite. Suppose that an object is moving in a video sequence. At first, each frame is divided into several homogeneous *regions*. These regions are classified into *dynamic regions* and *static regions* by the proposed algorithm. Dynamic regions are parts of moving object, which is moving at current frame. Next, *projected regions* are found among static regions by projecting the moving objects at previous frame. These projected regions are classified into *projected dynamic regions* and *uncovered regions* by the proposed algorithm. Projected dynamic regions are part of moving object, which was moving at previous frame but stops at current frame. At last, moving objects are formed from both dynamic regions and projected dynamic regions.

2.2. Adaptive sprite

In the proposed sprite generation, it is supposed that segmentation information and sprite information are unknown in advance. The proposed segmentation algorithm considers the first frame of a video sequence as initial sprite. After the segmentation information is obtained, the proposed algorithm updates the sprite adaptively.

Figure 3 shows the proposed adaptive sprite generation. Sprite-blocks can be divided into certain and uncertain. If a sprite-block has been updated or certain up to the previous frame, the sprite-block must be included in a sprite and is called certain sprite-block. But if a sprite-block has not been updated or certain up to the previous frame, the sprite-block might not be a sprite and is called uncertain sprite-block.

For the uncovered block, the sprite-block is updated and become certain if it is not the same as the uncovered block. But, the sprite-block is not updated if it is the same as the uncovered block. In this case, the sprite-block must be certain. For blocks except the uncovered block, the sprite-block might be uncertain if the block has never been an uncovered block up to now. And the sprite-block might be certain if the block has never been a covered block or a moving block up to now.

2.3. Temporal feature for segmentation

Let *active blocks* be inter-coded blocks with non-zero motion vector or intra-coded blocks, whereas *inactive blocks* be inter-coded blocks with zero motion vector.

$$\{active\,blocks\} = \{\mathbf{b}_{t,x,y} \mid INTER\ with\ (u_x, u_y) \neq 0\} \cup \{\mathbf{b}_{t,x,y} \mid INTRA\}$$

$$\{inactive\,blocks\} = \{\mathbf{b}_{t,x,y} \mid INTER\ with\ (u_x, u_y) = 0\}$$

where $\mathbf{b}_{t,x,y}$ is an 8×8 block at time t and location (x, y) with motion vector (u_x, u_y).

2.4. Spatial feature for segmentation

From DCT coefficients $v(k,l)$, $0 \le k,l \le 7$., of 8×8 block, spatial feature vector is defined as follows,

$$\vec{f}^{spatial} = (f_0, f_1, f_2, f_3) = \left(v(0,0), \sqrt{\sum_{l=1}^{7} v(0,l)}, \sqrt{\sum_{k=1}^{7} v(k,0)}, \sqrt{\sum_{k=1}^{7} \sum_{l=1}^{7} v(k,l)} \right)$$

Spatially similar blocks have similar spatial feature vector and form a homogeneous region.

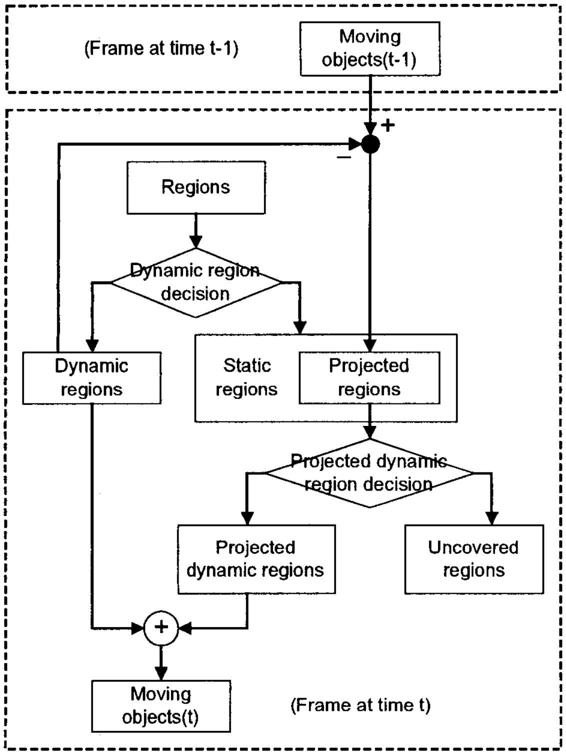

Figure 2. Block diagram of the proposed segmentation algorithm

2.5. Dynamic region detection

As shown in Figure 2, dynamic region detection classifies all regions into dynamic regionss and static regions. In a region, all blocks are classified into *type-I* and *type-II* blocks according to classification rules, which are differently applied to inter-coded dynamic blocks and intra-coded dynamic blocks.

The classification rule for intra-coded blocks is given as follows,

$$\begin{array}{ll} if\left(fd_{t,x,y} < sd_{t,x,y}\right) & type-I \\ else & type-II \end{array}$$

where

$$fd_{t,x,y} = MSE\left(\mathbf{b}_{t,x,y}, \mathbf{b}_{t-1,x,y}\right)$$
$$sd_{t,x,y} = MSE\left(\mathbf{b}_{t,x,y}, \mathbf{b}_{sprite,x,y}\right)$$

$fd_{t,x,y}$ and $sd_{t,x,y}$ are the inter-frame differences of $\mathbf{b}_{t,x,y}$ with previous frame and sprite, respectively. $MSE\left(\mathbf{b}_i, \mathbf{b}_j\right)$ is a mean of squared error between $\mathbf{b}_i$ and $\mathbf{b}_j$. The classification rule is preserved in DCT domain by Parseval's theorem [7].

The classfication rule for inter-coded blocks is given as follows,

$$\begin{array}{ll} if\left(mce_{t,x,y} < \min\left(fd_{t,x,y}, sd_{t,x,y}\right)\right) & type-I \\ else & type-II \end{array}$$

where

$$mce_{t,x,y} = MSE\left(\mathbf{b}_{t,x,y}, \mathbf{b}_{t-1,x-u_x,y-u_y}\right)$$

$mse_{t,x,y}$ is the motion-compensated error of $\mathbf{b}_{t,x,y}$, and it is preserved in DCT domain also.

In a region, if the number of true moving blocks is larger than the number of false moving blocks, then the region is a dynamic region. Otherwise, the region is a static region.

2.6. Projected dynamic region detection

As shown in Figure 2, projected dynamic region detection classifies all regions into projected dynamic regions and static regions.

In a region, all blocks are classified into *type-I* and *type-II* blocks according to classification rules, which are differently applied to certain sprite and uncertain sprite.

The classification rule for certain sprite is given as follows,

$$\begin{array}{ll} if\begin{bmatrix} \left\{fd_{t,x,y} < sd_{t,x,y}\right\} \; and \\ \left\{\left|sd_{t,x,y} - fd_{t,x,y}\right| > 2 \cdot QP\right\} \end{bmatrix} & type-I \\ else & type-II \end{array}$$

And the classification rule for uncertain sprite is given as follows,

$$\begin{array}{l} if\begin{Bmatrix} \left(\mathbf{b}_{t,x,y} \in dynamic\_region\right) \; and \\ \left(\mathbf{b}_{t,x-u_x,y-u_y} \in projected\_region\right) \end{Bmatrix} \\ \quad \mathbf{b}_{t,x-u_x,y-u_y} \text{ is } type-I \\ else \\ \quad \mathbf{b}_{t,x-u_x,y-u_y} \text{ is } type-II \end{array}$$

In a projected region, if the number of type-I blocks is larger than the number of type-II blocks, then the region is a projected dynamic region. Otherwise, the region is an uncovered region.

Finally, dynamic regions and projected dynamic regions form a moving object at current frame.

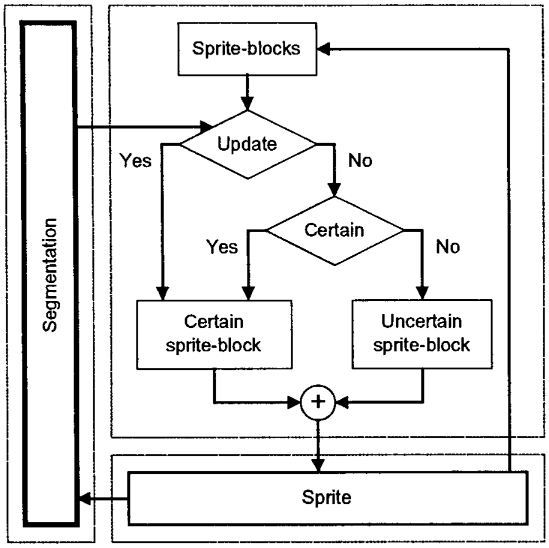

Figure 3. Proposed adaptive sprite generation

3. EXPERIMENTS

The proposed moving-object segmentation algorithm employing sprite-generation is implemented on the TD-ME coder. A sprite is initially the first frame of a video sequence and is updated during the coding. In the TD-ME coder of Figure 1, motion-compensated error is transmitted to the decoder after VLC. However, motion-compensated error is selectively transmitted in the proposed TD-ME coder. I.e., blocks in the moving objects are motion-compensated, coded and transmitted. Thus, the proposed video coder has to send a binary map of the segmented object shape to the decoder

Experiments are performed on "silent voice" QCIF sequence, which has a frame rate of 30 Hz with sampling rate of 10 Hz. And Figure 4 shows one of the experimental results.

4. CONCLUSIONS

We have developed an automatic and block-based moving-object segmentation algorithm, which uses adaptive sprite and operates in parallel with DCT-based video coder. The proposed DCT-based coder does not require any segmentation mask a priori because the proposed segmentation algorithm uses features that are essential for video coding. Experimental results show that the coding efficiency as well as the segmentation accuracy are improved.

5.REFERENCES

[1] "MPEG-4 video verification model version 15.0," ISO/IEC JTC1/SC29/WG11, December 1999.

[2] "Overview of the MPEG-4 standard," ISO/IEC JTC1/SC29/WG11, March 1999.

[3] F. Moscheni, S. Bhattacharjee, and M. Kunt, "Spatiotemporal segmentation based on region merging," IEEE Transactions on PAMI, Vol. 20, No. 9, pp. 897-915, September 1998.

[4] C. Gu, and M. C. Lee, "Semiautomatic segmentation and tracking of semantic video objects," IEEE Transactions on CSVT, Vol. 8, No. 5, pp. 572-584, September 1998.

[5] S. F. Chang, and D. G. Messerschmitt, "Manipulation and compositing of MC-DCT compressed video," IEEE Journal of Selected Areas in Communications, Vol. 13, No. 1, pp. 1-11, January 1995.

[6] M. Song, A. Cai, and J. A. Sun, "Motion estimation in DCT domain," ICCT-96, Vol. 2, pp. 670-674, 1996.

[7] A. K. Jain, "Fundamentals of digital image processing," Prentice Hall, Englewood Cliffs, NJ, 1989.

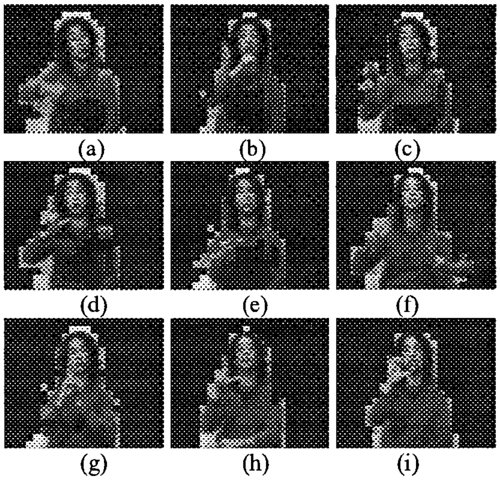

Figure 4. Segmented moving object of "silent voice" sequence: (a) 29th (b) 59th (c) 89th (d) 119th (e) 149th (f) 179th (g) 209th (h) 239th and (i) 269th frames