1—2

An Image Segmentation Technique For Very Low Bit Rate Video Coding

Seok Yoon Jung*, Rin Chul Kim**, Sang Uk Lee***
* Signal Processing Lab., Samsung AIT, Kiheung, Kyungki-Do, KOREA
** School of Information and Computer Eng., Hansung Univ. Seoul, KOREA
*** School of Electrical Eng., Seoul National Univ., Seoul, KOREA
e-mail: jsy@kate.snu.ac.kr

ABSTRACT

This paper describes an image segmentation technique for the object-oriented coding at very low bit rates. By noting that, in the object-oriented coding technique, each objects are represented by 3 parameters, namely, shape, motion, and color informations, we propose a segmentation technique, in which the 3 parameters are fully exploited. To achieve this goal, starting with the color space conversion and the noise reduction, the input image is divided into many small regions by the K-means algorithm on the O-K-S color space. Then, each regions are merged, according to the shape and motion information. In simultations, it is shown that the proposed technique segments the input image into relevant objects, according to the shape and motion as well as the colors. In addition, in order to evaluate the performance of the proposed technique, we introduce the notion of the interesting regions, and provide the results of encoding the image with emphasizing the interesting regions.

I. INTRODUCTION

In the conventional motion compensated image coding techniques, the input image is divided into square blocks, then the motion compensated transform coding technique are performed on each blocks. However, it is well known that those techniques suffer from severe blocking artifact at very low bit rates. In order to alleviate the blocking artifact, many efforts have been made so far[1,12]. Among them, the object-oriented coding technique[1] has received great interests in the related areas. In the object-oriented coding technique, the input image is divided into several regions, each of which identifies the objects having different characteristics. Then, the characteristics of each objects are described by 3 parameters, namely, the shape, motion, and color informations.

In implementing the object-oriented coding, one of the most crucial issues is the segmentation. Conventionally, the input image has been segmented by comparing the colors in the spatial domain or the prediction errors between neighbouring pixels[1-3]. However, those techniques are disadvantageous in that the spatio-temporal similarities between neigbourhoods, such as the shape, motion, etc, can not be taken into account effectively. Recently, Wu[4] proposed the spatio-temporal segmentation technique, based on the polygonal representation of the region shapes. In [4], the input image was first segmented in the spatial domain using the contour detection and the region growing approaches. Then, the neighbouring regions are merged, provided that the small motion compensated errors are yielded in this case. On the other hand, by comparing the segmentation results in both the spatial and the temporal domains separately prepared, Gu proposed the spatio-temporal segmentation technique[4].

In order to effectively exploit the color, shape, and motion in the segmentation, we propose a segmentation technique based on the split-and-merge technique. While the splitting is performed on the color space by using the K-means algorithm, the merging is achieved by using the shape and motion information. Moreover, in our approach, the noise reduction technique is also employed, to provide a proper segmentation result against the noise and irrelevant spatial details. In order to evaluate

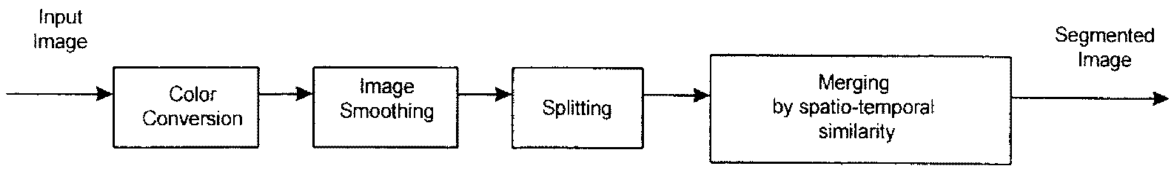


Fig. 1 Block diagram of the proposed technique.

the performance of the proposed technique, we introduce the notion of the interesting regions[11], and provide the coding results on those regions.

II. IMAGE SEGMENTATION

The proposed segmentation technique is based on the split-and-merge technique. Starting with the proper pre-processing, the input image is divided into many small regions by the splitting algorithm on the color space. Then, each regions are merged according to the spatio-temporal similarity between neighbourhoods. To achieve this goal, as shown in Fig. 1, the proposed technique consists of four stages; color space conversion, image smoothing, splitting, and merging. Subsequently, we will describe each stages in more details.

II.1 Color space conversion

In general, the widely employed color coordinate for image coding is the Y-U-V space. However, the O-K-S color space, proposed by Ohta, et al.[13], is known to be the most appropriate to the segmentation[6,7]. Hence, we choose to use the O-K-S space in our approach. Thus, it is necessary to convert the input image in the R-G-B or Y-U-V space into the one in the O-K-S space. The conversion can be achieved by using the relation, given by

$$\begin{pmatrix} R \\ G \\ B \end{pmatrix} = \begin{pmatrix} 1 & \frac{350}{256} & 0 \\ 1 & -\frac{178}{256} & -\frac{86}{256} \\ 1 & 0 & \frac{444}{256} \end{pmatrix} \begin{pmatrix} Y \\ U \\ V \end{pmatrix} + \begin{pmatrix} -175 \\ 132 \\ -222 \end{pmatrix} \tag{1}$$

$$\begin{pmatrix} O \\ K \\ S \end{pmatrix} = \begin{pmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ 1 & 0 & -1 \\ -\frac{1}{2} & 1 & -\frac{1}{2} \end{pmatrix} \begin{pmatrix} R \\ G \\ B \end{pmatrix} + \begin{pmatrix} 0 \\ 128 \\ 128 \end{pmatrix}$$

II.2 image smoothing

In order to alleviate the effects due to the noise, and to obtain better segmentation results, in our approach, the input image is first smoothed out by the median filter and the morphological filter. While the salt-and-pepper noise can be effectively reduced by the median filter[8], the small changes within the same objects can be effectively smoothed out by the morphological filter[2,5]. Note that the object boundaries are preserved by both filters.

The morphological filter, which is implemented by the opening and closing operators, removes spatial details smaller than the pre-defined structuring element, which is defined to be a circle with radius r in our approach. However, it is worth to note that, from the human visual system(HVS) view point, the objects with high temporal activity would be more noticeable than the stationary objects. By segmenting the input image, according to the temporal activity and encoding the input image with emphasizing the high temporal activity, we can obtain an improved subjective quality in the reconstructed image. In this context, we attempt to measure the temporal activity by the frame difference(FD), and adjust the radius r of the circle, according to the relation given by

$$\begin{cases} \text{if } |FD(x,y)| < Th_1, & r = R_1 \\ \text{else}, & r = R_2 < R_1 \end{cases} \tag{2}$$

II.3 Splitting

After smoothing the noise, the image in the O-K-S color space is splitted into small regions by the well-known K-means algorithm[10]. The K-means algorithm is advantageous in that the image is split into desired number of regions in a simple manner, without requiring *a priori* knowledge. As a distance measure for the K-means algorithm, we employ the mean square error(MSE), given by

$$d(X, X_{c,i}) = \sqrt{(o_{c,i}-o)^2+(k_{c,i}-k)^2+(s_{c,i}-s)^2} \quad (3)$$

where $X=(o, k, s)$ and $X_{c,i}=(o_{c,i}, k_{c,i}, s_{c,i})$ denote the current pixel values and the center of the i-th cluster, respectively, in the O-K-S color space.

II.4 Merging

In the splitting stage, we considered only the similarities of color components measured by the MSE. On the other hand, in the merging stage, each splitted regions are merged by taking into account the spatio-temporal similarities between neighbours. First, the neighbouring regions are merged according to the spatial similarity. The spatial similarity is measured by the edge, which is detected by the Laplacian and zero-crossing method[9]. In our approach, two regions whose boundaries do not lie on the edge are considered as a single object, merging them into one region.

Second, we take into account the temporal similarity in the merging stage. Let $\mathrm{DFD}_i(v_i)$ and $\mathrm{DFD}_i(v_j)$ be the displaced frame differences (DFD) of the i-th region by the motion vector for the i-th region and that for the j-th region, respectively. For the regions surrounding the i-th region, we examine that

$$|\,\mathrm{DFD}_i(v_i) - \mathrm{DFD}_i(v_j)| < Thr \quad (4)$$

where Thr is the prespecified threshold. Among those regions that satisfy (4), we merge the i-th region into the j-th region if two regions are combined to the same region in the previous frame, or the j-th region yields smaller difference in (4) than any other regions. Here, it is noted that, by considering the segmentation results in the previous frame, the connectivity of objects between frames can be provided by the proposed technique.

III. SEGMENT-BASED MOTION COMPENSATION

The conventional motion compensation techniques employ a simple translational motion model, in which the motion is described by the horizontal and vertical components $(t_x, t_y)^T$. However, since it is very common that the motion compensaton is performed on relatively large objects, in our approach, we employ rather complex motion model that can describe the rotation and scaling, as well as the translation. The motion model can be described by

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = a \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} t_x \\ t_y \end{pmatrix} \quad (5)$$

where $(t_x, t_y)^T$ is the translational components, θ is rotational component, and a is a scale, respectively. As a distance measure, we choose to use the mean absolute difference (MAD), given by

$$MAD = \frac{1}{N} \sum_{(x,y)\in \text{region}} |I^C(x, y) - I^P(x', y')| \quad (6)$$

where N is the number of pels in a region. And, I^C and I^P denote the current frame and the previous frame, respectively. For each regions, we attempt to find the parameter set (t_x, t_y, θ, a), which yields the minimum MAD among those sets within the search range provided in Table 1. In this paper, a full search method is employed to obtain the best performance.

	translation $(t_x, t_y)^T$	rotation θ	scaling a
range	-16～+15 step 0.5	-8～+7 step 0.002°	-8～+7 step 0.2

Table 1 Example of parameter search range

IV. CODING OF INTERESTING REGIONS

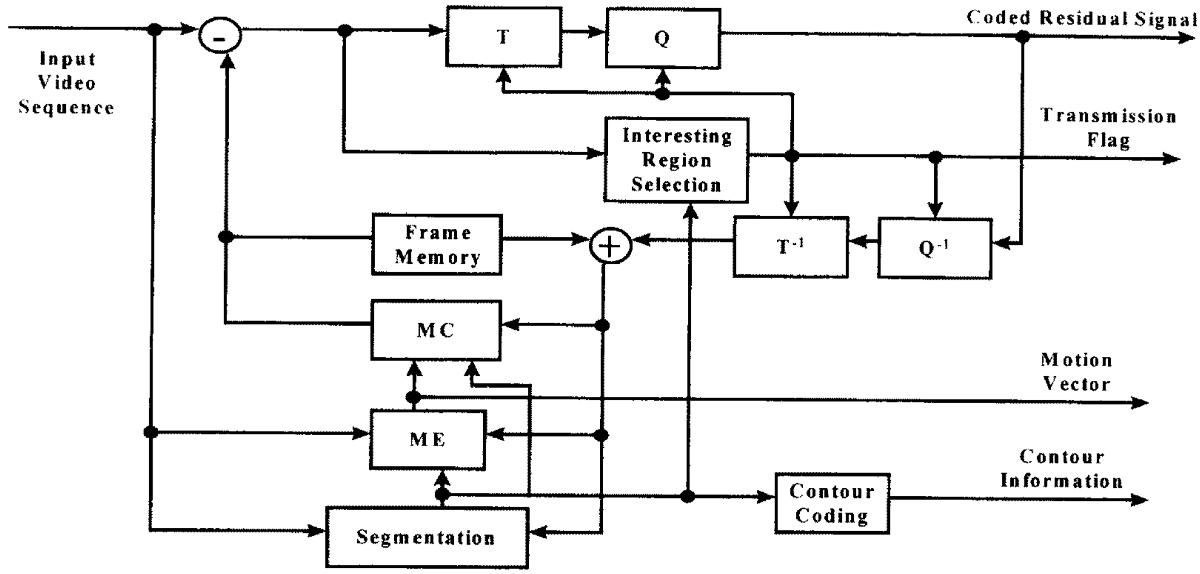


Fig. 2 A block diagram of the segment-based video coder.

In Fig. 2, we show a block diagram of the segment-based coder employing the proposed segmentation technique. In the object oriented coding technique, only the model failure regions are required to be transmitted in terms of the colors as well as model parameters, which consist of the motion and shape, while the model compliance regions are represented by only the model parameters.

In our approach, we employ the notion of the interesting regions proposed by Thierry[11]. Instead of providing similar level of performance on the entire image, the interesting regions are extracted. Then, the interesting regions are encoded more faithfully than other regions, resulting in an improved subjective quality. In the videophone applications, the human face in the center of the image would be an example of the interesting regions. In our approach, for the sake of simplicity, we determine the interesting regions, in terms of the image border criterion and the variance of the MC errors. More specifically, we classify the intersting regions into the regions in the center of the image and those having large MC errors. However, in our approach, the size of the interesting regions are restricted to be less than a half of the input image size. While the prediction errors of the interesting regions are encoded by using transform- based coding technique, those of the other regions are encoded simply by their average.

V. SIMULATION RESULTS

In order to evaluate the performance, we test the proposed technique on the CIF color image sequences, such as Claire and Foreman.

First, Fig. 3 shows examples of the segmented images, which is obtained by splitting the original color images using the K-means alogirhtm. It is observed, in Fig. 3, that there are many small isolated regions and long strips along the edges. Notice that those regions are irrelevant in the coding view point, since a large amount of bits is required for encoding those regions.

The segmenation results by the proposed technique are shown in Fig. 4. The advantage of the proposed technique can be demonstrated by comparing Fig. 4 with Fig. 3. As shown in Fig. 4, many small isolated regions, incurred by the noise, or irrelevant spatial details are effectively removed by the proposed technique. Moreover, by merging the neighbouring region considering edges, motion, and the results of the previous frame, we can significantly reduce the long strip rigions along the edges, while extracting the objects correctly.

In Fig. 4. the interesting regions are represented by shaded regions. As shown in Fig. 4(a), the interesting regions of the Claire include the woman's face, as is often the case with the videophone applications. However, on the Foreman in Fig. 4(b), the interesting regions include the background in the middle of the image, as well as the man's face. This result is due to the fact that the background also contains rather complex motion in this case. However, the background can be removed from the interesting regions if we consider the interframe relations such as hit score[11] as the criteria. The average PSNR performance of the coder is summarized in Table 2, in which the proposed technique yields good PSNR results cosidering its low bit rates.

Ⅵ. CONCLUSION

In this paper, we proposed a segmentation technique for the object-oriented coding, based on the split-and-merge. By noting that, in the object-oriented coding technique, each objects are represented by 3 parameters such as shape, motion, and color informations, we attempted to fully exploit the 3 parameters by the proposed technique. To achieve this goal, the proposed segmentation technique consisted of 4 stages; color space conversion, noise reduction, splitting and merging. Starting with the color space conversion, in order to smooth out the noise and irrelevant spatial details, noise reduction technique by the median filter and morphological filter was performed. Then, while the splitting was performed using the color information, the merging was achieved by the shape and motion informations. More specifically, the input image is divided into many small regions by the K-means algorithm on the O-K-S color space. Then, each regions were merged, according to the spatio-temporal similarity measured by the edge and the change in the DFD. In simulations, the performance of the proposed technique was examined on the real image sequence, and it was shown that the proposed technique provided good segmentation results. In addition, in order to evaluate the coding performance of the proposed technique, we introduced the notion of the interesting regions, and provided the results of encoding the image with emphasizing the interesting regions.

Ⅶ. REFERENCES

[1] H.G. Musmann, M. Hötter, and J. Ostermann, "Object-oriented analysis-synthesis coding of moving images", *Signal Processing: Image Communication 1*, pp. 117- 138, 1989.

[2] W. Li, M. Kunt, "Morphological segmentation applied to displaced frame difference coding", *Signal Processing* 38, pp.45-56, 1994

[3] H. Jozawa, "Segment-based video coding using an affine motion model", *SPIE* Vol. 2308, pp. 1605-1614.

[4] L. Wu, J. Benois-Pineau, Ph. Delagnes, D. Barda, "Spatio-temporal segmentation of image sequences for object-oriented low bit-rate image coding", *Signal Processing:Image Communica tion 8*, pp. 513-543, 1996.

[5] C. Gu, *Multivalued Morphology And Segmentation-based Coding*, Ph.D Thesis, EPFL, 1995.

[6] Y.-W. Lim, *A Study on the Color Image Segmentation Algorithm Using the Fuzzy Set Theory*, MS Thesis, Seoul National University, 1988.

[7] A. Chiari, S. Miceli, "Conversion of video signals from YUV to RGB representations in the digital domain", *Proc. the 2nd Int'l Workshops on Signal Processing of HDTV*, L'Aquila, Italy, March, 1988.

[8] J.S. Lim, *Two-Dimensional Signal and Image Processing*, Prentice-Hall, Inc., 1990.

[9] D. Cortes, et al., "Image segmentation towards new image representation methods," *Signal Processing: Image Communication 6*, pp. 485 - 498, 1995.

[10] R.O. Duda and R.E. Hart, *Pattern Classification and Scene Analysis*, New York, NY, Wiley, 1985.

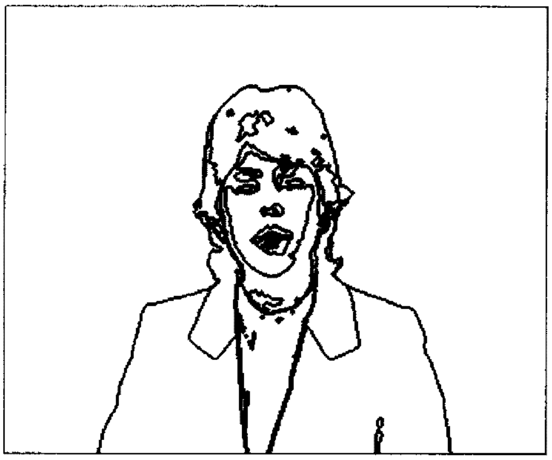
[11] D. Thierry, "Interesting regions determination", MPEG Doc. No. 372, Nov. 1995.

[12] H. Li, A. Lundmark, and R. Forchheimer, "Image sequence coding at very low bitrates: A review," *IEEE Trans. on Image Processing*, vol.

3, no. 5, pp. 589-609, Sept. 1994.
[13] Y. Ohta, T. Kanade, and T. Sakai, "Color information for region segmentation", *Computer Vision, Graphics and Image Processing*, Vol. 13, pp. 224-241, 1980.

Table 2 Average results of the coder employing the proposed technique.

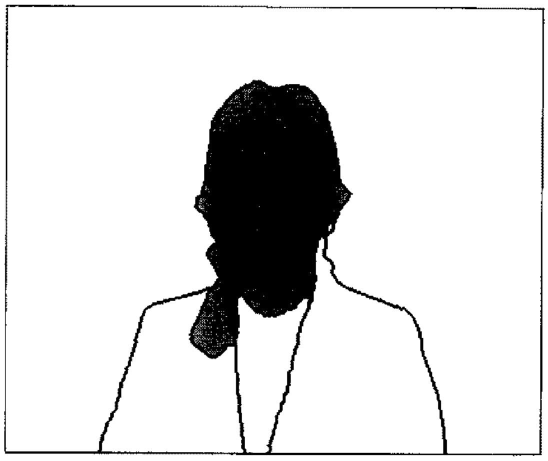
Sequence	Average PSNR [dB]	Average Bitrate [Bits/Sec]
Claire	32.02	~ 28000
Foreman	22.03	~ 86000



(a) Claire

(b) Foreman

Fig. 3 Examples of the segmented Claire and Foreman by the K-means algorithm.



(a) Claire

(b) Foreman

Fig. 4 Examples of the segmented images by the proposed technique.