Region-Based Video Segmentation Using DCT Coefficients

S. Ji and H. W. Park
Dept. of Electrical Engineering
Korea Advanced Institute of Science and Technology
Taejon, KOREA
hwpark@athena.kaist.ac.kr

Abstract

A region-based video segmentation algorithm, which is based on 8×8 block, is proposed to segment moving objects in a video sequence. The proposed video segmentation can be performed in parallel with the general video source coder. The proposed video segmentation uses block DCT (Discrete Cosine Transform) coefficients and block motion vector as segmentation features. The proposed method can be integrated with the video encoder so that different quantization factors can be applied to moving object and background, respectively.

1. Introduction

Video segmentation aims at partitioning a video sequence into several regions according to a given criterion and at making it possible to manipulate the video information. The proposed video segmentation defines a moving object as the meaningful object, since motion is the commonly accepted basic information of a video sequence.

Video segmentation techniques can be categorized into automatic and semi-automatic segmentations depending on a user interaction. The automatic segmentation identifies classes of moving objects according to some luminance homogeneity and motion coherence criterion [1]. On the contrary, the semi-automatic segmentation allows user to initially define regions of interest via a graphical user interface and finds out precise boundaries of the region [2]. In most off-line applications, the semi-automatic segmentation gives better segmentation results than the automatic segmentation. But in most on-line real-time applications, the transmission bandwidth and the user interaction are limited. Therefore, the automatic segmentation prevails in on-line applications. The proposed video segmentation is an automatic segmentation method.

To cope with general video source coder, block-based video segmentation is proposed. In MPEG-4, the video source coder may require the segmentation operation as a separate pre-processing without any influence on the compressed bitstream syntax. But the proposed video segmentation operates in parallel with the video source coder because it uses block motion vector and the DCT coefficients for segmentation features.

The proposed video segmentation uses block DCT coefficients of video input as a spatial feature since the DCT coefficients provide the frequency domain information of the block. Especially, high frequency features are used for edge and boundary detection.

In the limited transmission bandwidth applications or very low bit-rate coding applications, the proposed video segmentation can give successful result. The proposed method segments the moving objects from background, so that the video quality of the moving object can be improved in comparison with that of the background by applying different quantization factors to the moving objects and the background.

In this paper, chapter 2 describes the proposed video segmentation in detail. Chapter 3 shows the experimental results. Conclusions are given in chapter 4.

2. Video segmentation algorithm

2.1. Overview

Figure 1 (a) shows the general video source coder and Figure 1 (b) shows the proposed video source coder. As shown in Figure 1 (b), the proposed video segmentation uses the block motion vector as a temporal feature and uses the block DCT coefficients of input video as a spatial feature. Block motion vector is the result of the motion estimation in the video source coder, so any additional calculation is not required to get the temporal feature. Block DCT coefficients are calculated from input image, not from the residual signal of the motion compensation. Using the spatial and temporal feature, moving object is segmented in the scene.

0-7803-5467-2/99/ $ 10.00 © 1999 IEEE

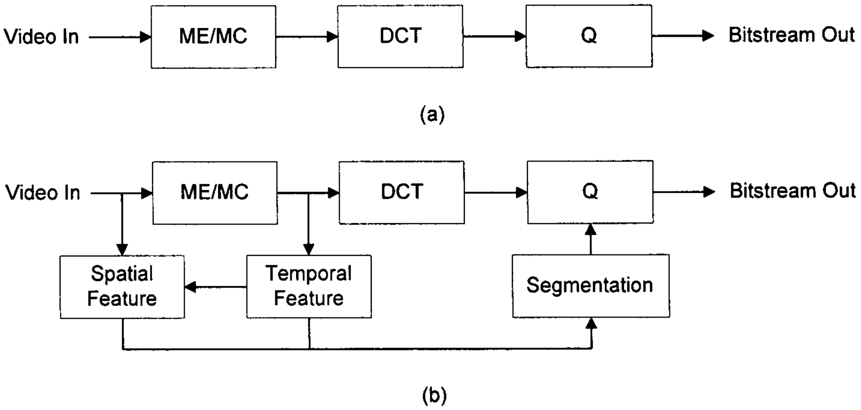

Figure 1. Conceptual diagram of the video source coders: (a) General and (b) Proposed

Compared to general video source coder, quantization factor can be adjusted to get a higher quality video for the moving object than that for the background in the proposed video source coder.

2.2. Block spatial feature

One of block spatial feature description is the block statistics and typical block statistics are the mean and the variance of color values in the region of interests. The mean and variance values can be described by the block DCT coefficients as follows:

Two-dimensional *N*×*N* DCT coefficients *v(k,l)* are given as,

$$
\begin{aligned}
& v(k,l) = \sum_{m=0}^{N-1}\sum_{n=0}^{N-1} a(k,l;m,n)u(m,n) \quad 0 \le k,l \le N-1 \\
& \textit{where} \\
& a(k,l;m,n) \\
& = \alpha(k)\beta(l)\cos\left(\frac{(2m+1)k\pi}{2N}\right)\cos\left(\frac{(2n+1)l\pi}{2N}\right) \\
& \alpha(0) = \beta(0) = \sqrt{\frac{1}{N}} \\
& \alpha(k) = \beta(l) = \sqrt{\frac{2}{N}} \quad \textit{for} \quad 1 \le k,l \le N-1
\end{aligned} \tag{1}
$$

where *u(m,n)* is the image in the spatial domain. The mean and variance are related with the DCT coefficients as follows,

$$\mu = \langle u(m,n) \rangle = \frac{1}{N} v(0,0) \tag{2}$$

$$\sigma^2 = \left\langle (u(m,n)-\mu)^2 \right\rangle = \frac{1}{N^2} \sum_{k=0}^{N-1} \sum_{\substack{l=0 \\ (k,l) \ne (0,0)}}^{N-1} |v(k,l)|^2 \tag{3}$$

In addition, the DCT coefficients have several different spatial features since DCT coefficients provide the frequency domain information in the block. Eq. (4) shows the high frequency DCT coefficients:

$$
\begin{aligned}
& v(0,l) = \alpha(0)\beta(l)\sum_{m=0}^{N-1}\sum_{n=0}^{N-1}\cos\left(\frac{(2n+1)l\pi}{2N}\right)u(m,n) \\
& v(k,0) = \alpha(k)\beta(0)\sum_{m=0}^{N-1}\sum_{n=0}^{N-1}\cos\left(\frac{(2m+1)k\pi}{2N}\right)u(m,n) \\
& v(k,l) = \alpha(k)\beta(l) \\
& \sum_{m=0}^{N-1}\sum_{n=0}^{N-1}\cos\left(\frac{(2m+1)k\pi}{2N}\right)\cos\left(\frac{(2n+1)l\pi}{2N}\right)u(m,n) \\
& 1 \le k,l \le N-1
\end{aligned} \tag{4}
$$

In eq. (4), $v(0,l)$ describes the horizontal high frequency, $v(k,0)$ describes the vertical high frequency, and $v(k,l)$ describes the mixed horizontal and vertical high frequencies for $1 \le k,l \le N-1$. Therefore, the block variance in eq. (3) can be divided into three features as follows,

$$
\begin{aligned}
& \sigma^2 = \sigma_{vedge}^2 + \sigma_{hedge}^2 + \sigma_{texture}^2 \\
& N^2\sigma_{vedge}^2 = \sum_{l=1}^{N-1}|v(0,l)|^2 \\
& N^2\sigma_{hedge}^2 = \sum_{k=1}^{N-1}|v(k,0)|^2 \\
& N^2\sigma_{texture}^2 = \sum_{k=1}^{N-1}\sum_{l=1}^{N-1}|v(k,l)|^2
\end{aligned} \tag{5}
$$

From eqs. (2) and (5), spatial feature vector is defined as eq. (6). The proposed video segmentation uses this spatial feature vector for spatial segmentation.

$$
\begin{aligned}
& f^{spatial} = (f^0, f^1, f^2, f^3) \\
& f^0 = N\mu = |v(0,0)| \\
& f^1 = N\sigma_{vedge} = \sqrt{\sum_{l=1}^{N-1} |v(0,l)|^2} \\
& f^2 = N\sigma_{hedge} = \sqrt{\sum_{k=1}^{N-1} |v(k,0)|^2} \\
& f^3 = N\sigma_{texture} = \sqrt{\sum_{k=1}^{N-1}\sum_{l=1}^{N-1} |v(k,l)|^2}
\end{aligned}
\quad (6)
$$

2.3. Block temporal feature

When an object moves, Figure 2 shows the definitions of moving area, covered area, uncovered area, and background area in two consecutive images. Moving area and covered area belong to the moving object, but uncovered area and background belong to the background at time t+1.

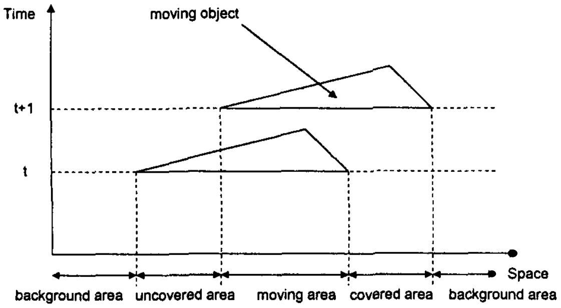

Figure 2. Definitions of moving area, covered area, uncovered area, and background area in two consecutive images.

If motion vector of a block is non-zero, the block is defined as a dynamic block. Otherwise, the block is defined as a static block. Dynamic blocks are classified into five classes according to the relationship between the motion vector of the block and its neighboring blocks as shown in Figure 3. In terms of motion estimation and compensation, INTRA MB (macroblock) has 4 blocks that have no motion vector. Those blocks are classified as class 0 block. INTER 1MV/4MV has 4 blocks that have motion vectors. Those blocks are classified as one of class 1 to class 4. Class 1 has dynamic blocks both ahead and behind. Class 2 has static block ahead and dynamic block behind. Class 3 block has dynamic block ahead and static block behind. Class 4 block has static blocks both ahead and behind.

The blocks of classes 0, 1, 2, and 4 have the high probability of being moving object. To the contrary, the block of class 3 has the high probability of being background.

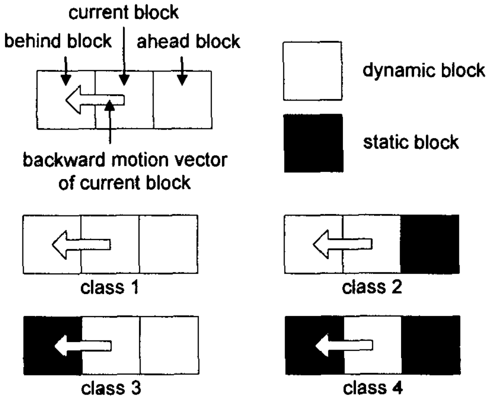

Figure 3. Classes of dynamic blocks

Using MB/block motion vector, temporal feature vector is defined as eq. (7). The proposed video segmentation uses this temporal feature vector for temporal segmentation.

$$
\begin{aligned}
& f^{temporal} = \left(f^4, f^5\right) \\
& f^4 = MB\,MV \\
& f^5 = block\,MV
\end{aligned}
\quad (7)
$$

2.4. Spatial segmentation

The proposed video segmentation uses the block-based region growing for spatial segmentation. Seed blocks are the dynamic blocks defined in section 2.3. The proposed spatial segmentation method checks if the spatial feature vectors of two neighboring blocks are similar according to eq. (8). If they are similar, they are merged into a region and the region growing of current region proceeds until there is no similar block.

$$
\begin{aligned}
& \left|f_i^{spatial} - f_j^{spatial}\right| = \sum_{k=0}^{3} \eta_k \left|f_i^k - f_j^k\right| \le T^{spatial} \\
& i \in N(j) = Neighborhood\ of\ block\ j \\
& \eta = (\eta_0, \eta_1, \eta_2, \eta_3) = metric
\end{aligned}
\quad (8)
$$

When the luminance value ranges between 0 and 255, the luminance smaller than 4 can be considered as noise. Since the spatial feature vector in eq. (6) has four feature components, threshold in eq. (8) is set to $T^{spatial} = 16N$ for $N \times N$ DCT. In eq. (8), the metric is set to $\eta = (1,1,1,1)$.

As a result, several homogeneous regions are obtained. If all blocks in a MB become the same region, the MB is called homogeneous MB. Otherwise, the MB is called inhomogeneous MB.

2.5. Temporal segmentation

The proposed temporal segmentation method checks statistically if the region from the previous spatial segmentation is moving object or background. If the number of true moving block (TB) is larger than that of false moving block (FB) in a region, the region is moving object. Otherwise, the region is background.

Using the temporal feature vector in eq. (7), every dynamic block in the region is examined. As mentioned in section 2.3, the dynamic blocks of classes 0, 1, 2, and 4 have the high probability of being moving block, whereas the dynamic block of class 3 has the high probability of being background. When the proposed video segmentation is realized with the video source coder that finds block motion vector around MB motion vector, MB homogeneity must be considered. When MB is homogeneous, the current dynamic block of class 3 becomes false moving block, and the current dynamic block of the other classes becomes true moving block.

When MB is inhomogeneous or MB homogeneity need not to be considered, the current dynamic block is decided to become true moving block or false moving block according to the relationship between the region of current dynamic block and those of neighboring (both ahead and behind) blocks. In case that the region of current dynamic block is the same as that of neighboring dynamic block and is different from that of neighboring static block, the current dynamic block becomes true moving block. In the opposite case, the current dynamic block becomes false moving block. In the other cases, if the block difference between current frame and previous frame is zero, the current dynamic block becomes false moving block and zero penalty (ZP) is increased, and if not zero, the current dynamic block becomes true moving block and uncertain penalty (UP) is increased.

To decide if the region is the moving object or background, region class R is checked as follows,

$$R = \begin{cases} moving\ object, & if\ \alpha \geq 1 \\ background, & otherwise \end{cases}$$
$$where \qquad (9)$$
$$\alpha = \frac{TB - \sigma(ZP)UP}{FB + \sigma(ZP)UP}, \sigma(ZP) = \begin{cases} 1, ZP > 0 \\ 0, ZP = 0 \end{cases}$$
$$\alpha = \infty\ if\ TB = FB = 0$$

As a result, the proposed temporal segmentation method classifies the region of moving object.

2.6. Object border adaptation

When the number of dynamic blocks in a region is small, the proposed temporal segmentation may result in wrong probability since the number of trial is small. It happens in the object border.

The proposed object border adaptation modifies the border of the region that is obtained from previous spatial and temporal segmentation as shown in Figure 4.

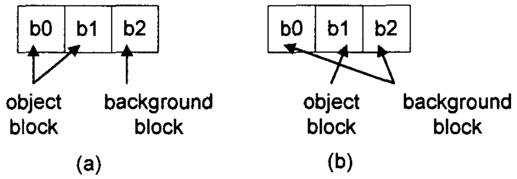

Figure 4. Object border adaptation.

Eq. (10) describes the block edge decision. If b2 (or b0 in Figure 4 (b)) has edge and b1 has no edge, then b2 (or b0 in Figure 4 (b)) is included in the object block.

$$\begin{array}{ll} if\left(\left(N^2\sigma^2_{vedge} + N^2\sigma^2_{hedge}\right) < T_{edge}\right), & edge\ block \\ else & no\ edge\ block \end{array} \qquad (10)$$

The final video segmentation is obtained after object border adaptation. Therefore, the proposed video segmentation is an automatic method for the moving object segmentation in a video sequence.

2.7. Quantization

In the general video source coder, quantization is applied to the DCT coefficients without the concept of object as shown in Figure 1 (a). But, if quantization is differently applied to each object in the scene, a different video quality can be obtained for each object.

Using the object segmentation obtained from the proposed video segmentation, different quantization factors can be applied to each object. So higher quality video for the moving object than for the background can be obtained. However, since the proposed video segmentation uses the motion vector, quantization for the first frame of a video sequence is the same as the general video source coder.

3. Experimental results

The proposed video segmentation is applied to 0th and 3rd frame of "mother and daughter" QCIF sequence and 1st and 2nd frame of "table tennis" SIF sequence. Figure 5 shows the experimental results.

Figure 5 (a) is the dynamic blocks and Figure 5 (b) is the moving objects of "mother and daughter" sequence. Figure 5 (c) is the dynamic blocks and Figure 5 (d) is the moving objects of "table tennis" sequence. These experimental results show that the moving objects are successfully extracted from the scene.

4. Conclusions

An 8×8 block-based video segmentation algorithm is proposed. It uses DCT coefficients of input video and motion vector from motion estimation for the segmentation features. The proposed video segmentation method successfully segments the moving objects from background. By applying different quantization factors to each object, moving object can get higher quality than background.

5. References

[1] F. Moscheni, S. Bhattacharjee, and M. Kunt, "Spatiotemporal Segmentation Based on Region Merging," IEEE PAMI, Vol. 20, No. 9, September 1998, pp. 897-915.

[2] C. Gu, and M. -C. Lee, "Semiautomatic Segmentation and Tracking of Semantic Video Objects," IEEE CSVT, Vol. 8, No. 5, September 1998, pp. 572-584.

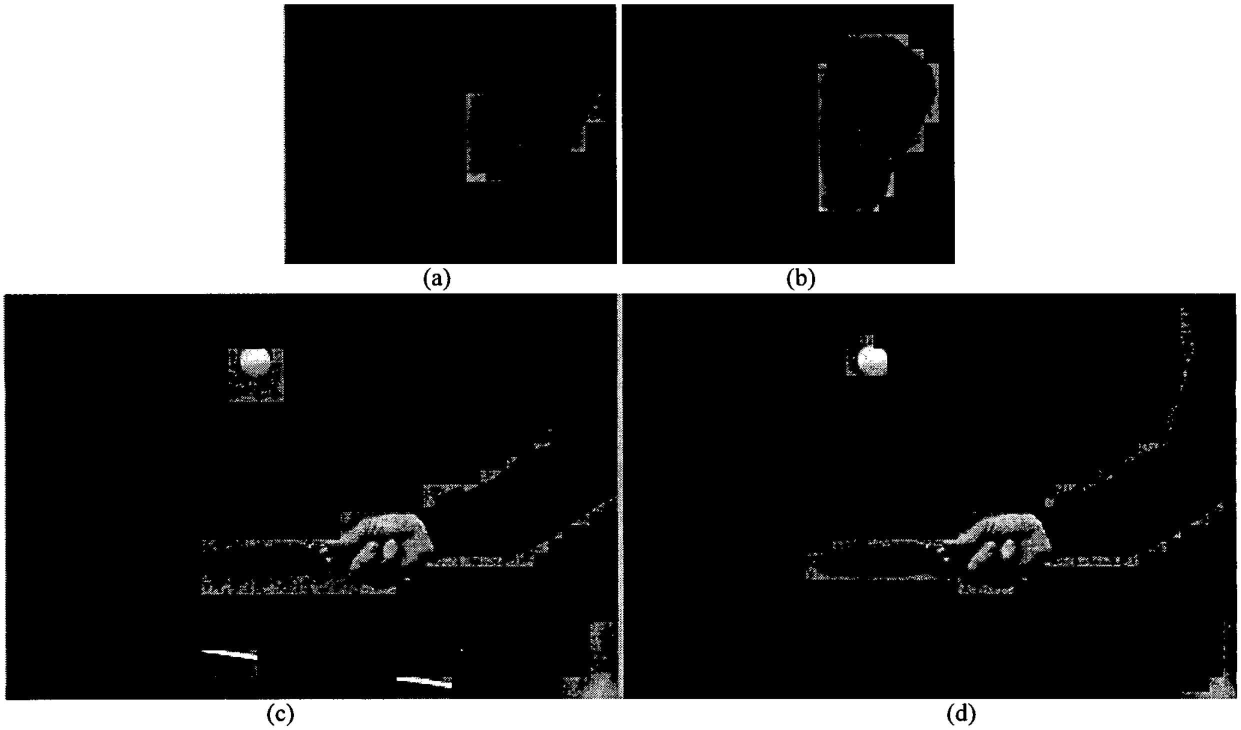

Figure 5. The proposed video segmentation: (a) dynamic blocks and (b) moving objects of "mother and daughter" sequence (c) dynamic blocks and (d) moving objects of "table tennis" sequence.