논문 99-24-9B-19

한국통신학회논문지 '99-9 Vol.24 No.9B

# Region-based H.263 Video Codec with Effective Rate Control Algorithm for Low VBR Video

Hwangjun Song* ***Regular Member***

## ABSTRACT

A region-based video codec based on the H.263+ standard is examined and its associated novel rate control scheme is proposed in this work. The region-based coding scheme is a hybrid method that consists of the traditional block DCT coding and the object-based coding. Basically, we adopt H.263+ as the platform, and develop a fast macroblock-based segmentation method to implement the region-based video codec. The proposed rate control solution includes rate control in three levels: encoding frame selection, frame-layer rate control and macroblock-layer rate control. The goal is to enhance the visual quality of decoded frames at low bit rates. The efficiency of proposed rate control schemes applied to the region-based video codec is demonstrated via several typical test sequences.

## I. Introduction

The video codec design plays an important role in the development of a visual communication system. The block-based DCT coding approach is widely used in video compression standards such as MPEG-1/2 and H.26x while the object-based approach is adopted in the new video coding standard MPEG-4. The object-based coding allows a more flexible choice to allocate bit rates in different regions of an image sequence. H.263+[1] is an emerging video compression standard for low-bit rate video communication and expected to play a key role in Internet video transmission. In this work, we examine a new region-based video coding scheme based on H.263+. The region-based H.263+ coding scheme is a hybrid method that consists of the traditional block DCT approach and the object-based approach. Simply speaking, we adopt H.263+ as the platform, and develop a fast macroblock-based segmentation method to implement the region-based video codec which separates the moving object from the background. The rate control unit in a video codec serves dual functions at the same time, i.e. regulating the compressed bitstream according to channel conditions and enhancing compressed video quality under various buffer and channel constraints. Thus, rate control algorithms should be designed by considering characteristics of communication channels and underlying video. Novel rate control scheme will be studied for the region-based H.263 video codec for low VBR video.

A video codec has sometimes to sacrifice spatial and/or temporal quality to meet the bit budget requirement and channel conditions. Blocking, ringing and texture deviation artifacts often appear in low bit rate video as a result of spatial quality degradation. Flickering (or blinking) and motion jerkiness are major artifacts observable due to temporal quality degradation. The flickering artifact is caused by the fluctuation of spatial

---

* The author is with the Integrated Multimedia Systems Center and the Department of Electrical Engineering-Systems, University of Southern California, Los Angeles, California 90089-2564. (hwangjun@sipi.usc.edu)

논문번호 : 99255-0628, 접수일자 : 1999년 6월 28일

image quality between adjacent frames while motion jerkiness occurs when there is an abrupt change of the coding frame rate or when the frame rate goes below a certain threshold. A considerable amount of effort has been devoted to compression artifact reduction at the decoder end. For example, blocking and ringing artifacts can be significantly reduced by postfiltering (e.g. the deblocking filter in H.263+), and motion jerkiness can be improved via frame interpolation. At the encoder end, the overlapped block motion compensation (OBMC) technique [2] can be used to improve the spatial qualit y at the expense of a higher computational complexity. In this work, we consider the use of rate control to achieve the same goal. That is, we seek an efficient trade-off between spatial and temporal quality to enhance the perceived visual quality under a given channel condition.

We present an efficient three-level rate control algorithm that improves human visual perceptual quality of low bit rate video. The three-level rate control algorithm includes the encoding frame selection, the frame-layer rate control and the macroblock-layer rate control. Compared with our previous work in [5], there is a unique feature in results in this work. That is, we consider a macroblock-based segmentation algorithm based on digital image processing techniques [6, 7] to result in a region-based H.263+ codec. Since it is feasible to choose the coding mode for each macroblock in H.263+ independently, the resulting bit stream is compatible with the H.263+ standard.

This paper is organized as follows. A macroblock-based moving region segmentation algorithm is proposed in Section 2, the frame-layer rate-distortion (R-D) model is studied in Section 3, which serves as the basis for the design of the rate control algorithm in the following section. Rate control algorithm is studied in Section 4. Experimental results are presented in Section 5. Finally, concluding remarks are given in Section 6.

## Ⅱ. Segmentation of Moving Region and Still Background

Compared with the traditional block-based approach such as MPEG-1,2 and H.26x, object-based video coding can potentially improve the perceived visual quality by assigning more bit rates to moving objects or regions of interest. MPEG-4 has been established based on this idea. There has been a large amount of research work in efficient object segmentation, including the spatial domain approach [9], the temporal domain approach [9] and the spatio-temporal domain approach [10, 11]. The spatial-domain approach segments a given image into several regions by using spatial homogeneity. This approach tends to generate too many fine regions which do not correspond to real objects. The temporal-domain approach utilizes motion information. The spatio-temporal domain approach employs both spatial and temporal information. Object segmentation is in general difficult to apply for real-time video applications due to the high computational complexity required.

Recently, several approaches that simplify the existing block-based video coding scheme with some approximations were proposed in [12, 13, 14, 15]. A simple change region detection was employed in [12], and a simple geometric face model using the shape of ellipse and rectangle was proposed in [13]. In [14], Fukuhara *et al.* proposed a coding scheme by combining H.263 and block partitioning with patterns such as vertical, horizontal, 45-degree and 135-degree diagonal edges. The neural network technology was employed in [15] to detect the region of interest to achieve MPEG-1 video segmentation. However, the computational complexity for the last three approaches is still high. Since the human visual system (HVS) is more sensitive to moving regions, we would like to improve the visual quality of moving regions. In H.263+, we can choose different coding modes for each macroblock, e.g. the intra-coded mode, the inter-coded mode with 1 motion vector, the inter-coded mode with 4 motion vectors, and uncoded. To be compatible with H.263+, our segmentation algorithm uses the macroblock as an elementary unit.

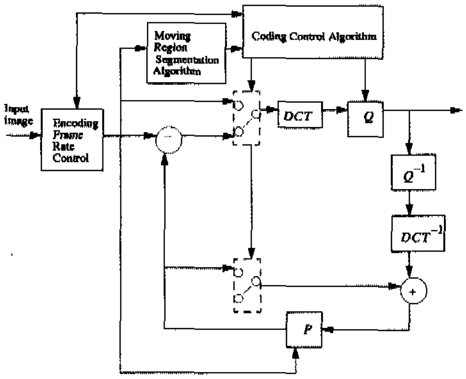


Fig 1. Black diagram of the proposed region-based coding algorithm.

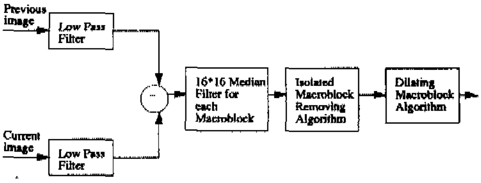


Fig 2. The block diagram of the proposed region segmentation algorithm.

In this research, we consider a simple yet effective moving region segmentation algorithm based on digital image processing techniques [6, 7]. The block diagram of the segmentation algorithm is shown in Fig. 2. By differencing the current frame and the previous frame, we can detect moving and still areas. However, some process is needed to remove background noise to simplify the segmentation result. To reduce the high frequency noise, we pass the input image sequences through a lowpass filter before taking the difference. An example of the resulting difference image is shown in Fig.4(a), where the employed lowpass filter is given in Fig.3. Each macroblock is then replaced by the median value in the difference image so that a preliminary macroblock-based region segmentation result can be obtained. It consists of the still, slowly moving and fast moving areas. The isolated moving macroblock is merged to the still region via simple morphological filtering. The result is shown in Fig.4(b). Finally, we apply the dilation process to remove the boundary discontinuity which tends to degrade perceived quality greatly. Without dilation, boundary discontinuity is occasionally observed in spite of clearer moving objects.

| | | |
|---|---|---|
| $\omega_3$ | $\omega_2$ | $\omega_3$ |
| $\omega_2$ | $\omega_1$ | $\omega_2$ |
| $\omega_3$ | $\omega_2$ | $\omega_3$ |

$$\hat{p}_{i,j} = \frac{1}{\omega}\sum_{k=-1,l=-1}^{k=1,l=1} \omega_{k,l} P_{k,l}$$

$$\omega_{0,0} = \omega_1$$

$$\omega_{-1,0} = \omega_{1,0} = \omega_{0,-1} = \omega_{0,1} = \omega_2$$

$$\omega_{-1,-1} = \omega_{-1,1} = \omega_{1,-1} = \omega_{1,1} = \omega_3$$

$$\omega = \omega_1 + 4\omega_2 + 4\omega_3$$

Fig. 3 An example of 3×3 lowpass filter.

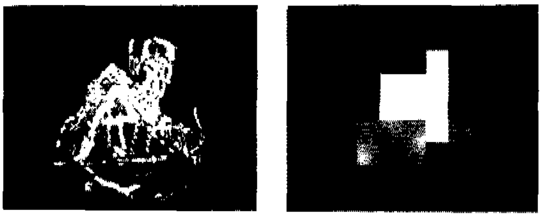

(a) (b)

Fig. 4 The proposed region segmentation process
(a) the difference image between lowpass filtered $197_{th}$ and $199_{th}$ images of QCIF Salesman,
(b) the resulting image after median filtering and isolated macroblock removal.

## Ⅲ. Frame Layer R-D Models

The block diagram of the overall rate control scheme is given in Fig.1. The encoding frame selection scheme is compatible with the H.263+ standard, since it allows the encoder to drop frames when needed. It is designed to provide acceptable temporal quality. The region-based coding scheme and the frame-layer rate control handle spatial quality by segmenting moving regions and adjusting the quantization parameter (QP).

The encoding time delay and the computational complexity are critical to real-time video applications. R-D modeling is important to the derivation of fast rate control algorithms.

Generally speaking, there are two methods to achieve rate-distortion (R-D) modeling: statistical and experimental methods. One commonly used statistical model is to assume that the source signal has the generalized Gaussian distribution. For a Gaussian source, which is a special case of the generalized Gaussian distribution, a closed form of the R-D curve can be found [16]. Other simplified models have also been examined. Some models were derived experimentally, e.g. the quadratic rate model [17], the exponential model [18], the spline approximation model [19] and the normalized rate-distortion model [20], etc. Even though statistical models demand a lower computational complexity than experimental models, experimental models can provide a more accurate model through a data fitting process. TMN8 [21] adopts a statistical R-D model for the coding of a macroblock. This approach is however too simple to characterize statistics of a frame by estimating only the variance of a frame [21].

In this paper, the frame-layer R-D model is obtained with coding results based on the existing H.263+ macroblock layer rate control algorithm. To be more specific, we derive the R-D model *w.r.t.* the average quantization parameter (QP) in a frame, where the macroblock layer rate control algorithm of TMN8 is adopted as an auxiliary control component. We examine distortion measures in Section 3.1 and then derive the R-D models in Section 3.2.

### 3.1 Distortion Measure

Human visual perceptual quality is very complex. For example, not all frequencies in an image have the same importance with respect to human perceptual characteristics [22, 23]. Up to now, the mean square error (MSE) has been widely employed as a spatial distortion measure although it does not correlate to the human perceived quality exactly. Obviously, MSE is not a proper distortion measure in the region-based video coding. In this work, we consider a weighted MSE of a low computational complexity. It is defined as:

$$D_w = \frac{1}{w} \sum_{i=1}^{N_{wd}} \sum_{j=1}^{N_{hgt}} \mu_{i,j}(p_{i,j} - \widehat{p_{i,j}})^2, \tag{1}$$

$$\omega = \sum_{i=1}^{N_{wd}} \sum_{j=1}^{N_{hgt}} \mu_{i,j} \tag{2}$$

where $N_{wa}$ and $N_{hgt}$ are pixel numbers of the width and the height, $p_{i,j}$ and $\widehat{p_{i,j}}$ are pixel values of the original image and the reconstructed image, respectively, and

$$\mu_{i,j} = \begin{matrix} \mu_m \, if \, (i,j) \in \text{movingregion}, \\ \mu_s \in \text{otherwise} \end{matrix}$$

Generally speaking, it is more reasonable to set $\mu_m \geq \mu_s$ by considering the human visual effect. Subjective measure may also be required to determine the proper values of $\mu_m$ and $\mu_s$.

### 3.2 Rate-distortion Modeling

In this section, we examine a frame layer R-D modeling approach which constructs both the rate and distortion models with respect to the averaged quantization parameter (QP) of all moving regions in each frame. In terms of mathematics, the rate and distortion models can be written, respectively, as:

$$\widehat{R}(\overline{q}) = (a\overline{q^{-1}} + b\overline{q^{-2}})M_{w(f_{ref},\ f_{cur})}, \tag{4}$$

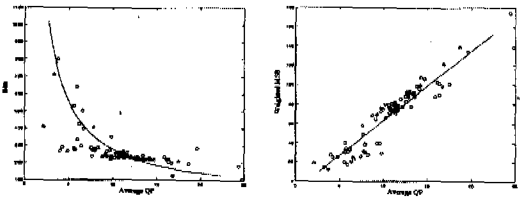
$$\widehat{D_w}(\overline{q}) = a'\overline{q} + b', \tag{5}$$

where $a$, $b$, $a'$ and $b'$ are model parameters, $f_{ref}$ is the reconstructed reference frame at the previous time instance, $f_{cur}$ is the original image at the current time instance, $\overline{q}$ is the average QP of all macroblocks in a frame and

$$M_w(f_{ref}, f_{cur}) = \frac{1}{\omega} \sum_{i=1}^{N_{wd}} \sum_{j=1}^{N_{hgt}} \mu_{i,j} \mid \widehat{p_{i,j}^{ref}} - p_{i,j}^{cur} \mid,$$

where $\widehat{p_{i,j}^{ref}}$ and $p_{i,j}^{cur}$ are pixels in the reconstructed reference frame and the current frame, respectively. Note that $M_w(f_{ref}, f_{cur})$ takes into account the dependency among frames. Coefficients *a*, *b*, *a'* and *b'* are determined by using the linear regression method.

Conventionally, the R-D curve is computed based on integer QPs. In our case, $\bar{q}$ can be a floating-point number since $\bar{q}$ is the average QP of all macroblocks in a frame. We use an outlier removal process to improve the model accuracy as done in MPEG-4 Video Verification Model version 10.0 [24]. That is, if the difference between a data point and the derived model is greater than one standard deviation, the datum is removed. Based on filtered data, we can derive the rate and distortion models again. We show the rate and distortion models in Fig.5(a) and (b), respectively, for the QCIF Salesman sequence, where the circle denotes the measured data points while the solid curve is the computed model. As shown in these two figures, the rate and distortion models work reasonably well. The R-D modeling approach presented above provides a good approximation of the rate and distortion performances w.r.t. the average QP for all test sequences in our experiment.



(a) (b)

Fig. 5 Frame layer R-D modeling for the QCIF Salesman sequence :
(a) the rate model and (b) the distortion model as a function of the average QP of macroblocks($\mu_m = 100\mu_s$)

## Ⅳ. Proposed Rate Control Algorithm

For unconstrained VBR, if there are sufficient buffers at both the encoder and the decoder, rate control can be formulated as an optimization problem constrained to the bit budget only. Many video rate control algorithms have been developed under this scenario. Even though it cannot provide the optimal solution under multiple channel constraints, it does provide an operating point where rate and distortion are traded off for a given bit budget. For the bit-budget constrained rate control, we need a basic unit to perform rate control. For example, the group of pictures (GOP) is generally used as a basic unit for rate control in MPEG, which consists of one I-frame and several predictive frames (i.e. P-and B-frames) repeated periodically. Generally speaking, I-frames require a higher rate than predictive frames since motion estimation and compensation are not employed. For a very low bit rate environment, we have to reduce the number of I-frames, and the number of frames in a GOP can become larger. This explains why existing MPEG-1/2 frame layer rate control algorithm can not be straightforwardly extended to H.263+.

Up to now, there is little work about frame layer rate control for H.263+ even though several macroblock layer rate control algorithms were proposed before [25, 26, 27, 20, 28, 29, 30]. Especially, since TMN8 rate control focuses on CBR channels and low-latency, frame layer rate control is not needed. However, frame layer rate control is required for efficient coding for VBR video. In this work, we define a GOP as a *group of predictive frames* without I-frame. It is not efficient to consider I-frames and predictive frames simultaneously in H.263+. This is a general trend. The bit rate constraint is satisfied in each GOP. The proposed rate control scheme uses this new GOP as a basic rate control unit.

The proposed rate control algorithm consists of three layers: encoding frame selection, frame layer rate control and macroblock layer rate control. As mentioned earlier, the proposed rate control algorithm includes the existing H.263+ macroblock layer rate control algorithm as a tool. In our previous work [5], we considered an encoding frame rate control which adjusts the frame rate by detecting motion change in video. This approach focused on the spatial quality of frames without noticeable motion unsmoothness. We reduced the encoding frame rate in the fast motion interval to keep the spatial quality in the tolerable range. However, it does not help the decoder to interpolate skipped frames.

Efficient postprocessing techniques at the decoder end have been studied intensively. One of them is the frame interpolation technique. Due to the availability of powerful hardware, they are feasible to implement in a practical system. In the following discussion, we assume that the decoder has the frame interpolation capability to improve video quality. We will propose an encoding frame selection algorithm that is friendly to the decoder with frame interpolation capability in Section 4.1, then a frame layer rate control scheme with a low computational complexity in Section 4.2.

### 4.1 Encoding Frame Selection w.r.t. Decoders with Frame Interpolation Capability

Motion compensated frame interpolation at the decoder end depends on motion change between adjacent encoded frames. If we adopt a fixed encoding frame rate, motion change between adjacent encoded frames is not constant. Hence, it is difficult for the decoder to interpolate the skipped frames for fast and non-uniform motion change. The quality of interpolated frames can be poor. In this work, we propose an encoding frame selection algorithm with a low computational complexity based on the observation that motion change is greatly related to the decoder's frame interpolation capability in video conferencing. To detect motion change, we adopt *HOD* (the histogram of difference) since it is sensitive to local motion.

- If $HOD(f^{org}_{ref}, f_{cur}) \geq TH_1$, encode the current frame.
- If $HOD(f^{org}_{ref}, f_{cur}) < TH_1$, skip the current frame.

where $f^{org}_{ref}$ is the last encoded original frame, $f_{cur}$ is the current original image, $TH_1$ is the threshold value, and *HOD* is defined as

$$HOD(f_{n}, f_{m}) = \frac{\sum_{i > |TH_0|} hod(i)}{N_{pixel}}, \tag{6}$$

where $i$ is the index of the quantization bin, *hod(i)* is the histogram of the difference image, $TH_0$ is the threshold value for detecting the closeness of the position to zero, and $N_{pixel}$ is the number of pixels. More frames are encoded in fast motion intervals than slow motion intervals. As a result, motion change between adjacent encoded frames is almost the same. This fact helps the decoder to interpolate skipped frames successfully. In Figs. 6 and 8, we show the trace of the HOD value and encoded frame positions of the QCIF Salesman and Silent Voice sequences. Even though the decoder does not have a smart frame interpolation capability, the proposed encoding frame selection algorithm can still improve video quality by simple frame repetition (i.e. *intrafiltering*).

### 4.2 Frame Layer Rate Control of Low Complexity

After encoded frame positions are determined, we should perform frame-layer and macroblock-layer rate control to allocate bits to a selected frame and its associated macroblocks. For macroblock layer rate control, we adopt the efficient algorithm proposed in [26] as a component in our overall scheme. For frame-layer rate control, we consider an algorithm of a low computational complexity and a low encoding time-delay in this section. The frame layer rate control problem can be formulated as follows.

Problem Formulation : Determine $\overline{q_i}$, $i=1,2,\cdots,N_{gop}$, to minimize

$$\sum_{i=1}^{N_{gop}} \widehat{D_{w,i}}(\overline{q_i}), \tag{7}$$

subject to

$$\sum_{i=1}^{N_{gop}} R_i \leq B_{gop}, \tag{8}$$

where $N_{gop}$ is the encoded frame number in a GOP, $\widehat{D_{w,i}}$ is the estimated distortion of the $i_{th}$ frame, and $R_i$ is the bit budget for the $i_{th}$ frame. The Lagrange multiplier method has been widely employed for bit rate allocation in video coding [31,32,33]. However, to find the optimal Lagrange multiplier usually requires a high computational complexity and leads to a long encoding delay. To simplify the search process, adaptive algorithms for Lagrange multiplier selection were examined in [34,35,27]. In this work, we consider a suboptimal scheme that consists of two steps. They work iteratively to reduce the computational complexity and encoding delay. The two steps are described below.

*Step 1 : Optimization with R-D models.*

By using the Lagrange multiplier method, we can define a penalty function for the $i_{th}$ frame by combining the cost function and the constraint through the Lagrange multiplier, i.e.

$$P_k(\overline{q_k}) = \widehat{D_{w,k}} + \lambda_k \max\{0, \widehat{B_k^{res}}\}, \tag{9}$$

$$\widehat{B_k^{res}} = \sum_{i=1}^{k-1} R_i + \widehat{R_k}(\overline{q_k}) - k \cdot R_{target}, \tag{10}$$

$$R_{target} = \frac{B_{gop}}{N_{gop}}, \tag{11}$$

where $P_k(\overline{q_k})$ is the cost function for the *k*th frame and $\lambda_k$ is the Lagrange multiplier for the *k*th frame. It is assumed that a given bit rate is assigned to every frame equally. Based on the rate and distortion models in Section 3, we can determine the optimal QP to minimize the above penalty function. We can get the optimal solution by using the gradient method under the convex hull assumption:

$$\overline{q_k^*} = arg \min_{\overline{q_k}} P_k(\overline{q_k}). \tag{12}$$

What we actually need is not $\overline{q_k^*}$ but the target bit budget $\widehat{R_k}(\overline{q_k^*})$ for the *k*th frame.

*Step 2 : Lagrange Multiplier Adaptation*

We adopt an adaptive Lagrange multiplier selection rule based on the LMS method[27]. The Lagrange multiplier is updated based on residual bits after the coding of a frame. The adaptation algorithm is:

$$\lambda_{k+1} = \lambda_k + \Delta\lambda_k, \tag{13}$$

$$\Delta\lambda_k = \frac{B_{used,k}}{B_{target,k}} - 1, \tag{14}$$

where $\lambda_k$ is the Lagrange multiplier for the $k_{th}$ frame,

$$B_{used,k} = \sum_{i=1}^{k} R_i, \tag{15}$$

$$B_{target,k} = R_{target} \cdot t_k. \tag{16}$$

Finally, the macroblock layer rate control in TMN8 is employed in the overall rate control scheme.

## V. Experimental Results

In the experiment, the macroblock layer rate control in TMN8 [26] is employed, and the implementation is based on the UBC H.263+ source code [36]. We use the deformable block-based frame motion-compensated interpolation proposed in [37]. The performance comparison is made based on subjective as well as objective evaluation, with more emphasis on subjective evaluation.

It is assumed that the I-frame is encoded at a predetermined bit rate. The quality of the I-frame can affect the overall objective gain since the error of the background is not updated in the region-based coding scheme. However, the bit rate for the I-frame is related to time-delay. H.26L Evaluation Delay Model User Guide recommends that the bit rate for the I-frame must not be greater than one second worth of bit transmission at the assumed channel bit rate. For example, for 24 kbps channel, the bit rate for the I-frame can not exceed 24 kbits. In this experiments, the

I-frame is encoded with QP = 15, which satisfies the above recommendation.

Now, experimental results are presented to demonstrate the performance of the proposed region-based coding scheme and rate control scheme for low VBR video in comparison with TMN8. The test sequences are QCIF "'Salesman" and "Silent Voice" and the average target bit rate is 24 kbps.

*$TH_1$* is determined by analyzing GOP with the initial frame skip interval. That is, *$TH_1$* is set to average *HOD* in a GOP. *$TH_1$* is set to 0.005695 for *Salesman* and 0.019030 for *Silent Voice*, which are the average *HOD* value at 15 fps. We have Figs. 6 and 8, and Table 2 for QCIF Salesman and Silent Voice. In (6), *$TH_1$* is set to 32, and one GOP consists of 200 frames. In Figs. 6 and 8, we observe the *HOD* value trace and encoding frame position. The congestive areas in these figures indicate the fast motion change interval while sparse areas denote the slow motion change interval. It is observed in Fig.6 that *HOD* is decreasing in frame no. 70-80. It is difficult for the decoder to interpolate skipped frames because of non-uniform motion change. In this case, we have to encode a frame. We can see the exactly encoded frame positions in Table 2. More frames are encoded in the fast motion change interval than the slow motion change interval. As shown in Table 1, the proposed encoding frame selection algorithm decreases the standard deviation of HOD values significantly.

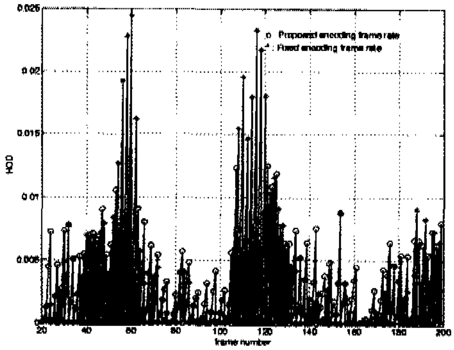
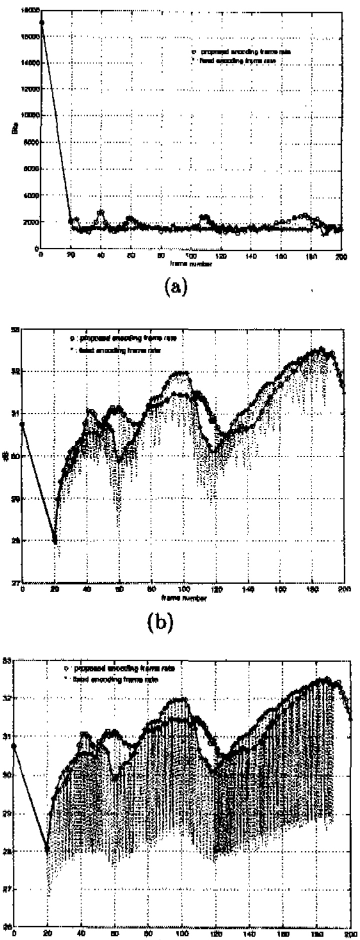


Fig. 6 Encoded frame positions and HOD plot of QCIF Salesman sequence.



(a)

(b)

(c)

Fig. 7 (a) Rate plot, (b) PSNR plot of Intrafiltering and (c) PSNR plot of motion compensated frame interpolation of QCIF Salesman under unconstrained VBR.

The performances of coded video are shown in Figs. 7 and 9. The output bit rate plots are shown in Figs.7(a) and 9(a), and PSNR plots are shown in Figs.7(b) and 9(b).

Table 1. HOD comparison with fixed frame rate and proposed encoding frame selection algorithm which is friendly to decoder with frame interpolation capability.

| Method | Sequence | Avg HOD | STD of HOD |
|---|---|---|---|
| TMN8 | Salesman | 0.0046 | 0.00003525 |
|  | Silent Voice | 0.0179 | 0.00008252 |
| Proposed frame rate control | Salesman | 0.0058 | 0.00000847 |
|  | Silent Voice | 0.0202 | 0.00003206 |

Table 2. The encoded frame positions for encoding frame selection algorithm friendly to decoder with frame interpolation capability.

| Sequence | Encoded frame positions |
|---|---|
| Salesman | 20 24 27 30 32 34 37 39 41 43 45 47 49<br>51 53 54 55 56 57 58 59 60 61 63 66 69<br>72 78 82 85 91 96 102 105 107 108 109<br>110 111 112 113 114 115 116 117 118 119<br>121 123 125 127 129 131 134 136 139 143<br>148 150 153 155 160 166 171 175 177 180<br>184 187 189 191 193 195 197 199 |
| Silent Voice | 20 23 25 26 27 28 31 33 35 37 41 42 43<br>44 46 49 51 53 54 55 57 60 62 64 65 66<br>67 69 72 74 77 79 81 84 86 88 94 97 100<br>103 106 107 108 109 110 111 113 115 117<br>119 121 124 126 128 130 133 138 141 147<br>152 155 157 159 161 164 166 168 170 175<br>177 179 181 184 186 188 190 192 194 196<br>197 198 199 |

Table 3. Performance comparison with TMN8 under unconstrained VBR when Intrafiltering frame interpolation is employed. Target average rate is 24kbps.

| Method | Sequence | Avg PSNR | STD of PSNR |
|---|---|---|---|
| TMN8 | Salesman | 30.7998 | 1.0217 |
| | Silent Voice | 29.4500 | 1.2118 |
| Region \&proposed frame rate control | Salesman | 30.7055 | 0.8081 |
| | Silent Voice | 29.2233 | 1.1914 |

Table 4. Performance comparison with TMN8 under unconstrained VBR when motion compensated frame interpolation is employed. Target average rate is 24kbps.

| Method | Sequence | Avg PSNR | STD of PSNR |
|---|---|---|---|
| TMN8 | Salesman | 29.6516 | 1.5950 |
| | Silent Voice | 28.3928 | 2.0304 |
| Region \&proposed frame rate control | Salesman | 29.3612 | 1.4970 |
| | Silent Voice | 28.0699 | 2.0699 |

The statistical data are summarized in Table 3 and 4. It is observed that the proposed approach can decrease the PSNR fluctuation although the average PSNR is almost the same for both sequences with frame interpolation. Since the proposed region-based approach does not update the error in the background, the error is propagated to the following frames. It can decrease the PSNR gain. This phenomena is more obvious for Salesman than Silent Voice since the background of Salesman is relatively more complicated. However, it is not important visually. We cannot use enough bits for the coding of I-frames due to the increase in latency. Even though the PSNR gain is not good enough, subjective quality is obviously improved. However, it is observed that the proposed region segmentation algorithm does not work as well for Foreman as for Salesman and Silent Voice since the background of the Foreman sequence is moving. Hence, the performance of region-based coding is almost the same as the traditional approach. In this case, a more complicated segmentation algorithm is required. In addition, the proposed algorithm improves motion smoothness than a fixed encoding frame rate when the frame interpolation technique is employed at the decoder. The ghost artifact appears and motion is not smooth where motion change is fast, e.g. the hand part in Silent Voice. The phenomenon is obvious even under a fixed encoding frame rate (TMN8,15fps). However, the proposed encoding frame selection designed based on motion change information can reduce these phenomena. As a result, human visual perception is improved. We conclude that the proposed encoding frame selection algorithm helps the decoder to interpolate skipped frames. Even though the decoder does not have an advanced frame interpolation capability, temporal quality is still improved by frame repetition (intrafiltering).

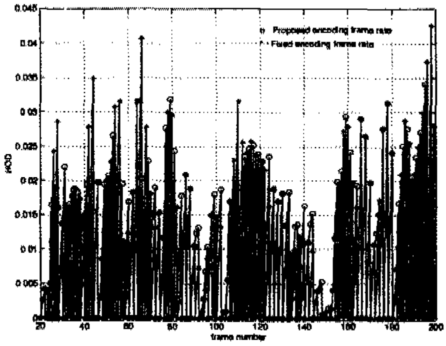
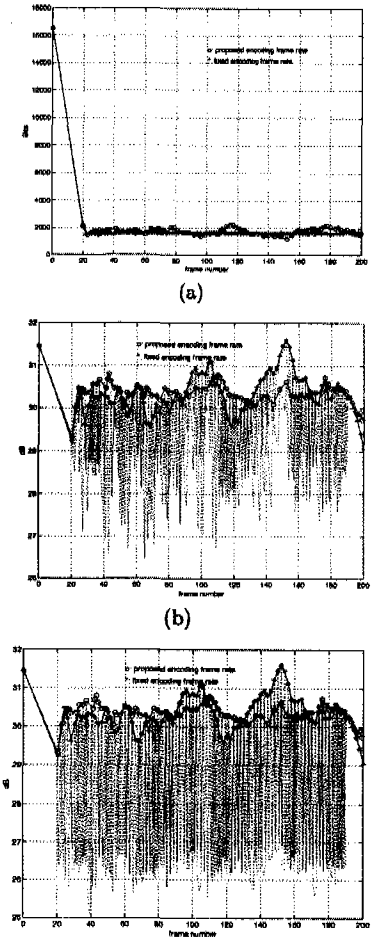


Fig. 8 Encoded frame positions and HOD plot of QCIF Silent Voice sequences



(a)

(b)

(c)

Fig. 9 (a) Rate plot, (b) PSNR plot of Intrafiltering and (c) PSNR plot of motion compensated frame interpolation of QCIF Silent Voice under unconstrained VBR

## Ⅵ. Conclusion

Region-based video coding for video conferencing with a three-layer rate control scheme is proposed. The proposed region-based video coding scheme is a hybrid block- and object-based coding scheme. For rate control, we treat the encoding frame interval (or rate) as a control variable. For low VBR video, we proposed an encoding frame selection scheme which is friendly to the decoder with a frame interpolation capability and a frame-layer rate control algorithm, which has a low computational complexity and coding latency. It was observed that the proposed rate control scheme can improve the spatial/temporal qualities for low VBR video. However, its detailed implementation and performance analysis require further research in the future.

## References

[1] ITU-T, "Video coding for low bitrate communication," ITU-T Recommendation H.263 Version 2, January 1998.

[2] M.T.Orchard and G.J.Sullivan, "Overlapped block motion compensation: an estimation-Threoretic approach," *IEEE Trans. on Image Processing,* vol. 3, pp. 693-699, Sept. 1994.

[3] T. V. Lakshman, A. Ortega, and A. R. Reibuman, "VBR video: Tradeoffs and potentials," *Proceeding of the IEEE,* vol. 86, pp. 952-973, May 1998.

[4] M. Grossglauser, S. Keshay, and D. C. Tse, "RCBR : A simple and efficient service for multiple time-scale traffic," *IEEE/ACM Trans. on Networks,* vol.5, pp. 741-755, Dec. 1997.

[5] H. Song, J. Kim, and C. C. J. Kuo, "Real-time encoding frame rate control for H.263+ video over the internet," *Signal Processing : Image Communication,* vol. 15, to be appeared 1999.

[6] W. K. Pratt, *Digital image processing.* Wiley Interscience, 1991.

[7] A. K. Jain, *Fundamentals of digital image processing.* Prentice-Hall, 1989.

[8] P. William, T. Reed, and M. Kunt, "Image sequence coding by split and merge," *IEEE Trans, on Communications,* vol.39, pp. 1845-1855, Dec. 1991.

[9] H.G. Musman, M. Hotter, and J. Ostermann, "Object-based analysis-synthesis of moving images," *Signal Processing : Image Communication,* vol. 1, pp, 117-138, Oct. 1989.

[10] J. Benois, L. Wu, and D. Barba, "Joint contour-based and motion-based image

sequences segmentation for TV image coding at low bit rate," in *Proc. of SPIE Visual Communication and Image Processing*, pp. 1074-1085, Sept, 1994.

[11] C. Gu and M. Kunt, "Very low bit-rate video coding using multi-criterion segmentation," in *Proc. of IEEE International Conference on Image Processing*, vol.2, pp.418-422, Nov. 1994.

[12] H. Chen, P. Wu, Y. Lai and L. Chen, "A multimedia video conference system:using region base hybrid coding," *IEEE Trans, on Consumer Electronics*, vol. 42, pp 781-786, Aug. 1996.

[13] J. hartung, A. Jacquin, J. Pawlyk, J. Rosenberg, H. Okada, and P. E. Crouch, "Object-oriented H.263 compatible video coding platform for conferencing applications," *IEEE Jouranl on Selected Areas in Communication*, vol.16, pp. 42-55, Jan. 1998.

[14] T. Fukuhara, K. Asai, and T. Murakami, "Very low bit-rate video coding wihth block partitioning and adaptive selection of two time-differential frame memories," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 7, pp. 212-220, Feb. 1997.

[15] N. Doulamis, A. Doulamis, D. Kalogeras, and S. Kolias, "Low bit-rate coding of image sequences using adaptive regions of interest," *IEEE Trans. on Circuits and Systems for Video Technology*, vol 8, pp.928-934, Dec. 1998.

[16] T. M. Cover and J. A. Thomas, *Elements of information theory.* John Wiley & Sons Inc., 1992

[17] T. Ding and Y.-Q. Zhang, "A new rate contro scheme using quadratic rate distortion model," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 7, pp. 246-250, Sept. 1997.

[18] W. Ding and B. Liu, "Rate control of MPEG video coding and recording by rate-quantization modeling," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 6, pp. 12-20, Feb.

[19] L. J. Lin, A. Ortega, and C. C. J. Kuo, "Rate control using spline interpolated R-D characteristics," in *Proc. of SPIE Visual Communication and Image Processing*, pp.111-122, Mar. 1996.

[20] K. H. Yang, A. Jacquin, and N. S. Jayant, "A normalized rate-distortion model for H.263- compatible codecs and its application to quantizer selection," in *Proc. of IEEE International Conference on Image Processing*, vol. 2, pp. 41-44, Oct. 1997.

[21] ITU-T "Video codec test model, near-term, version 8 (TMN8)," tech. rep., H.263 Adhoc Group, Portland, Jun 1997.

[22] A. ortega and K. Ramchandran, "Rate-distortion for image and video compression," *IEEE Signal Processing magazine*, vol. 15, pp. 23-50, Nov. 1998.

[23] N. Jayant, J. Johnson, and R. Safranek, "Signal compression based on models of human perception," *Proceedings of the IEEE*, Oct. 1993.

[24] Moving Picture Expert Group Video group, "MPEG-4 video verification model version 10.0," draft in progress, ISO/IEC JTC1/SC29/WG11, Feb. 1998.

[25] D. Mukherjee and S. K. Mitra, "Combined made selection and macroblock step adaptation for H.263 video encoder," in *Proc. of IEEE International Conference on Image Processing*, vol. 2, pp. 37-40, Oct. 1997.

[26] J. Ribas-Corbera and S. Lei, "Rate control in DCT video coding for low-delay video communication," *IEEE Trans. on Circuits and Systems for Video Technogy.* vol. 9, pp. 172-185, Feb. 1999.

[27] T. Wiegand, M. Lightston, D. Mukherjee, T.G.Campbell, and S.K.Mitra, "Rate-distortion optimized made for very low bit rate video coding and emerging H.263 standard," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 6, pp. 182-190, Apr. 1996.

[28] K. T. Ng, S. C. Chan, and T. S. Ng, "Buffer contro algorithm for low bit rate video compressionm," in *Proc. of IEEE International Conference on Image Processing*, Sept. 1996.

[29] K. Oehler and J. L. Webb, "Macroblock quantizer selection for H.263 video coding," in *Proc. of IEEE International Conference on Image Processing*, vol. 1, pp.365-368, Oct. 1997.

[30] G. Schuster and A Katsaggelos, "Fast and effcient mode and quantizer selection in the rate and distortion sense for H.263," in Proc. of SPIE Visual Communication and Image Processing, 1996.

[31] A. Ortega, K. Ramchandran, and M. Vetterli, "Optimal trellis-based buffered compression and fast approximation," IEEE Trans. on Image processing, vol. 3, pp.26-40, Jan. 1994.

[32] K. Ramchandran, A. Ortega, and M. Vetterli, "Bit allocation for dependent quantization with application to multiresolution and MPEG video coder," *IEEE Trans, on Image Processing*, vol. 3, pp. 533-545, Sept. 1994.

[33] Y. Shoham and A. Gersho, "Efficient bit allocation for an arbitrary set of quantizers," *IEEE Trans, on Acoustic, Speech, and Signal Processing*, vol. 36, pp. 1445-1453, Sept. 1988.

[34] J. Choi and D. Park, "A stable feedback control of the buffer state using the controlled multiplier method," IEEE Trans. on Image Processing, vol. 3, pp. 546-558, Sept. 1994.

[35] D. W. Lin and J. J. Chen, "Efficient optimal rate-distortion coding of video sequences under multiple rate constraints," in Proc. of IEEE International Conference on Image Processing, vol. 2, pp. 29-32, Oct. 1997.

[36] Image Processing Lab, University of British Columbia, "H.263+ encoder/decoder," TMN (H.263) codec, Feb 1998.

[37] T. Kuo and C.-C. J. Kuo, "Motion-compensated interpolation for low-bit rate video quality enhancement," in *Proc. of SPIE International Symposium on Optical Science, Engineering and Instrumentaion*, May 1998.

**송 황 준(Hwangjun Song)** 정회원

received the B.S and M.S. degrees in 1990 and 1992 from Dept. Control and Instrumentation, Seoul National University, Korea. He has completed his Ph. D. degree in Electrical Engineering-Systems, University of Southern California, Los Angeles, CA, In 1999, he was a research engineer at LG Industrial Lab. in Korea. From 1992 to 1994, he served in Korea Army as a full-time. In 1995, he was a part-time lecturer at college. From 1995 to 1999, he was a research assistant in SIPI (Signal and Image Processing Institute) and IMSC (Integrated Media Systems Center) sponsored by NSF(National Science Foundation). His research interests include multimedia signal processing and communication, image/video compression, digital signal processing, network protocols necessary to implement a functional image/video applications, control system and fuzzy-neural system. Dr. Song received Outstanding Academic Achievement Award and nominated for Outstanding Research Paper Award by University of Southern California.