A Region-Based H.263+ Codec and Its Rate Control for Low VBR Video

Hwangjun Song and C.-C. Jay Kuo, *Fellow, IEEE*

***Abstract*—This paper presents a region-based video codec, which is compatible with the H.263+ standard, and its associated rate control algorithm for low variable-bit-rate (VBR) video. The proposed region-based coding scheme is a hybrid method that incorporates traditional block DCT coding as well as object-based coding. To achieve this, we adopt H.263+ as the platform, and develop a fast macroblock-based segmentation method to implement the new region-based codec. The associated 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 improve human visual perceptual quality at low bit rates. The efficiency of the proposed rate control algorithm applied to the region-based H.263+ codec is demonstrated via several typical test sequences.**

I. Introduction

THE VIDEO CODEC plays an important role in the development of a video 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 emerging video coding standard MPEG-4. Object-based coding allows a more flexible choice to allocate bit rates in different regions of an image sequence. H.263+ [1] is a 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, H.263+ is adopted as the platform, and the region-based video coding scheme that separates the moving object from the background by a macroblock-based morphological process is developed. Furthermore, the rate control unit in a video codec serves dual functions at the same time, i.e., regulating the compressed bit-stream according to channel conditions and enhancing compressed video quality under various buffer and channel constraints. Rate control algorithms should be designed by considering characteristics of communication channels and underlying video, and a corresponding rate control algorithm is studied for the proposed region-based H.263+ codec at low bit rates with these considerations.

A video codec has often to sacrifice spatial and/or temporal quality to meet the bit budget requirement and channel conditions. Blocking, ringing and texture deviation artifacts tend to 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 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 reduce compression artifacts 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 spatial quality at the expense of a higher computational complexity. In this work, we consider the use of rate control to achieve the same goal.

We present a region-based video coding scheme compatible with H.263+ and its efficient three-level rate control algorithm that pursuit an efficient trade-off between spatial and temporal qualities to enhance the perceived visual quality at low bit rates. The block diagram of the proposed system is given in Fig. 1. As shown in the figure, two components are added to the H.263+ baseline encoder: a moving region segmentation algorithm to improve the quality of moving regions of the underlying video and an encoding frame selection algorithm to enhance the quality of interpolated frames. Compared with our previous work in [33], there are several unique features in our current work. First, we consider a macroblock-based segmentation algorithm based on digital image processing techniques [5], [6] 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. Second, we study an effective encoding frame selection algorithm at low bit rates that helps the decoder interpolate skipped frames to improve motion smoothness, which is also compatible with the H.263+ standard since it allows the encoder to drop frames when needed. Finally, an efficient frame layer rate control algorithm with a low computational complexity is proposed.

The rest of this paper is organized as follows. A macroblock-based moving region segmentation algorithm is proposed

Manuscript received May 2, 2001; revised September 10, 2002. This work was supported by the Korea Research Foundation under Grant KRF-2002-003-D00296, the Integrated Media Systems Center, a National Science Foundation Engineering Research, and the Annenberg Center for Communication at the University of Southern California and the California Trade and Commerce Agency. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Hong-Yuan Mark Liao.

H. Song is with School of Electrical Engineering, Hong-ik University, Seoul, Korea. (e-mail: hwangjun@wow.hongik.ac.kr).

C.-C. J. Kuo is with the Integrated Multimedia Systems Center and the Department of Electrical Engineering-Systems, University of Southern California, Los Angeles, CA 90089-2564 USA (e-mail: cckuo@sipi.usc.edu).

Digital Object Identifier 10.1109/TMM.2004.827488

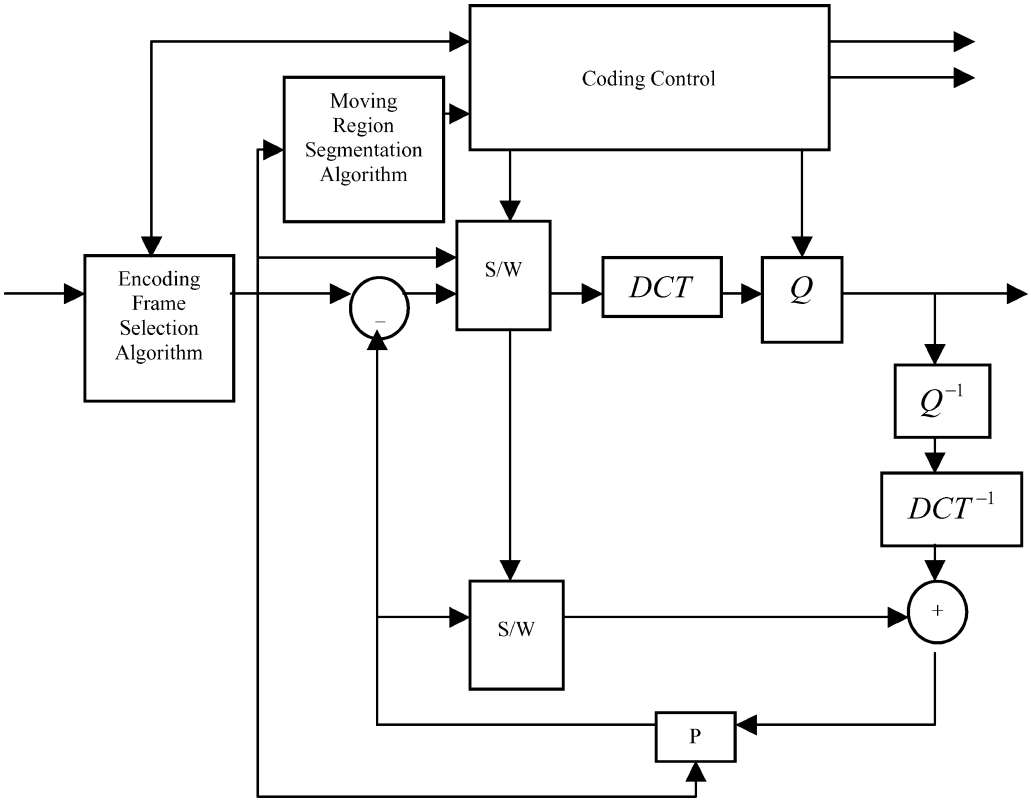

Fig. 1. Block diagram of the proposed region-based coding algorithm.

in Section II, the frame-layer rate-distortion (R-D) model is studied in Section III, which serves as the basis for the design of the proposed rate control algorithm in the following section. An efficient rate control algorithm for low VBR video is studied in Section IV, and experimental results are presented in Section V. Finally, concluding remarks are given in Section VI.

II. Fast 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 bits to moving objects and/or regions of interest (ROI). 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 [7], the temporal domain approach [8] and the spatio-temporal domain approach [9], [10]. The spatial-domain approach segments a given image into several regions by using spatial homogeneity. This approach tends to generate too many fine regions that do not correspond to real objects well. 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 [11]–[14], [42]. A region-change detection was employed in [11], and a simple geometric face model using the shape of ellipse and rectangle was proposed in [12]. In [13], Fukuhara *et al.* proposed a coding scheme by combining H.263 and block partitioning with patterns such as vertical, horizontal, 45°and 135° diagonal edges. The neural network technology was employed in [14] 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 intracoded mode, the intercoded mode with one motion vector, the inter-coded mode with four motion vectors, and uncoded. To be compatible with H.263+, our segmentation algorithm uses the macroblock as an elementary unit.

Here, we consider a simple yet effective moving region segmentation algorithm based on morphological processing techniques [5], [6]. 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 processes are needed to remove background noise to simplify the segmentation result. To reduce the high frequency noise, we pass the input image sequence through a low-pass filter before taking the difference, where the employed low-pass filter can be a simple 3×3 filter. 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. Finally,

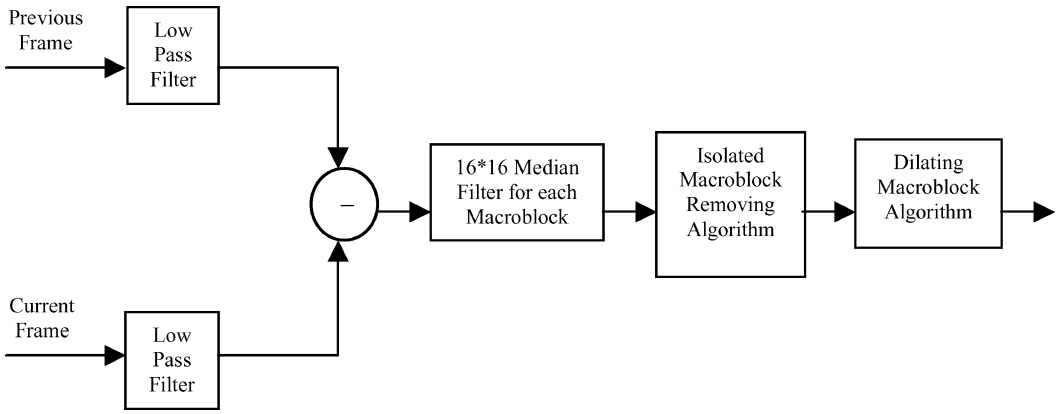

Fig. 2. Block diagram of the proposed region segmentation algorithm.

we apply the dilation process to remove the boundary discontinuity of moving objects that tends to degrade perceived quality greatly. Without dilation, boundary discontinuity is occasionally observed in spite of clearer moving objects. It is also observed that the proposed region segmentation algorithm does not work as well for the Foreman sequence as for the Salesman and the Silent Voice sequences since the background of the Foreman sequence is also moving. In this case, a more complicated segmentation algorithm is required.

III. Frame Layer R-D Models

Rate-distortion (R-D) modeling is important to the derivation of fast rate control algorithms with a low computational complexity. Generally speaking, there are two methods to achieve R-D modeling: statistical and experimental methods. One commonly used statistical model is to assume that the source signal has a 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 [15]. Other simplified models have also been examined. Some models were derived experimentally, e.g., the quadratic rate model [16], the exponential model [17], the spline approximation model [18] and the normalized rate-distortion model [19], etc. Even though statistical models demand a lower computational complexity than experimental models, experimental models can provide a more accurate R-D curve through a data fitting process. TMN8 [20] 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 the frame [18].

Here, 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 with respect to 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 III-A, and then derive the R-D models in Section III-B.

A. 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 [21], [22]. 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. Recently, several weighted MSE schemes have been proposed to take into account human visual characteristics in measuring the distortion of an image [22], [23]. In this work, we consider a moving-region-weighted MSE of a low computational complexity. It is defined as

$$D_w = \frac{1}{\omega}\sum_{i=1}^{N_{wd}}\sum_{j=1}^{N_{hgt}} \mu_{i,j}\,(p_{i,j}-\hat{p}_{i,j})^2, \tag{1}$$

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

where N_{wd} and N_{hgt} are pixel numbers of the width and the height of the frame, $p_{i,j}$ and $\hat{p}_{i,j}$ are pixel values of the original and the reconstructed images, respectively, and

$$\mu_{i,j} = \begin{cases} \mu_m & \text{if } (i,j) \in \text{moving region,} \\ \mu_s & \text{otherwise.} \end{cases} \tag{3}$$

Generally speaking, it is more reasonable to set $\mu_m \geq \mu_s$ by considering the human visual effect. Some subjective measure may also be required to determine the proper values of μ_m and μ_s. In the experimental section, we set $\mu_m = 100\ \mu_s$ empirically for all test sequences.

B. Rate-Distortion Modeling

We examine a frame layer R-D modeling approach that constructs both the rate and distortion models with respect to the averaged quantization parameter (QP) in each frame. In this paper, the quadratic rate [16] and the affine distortion models are employed. In terms of mathematics, they can be written as follows:

$$\hat{R}(\bar{q}) = (a\bar{q}^{-1} + b\bar{q}^{-2})M_w(f_{ref}, f_{cur}), \tag{4}$$

$$\hat{D}_w(\bar{q}) = a'\bar{q} + b' \tag{5}$$

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, $\bar{q}$ is the average QP of all macroblocks in a frame and

$$M_w(f_{ref}, f_{cur}) = \frac{1}{\omega} \sum_{i=1}^{N_{wd}} \sum_{j=1}^{N_{hgt}} \mu_{i,j} \left| \hat{p}_{i,j}^{ref} - p_{i,j}^{cur} \right| \quad (6)$$

and where $\hat{p}_{i,j}^{ref}$ 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 [16].

Conventionally, the R-D curve is computed based on integer QP's. In our case, $\bar{q}$ can be a floating-point number since it 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 [25]. 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 the distortion models again. We show the rate and distortion models in Fig. 3(a) and (b), respectively, for the QCIF Salesman sequence, where the circle denotes measured data points while the solid curve corresponds to the computed model. As shown in these two figures, the rate and the distortion models work reasonably well. The R-D modeling approach presented above provides a good approximation of the rate and the distortion performances with respect to the average QP for all test sequences in our experiment.

IV. Proposed Rate Control Algorithm for Low VBR Video

The rate control algorithm is often not standardized since it can be independent of the decoder structure and should be designed by the channel characteristics as well as the underlying video. 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 the rate and the 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 low-bit-rate environment, we have to reduce the number of I-frames, and the number of frames in a GOP can become larger. Thus, the basic encoding frame structure of H.263+ is *IPPPP*— even though the PB-mode is supported as an annex. This explains why existing MPEG-1/2 frame layer rate control algorithms cannot be straightforwardly extended to H.263+. In general, H.263/+ rate control deals with only predictive frames except I-frames [40].

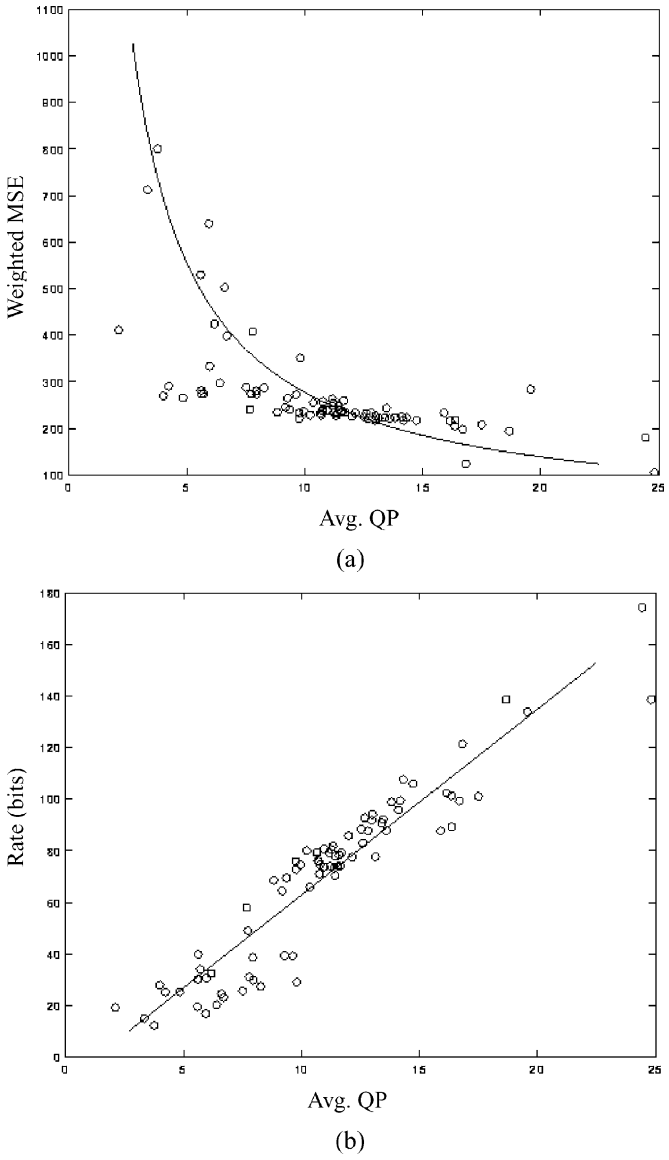

Fig. 3. Frame layer R-D modeling for the QCIF Salesman sequence: (a) rate model and (b) distortion model as a function of the average QP of macroblocks ($\mu_m = 100\ \mu_s$).

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 [26]–[32]. 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 of VBR video. In this work, we define a GOP as a *group of predictive frames* without I-frame in a fixed time interval. The proposed rate control scheme uses this new GOP as a basic rate control unit. The bit rate constraint is specified for each GOP.

As mentioned earlier, the proposed rate control algorithm contains three parts: 1) encoding frame selection, 2) frame layer rate control, and 3) macroblock layer rate control. Recently, some encoding frame rate control algorithms [33], [41] have been proposed. In [41], Hwang *et al.* adopted the dynamic frame skipping algorithm based on the motion vector information to

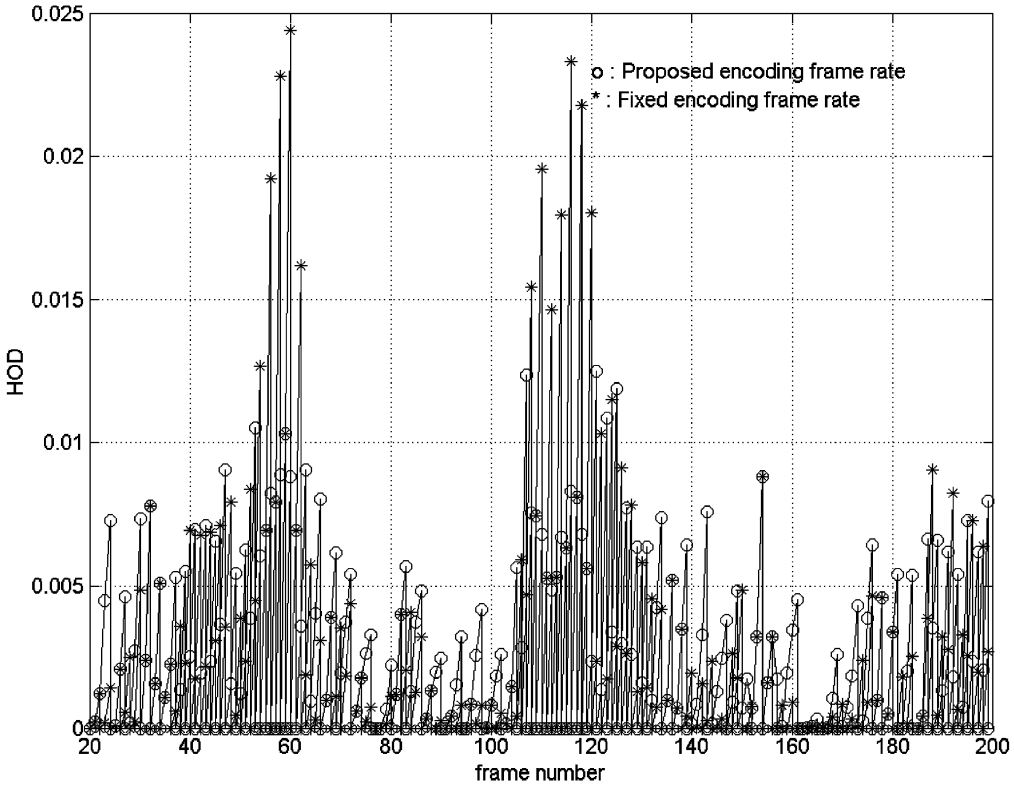

Fig. 4. Encoded frame positions and the HOD plot of the QCIF Salesman sequence.

TABLE I
HOD COMPARISON WITH THE FIXED FRAME RATE AND THE PROPOSED ENCODING FRAME SELECTION ALGORITHM THAT IS FRIENDLY TO THE DECODER WITH FRAME INTERPOLATION CAPABILITY

Method	Test Video	Avg HOD	STD of HOD
TMN8	Salesman	0.0046	0.00003525
	Silent Voice	0.0179	0.00008252
Prop. rate control	Salesman	0.0058	0.00000847
	Silent Voice	0.0202	0.00003206

make the motion of the decoded sequences smoother. In our previous work [33], we considered an encoding frame rate control that adjusts the frame rate by detecting motion change in video. This approach focused on spatial quality of frames without noticeable motion unsmoothness. We reduced the encoding frame rate in the fast motion interval to keep spatial quality in the tolerable range. However, it does not help the decoder to interpolate skipped frames.

Efficient post-processing 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 propose an encoding frame selection algorithm that is friendly to the decoder with the frame interpolation capability in Section IV-A, and a frame layer rate control scheme with a low computational complexity in Section IV-B.

A. Encoding Frame Selection w.r.t. Decoder With Frame Interpolation Capability

Generally speaking, motion compensated frame interpolation at the decoder exploit motion variation information between adjacent encoded frames. If we adopt a fixed encoding frame rate, the degree of motion between adjacent encoded frames is not constant and it is difficult for the decoder to interpolate skipped frames for intervals with fast and nonuniform motion so that the quality of interpolated frames can be poor. In this section, we propose an encoding frame selection algorithm with a low computational complexity to leverage the fact that motion variation is greatly related to the decoder's frame interpolation capability.

To detect motion change, many measures have been proposed [27]. Here, we adopt *HOD* (histogram of difference) since it is sensitive to local motion.

- If $HOD(f_{prev}, f_{cur}) \geq TH_1$, encode the current frame.

TABLE II
THE ENCODED FRAME POSITIONS FOR ENCODING FRAME SELECTION ALGORITHM FRIENDLY TO THE DECODER WITH FRAME INTERPOLATION CAPABILITY

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

TABLE III
THE ENCODED FRAME POSITIONS FOR TMN8

Test Video	Encoded frame positions
Salesman	20 22 24 26 28 30 32 34 36 38 40 42 44 46 48 50 52 54 56 58 60 62 64 66 68 70 72 74 76 78 80 82 84 86 88 90 100 102 104 106 108 110 112 114 116 118 120 122 124 126 128 130 132 134 136 138 140 142 144 146 148 150 152 154 156 158 160 162 164 166 168 170 172 174 176 178 180 182 184 186 190 192 194 196 198
Silent Voice	20 22 24 26 28 30 32 34 36 38 40 42 44 46 48 50 52 54 56 58 60 62 64 66 68 70 72 74 76 78 80 82 84 86 88 90 100 102 104 106 108 110 112 114 116 118 120 122 124 126 128 130 132 134 136 138 140 142 144 146 148 150 152 154 156 158 160 162 164 166 168 170 172 174 176 178 180 182 184 186 190 192 194 196 198

- If $HOD\,(f_{prev}, f_{cur}) < TH_1$, skip the current frame.
- If *HOD* is decreasing (which implies nonuniform motion change), encode the current frame.

In above, f_{prev} is the last encoded original frame, f_{cur} is the current original image, TH_1 is the threshold value, and *HOD* is defined as

$$HOD\,(f_n, f_m) = \frac{\sum_{i>|TH_0|} hod(i)}{N_{pixel}} \tag{7}$$

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 those in slow motion intervals. As a result, motion change between adjacent encoded frames is approximately of the same degree. This in turn helps the decoder to interpolate skipped frames successfully.

The *HOD* value is generally monotonically increasing. However, it can decrease when the nonuniform motion change exists. In this case, it is very difficult to interpolate the skipped frame even if the *HOD* is less than the threshold value. Thus, the third condition is added above to take into account the nonuniform motion change case.

B. Low-Complexity Frame-Layer Rate Control

After positions of encoded frames 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 [28] as a component in our overall scheme. For frame-layer rate control, we consider an algorithm of a low computational complexity in this section. The frame layer rate control problem can be formulated as follows.

Problem Formulation: Determine $\bar{q}_i$, $i = 1, 2, \ldots, N_{gop}$, to minimize

$$\sum_{i=1}^{N_{gop}} \hat{D}_{w,i}(\bar{q}_i) \tag{8}$$

subject to

$$\sum_{i=1}^{N_{gop}} R_i \le B_{gop} \tag{9}$$

where N_{gop} and B_{gop} are the encoded frame number and the target bit rate in a GOP, respectively, $\hat{D}_{w,i}$ 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 [34]–[36]. 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 [29], [37], [38]. Here, we adopt a suboptimal scheme that consists of two steps. They work iteratively to reduce the computational complexity and the coding delay. The two steps are described as follows.

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.,

$$P_k(\bar{q}_k) = \hat{D}_{w,k}(\bar{q}_k) + \lambda_k \cdot \max\left\{0, \hat{B}_k^{res}(\bar{q}_k)\right\} \tag{10}$$

where $P_k(\bar{q}_k)$ is the cost function for the k_{th} frame and λ_k is the Lagrange multiplier for the k_{th} frame, and

$$\hat{B}_k^{res}(\bar{q}_k) = B_{k-1}^{used} + \left\{\hat{R}_k(\bar{q}_k) - R_k^{ref}\right\}, \tag{11}$$

$$B_{k-1}^{used} = \sum_{i=1}^{k-1} R_i, \tag{12}$$

$$R_k^{ref} = \left(1 + \mu \cdot \frac{HOD_k - HOD_{avg}}{HOD_{avg}}\right) \cdot \frac{B_{gop}}{N_{gop}} \tag{13}$$

where HOD_k is the HOD value of the k_{th} encoded frame, HOD_{avg} is the average HOD value of the GOP, and μ is an weighting factor (μ is set to 1 in this work). The reference bit rate for each frame is determined by its motion change since the frame with larger motion change needs more bit rates generally. Based on the rate and the distortion models in Section III, 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:

$$\bar{q}_k^* = \arg\min_{\bar{q}_k} P_k(\bar{q}_k). \tag{14}$$

What we actually need is not $\bar{q}_k^*$ but the target bit budget $\hat{R}(\bar{q}_k^*)$ for the k_{th} frame.

Step 2: Lagrange Multiplier Adaptation: We adopt an adaptive Lagrange multiplier selection rule based on the LMS method [29]. The Lagrange multiplier is updated by using residual bits after the coding of a frame. The adaptation algorithm is stated as follows:

$$\lambda_{k+1} = \lambda_k + \Delta\lambda_k \tag{15}$$

$$\Delta\lambda_k = \frac{B_k^{used}}{B_k^{ref}} - 1 \tag{16}$$

where λ_{k+1} is the updated Lagrange multiplier for the $(k+1)_{th}$ frame, and

$$B_k^{ref} = \sum_{i=1}^{k} R_k^{ref}. \tag{17}$$

Finally, the macroblock layer rate control in TMN8 is employed as the third component of the overall rate control scheme.

We would like to give some remarks on the computational complexity of the proposed algorithm. The proposed algorithm needs to analyze of the motion change of the underlying video to determine the encoded frame positions, thus the encoding time delay may occur. However, the main idea of the proposed encoding frame selection algorithm can be extended to the real-time version with some modifications such as a sliding window approach. Furthermore, the complexity for the proposed segmentation algorithm is very low compared with other segmentation algorithms in the literatures [11]–[14], [42], and the additional complexity of rate-distortion modeling and frame layer rate control can be very low compared with those of other encoding processes.

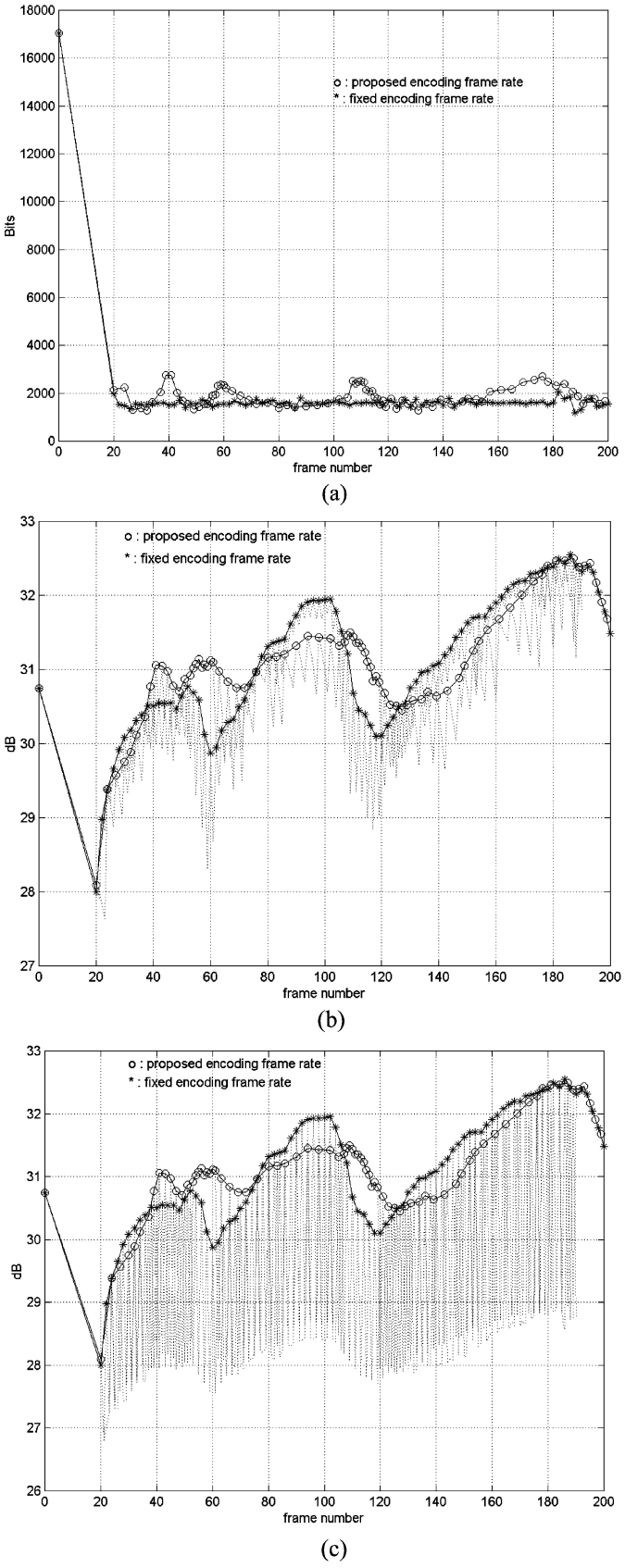

Fig. 5. Performance comparison for the QCIF Foreman: (a) rate plot, (b) PSNR plot of Intrafiltering, and (c) PSNR plot of motion compensated frame interpolation.

V. Experimental Results

In the experiment, the macroblock layer rate control adopted by TMN8 was employed and our implementation was based on the UBC H.263+ source code [39]. The performance comparison was made based on subjective as well as objective evaluation with more emphasis on subjective evaluation. The PSNR (not weighted PSNR) value is employed as the objective measure.

The I-frame was encoded at a predetermined bit rate in the experiment. 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 the 24 kbps channel, the bit rate for the I-frame cannot exceed 24 kbits. In our experiment, the I-frame was encoded with $QP = 15$, which satisfied the above recommendation. Region-based video coding and the proposed rate control algorithm is compared with TMN8.

Experimental results are presented to demonstrate the performance of the proposed region-based coding scheme and its associated 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. The parameter TH_1 was determined by analyzing GOP with the initial frame skip interval. That is, TH_1 was set to the average *HOD* in a GOP. As a result, TH_1 was set to 0.005 695 for Salesman and 0.019 030 for Silent Voice. These were the average *HOD* values at a frame rate of 15 fps for the two test sequences. The value TH_0 in (7) was set to 32, and one GOP consists of 200 frames.

The comparison of *HOD* results of TMN8 and the proposed encoding frame selection algorithm for QCIF Salesman and Silent Voice sequences are summarized in Table I. As shown in Table I, the proposed encoding frame selection algorithm decreases the standard deviation of the *HOD* values significantly. The encoded frame positions of the proposed encoding frame selection algorithm are given in Table II while those of TMN8 are listed in Table III. As shown in these tables, more frames are encoded in the fast motion change interval than the slow motion change interval. Furthermore, as shown in Fig. 4, we observed the *HOD* value trace and encoding frame positions and found that the *HOD* value was decreasing between frame numbers 70–80. It is difficult for the decoder to interpolate skipped frames due to nonuniform motion change. For this case, we follow the 3rd condition of the HOD test as described in Section IV.A to encode the frames.

The objective performances of coded video are shown in Figs. 5 and 6. The output bit rate plots are given in Figs. 5(a) and 6(a), and the PSNR plots are depicted in Figs. 5(b) and 6(b). The performances of two frame interpolation schemes are compared in Tables IV and V. They are the advanced motion compensated frame interpolation [4] and the simple

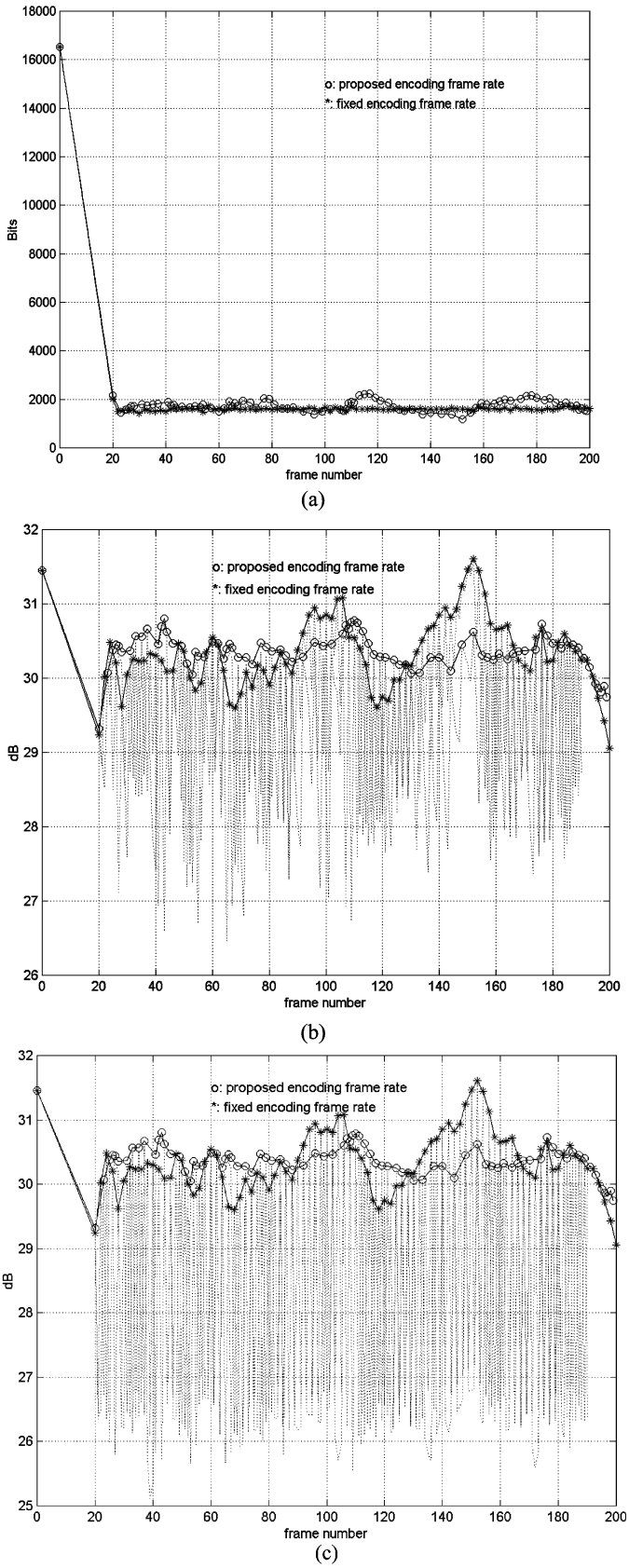

Fig. 6. Performance comparison for the QCIF Foreman: (a) rate plot, (b) PSNR plot of Intrafiltering, and (c) PSNR plot of motion compensated frame interpolation of QCIF Silent Voice under unconstrained VBR.

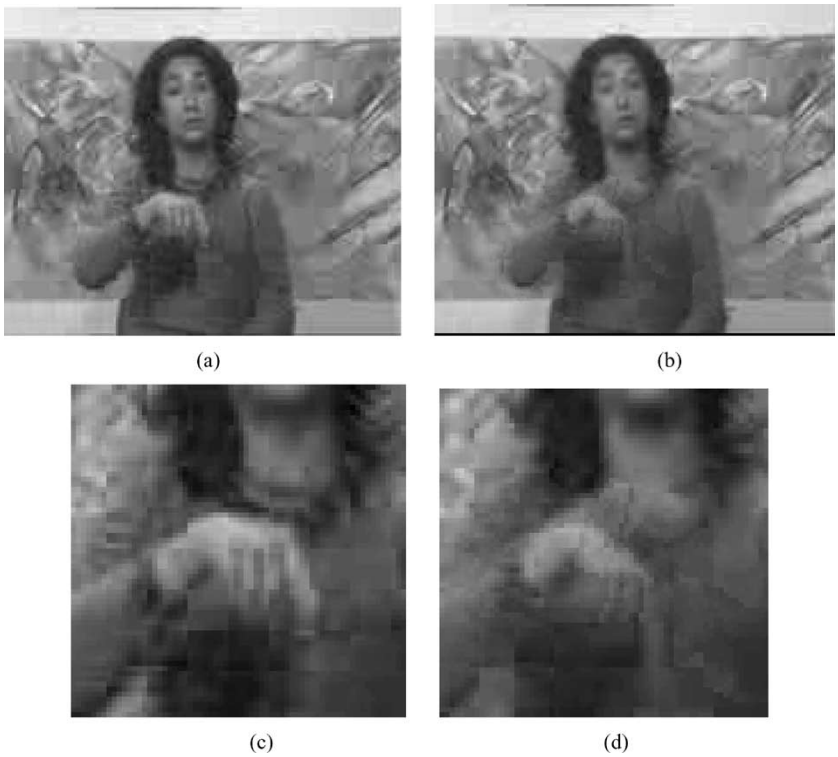

Fig. 7. Subjective quality comparison of the frames interpolated by FMCI [5]: (a) the 47th frame when encoded by the proposed algorithm, and (b) the 47th frame when encoded by TMN8 with a fixed time interval, (c) the enlarged hand part of (a), and (d) the enlarged hand part of (b).

TABLE IV
PERFORMANCE COMPARISON WITH TMN8 UNDER UNCONSTRAINED VBR WHEN THE INTRAFILTERING FRAME INTERPOLATION IS EMPLOYED, WHERE THE TARGET AVERAGE RATE IS 24 kbps

Method	Test Video	Avg PSNR	STD of PSNR
TMN8	Salesman	30.7998	1.0217
	Silent Voice	29.4500	1.2118
Region & Prop. rate control	Salesman	30.7055	0.8081
	Silent Voice	29.2233	1.1914

TABLE V
PERFORMANCE COMPARISON WITH TMN8 UNDER UNCONSTRAINED VBR WHEN THE MOTION COMPENSATED FRAME INTERPOLATION IS EMPLOYED, WHERE THE TARGET AVERAGE RATE IS 24 kbps

Method	Test Video	Avg PSNR	STD of PSNR
TMN8	Salesman	29.6515	1.5950
	Silent Voice	28.3928	2.0304
Region & Prop. rate control	Salesman	29.3612	1.4970
	Silent Voice	28.0699	2.0699

frame repetition (*intrafiltering*) scheme. As shown by these experimental results, the proposed approach decreases the PSNR fluctuation although the average PSNR is almost the same for both sequences with frame interpolation. This phenomenon comes from the following fact. While the quality of moving regions (face and hand parts) is much improved, the quality of still regions (telephone and background) is degraded compared with TMN8 as shown in Fig. 7. The reason is that the proposed algorithm does not update still regions. Therefore, the overall objective quality (PSNR) can be degraded although the subjective quality is improved.

For subjective quality evaluation, interpolated frames are given in Figs. 7 and 8. As shown in Fig. 7(b) and (d), the ghost artifact appears obviously in the hand part of Silent Voice

Fig. 8. Subjective quality comparison of the frames interpolated by FMCI [5]: (a) the 103rd frame when encoded by the proposed algorithm, (b) the 103rd frame when encoded by TMN8 with a fixed time interval, (c) the enlarged face part of (a), (d) the enlarged face part of (b), (e) the enlarged telephone part of (a), and (d) the enlarged telephone part of (b).

under a fixed encoding frame rate. Whereas, the proposed encoding frame selection designed based on motion change information can reduce this phenomenon so that the moving regions including face and body parts are clearer than those from TMN8. In addition, we can see a much-improved facial region and the hand part with the proposed region-based coding and the rate control algorithm in comparison of the TMN8 result as shown in Fig. 8(b) and (d). On the other hand, the quality in still regions such as the bookshelf and telephone is degraded as shown in Fig. 7(e) and (f) since these regions are not updated. Thus the overall PSNR value can be decreased compared with that of TMN8. Furthermore, the proposed algorithm improves motion smoothness than a fixed encoding frame rate when the frame interpolation technique is employed at the decoder. Therefore, the subjective quality is even much more improved compared with the objective quality.

VI. Conclusion and Future Work

Region-based video coding scheme compatible with H.263+ and a three-layer rate control algorithm were studied in this work. The proposed region-based video coding scheme is a hybrid block- and object-based coding scheme, and the proposed rate control treated the encoding frame position as a control variable. It consists of an encoding frame selection scheme that is friendly to the decoder with a frame interpolation capability and a frame-layer rate control algorithm with a low computational complexity. By the experimental results, it was observed that the proposed region-based coding scheme enhanced the subjective spatial quality in a frame and the proposed rate control algorithm improved obviously the human visual perceptual quality of the interpolated frames at the decoder end based on the enhanced spatial quality of the encoded frames. Therefore, the subjective performance is more improved than the objective performance.

However, its detailed performance analysis requires further research in the future.

REFERENCES

[1] ITU-T, "Video coding for low bit-rate communication," ITU-T Rec. H.263 Ver. 2, 1998.

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

[3] A. M. Tekalp, *Digital Video Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1995.

[4] T. Kuo and C.-C. J. Kuo, "Motion-compensated interpolation for the low-bit-rate video quality enhancement," in *Proc. SPIE Int. Symp. Optical Science, Engineering and Instrumentation*, May 1998.

[5] W. K. Pratt, *Digital Image Processing*. New York: Wiley, 1991.

[6] A. K. Jain, *Fundamentals of Digital Image Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1989.

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

[8] H. G. Musman, M. Hotter, and J. Ostermann, "Object-based analysis-synthesis of moving images," *Signal Process.: Image Commun.*, pp. 117–138, Oct. 1989.

[9] J. Benois, L. Wu, and D. Barba, "Joint contour-based and motion-based image sequences segmentation for TV image coding at low bit rate," *Proc. SPIE, Vis. Commun. Image Process.*, pp. 1074–1085, Sept. 1994.

[10] C. Gu and M. Kunt, "Very low-bit-rate video coding using multi-criterion segmentation," in *Proc. Int. Conf. Image Processing*, vol. 2, Nov. 1994, pp. 418–422.

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

[12] J. Hartung, A. Jacquin, J. Pawlyk, J. Rosenburg, H. Okada, and P. E. Crouch, "Object-oriented H.263 compatible video coding platform for conferencing applications," *IEEE J. Select. Areas Commun.*, vol. 16, pp. 42–55, Jan. 1998.

[13] T. Fukuhara, K. Asai, and T. Murakami, "Very low bit-rate coding with block partitioning and adaptive selection of two time-differential frame memories," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 7, pp. 212–220, Feb. 1997.

[14] N. Doulamis, A. Doulamis, D. Kalogeras, and S. Kollias, "Low bit-rate coding of image sequences using adaptive regions of interest," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 8, pp. 928–934, Dec. 1998.

[15] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley, 1992.

[16] T. Chiang and Y. -Q. Zhang, "A new rate control scheme using quadratic rate distortion model," *IEEE Trans. Circuits Systems Video Technol.*, vol. 7, pp. 246–250, Sept 1997.

[17] W. Ding and B. Liu, "Rate control of MPEG video coding and reordering by rate-quantization modeling," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 6, pp. 12–20, Feb. 1996.

[18] L. J. Lin, A. Ortega, and C. -C. J. Kuo, "Rate control using spline interpolated R-D characteristics," *Proc. SPIE, Vis. Commun. Image Process.*, pp. 111–122, Mar. 1996.

[19] K. H. Yang, A. Jacquin, and N. S. Jayant, "Normalized rate-distortion model for H.263-compatible codecs and its application to quantizer selection," in *Proc. IEEE Int. Conf. Image Processing*, vol. 2, Oct. 1997, pp. 41–44.

[20] "ITU-T, "Video Codec Test Model, Near-Term, Version 8 (TMN8)," Tech. Rep.," AdHoc Group, Portland, H.263, 1997.

[21] A. Ortega and K. Ramchandran, "Rate-distortion for image and video compression," *IEEE Signal Processing Mag.*, vol. 15, pp. 23–50, Nov. 1998.

[22] N. Jayant, J. Johnson, and R. Safranek, "Signal compression based on models of human perception," *Proc. IEEE*, vol. 81, pp. 1385–1422, Oct. 1993.

[23] R. Balasubramanian, C. A. Bouman, and J. P. Allebach, "Sequential scalar quantization of color images," *J. Electron. Imag.*, vol. 3, no. 1, pp. 45–59, Jan. 1994.

[24] J. K. Su, J. J. Eggers, and B. Girod, "Optimal attack on digital watermarks and its defense," in *Proc. 34th Asilomar Conference on Signal, Systems, and Computers*, Pacific Grove, CA, Oct. 2000.

[25] "MPEG-4 Video Verification Model Version 10.0," Moving Picture Expert Group Video group, ISO/IEC JTC1/SC29/WG11, 1998, to be published.

[26] D. Mukherjee and S. K. Mitra, "Combined mode selection and macroblock step adaptation for H.263 video encoder," in *Proc. IEEE Int. Conf. Image Processing*, vol. 2, Oct. 1997, pp. 37–40.

[27] J. Lee and B. W. Dickenson, "Temporally adaptive motion interpolation," *IEEE Trans. Image Processing*, vol. 3, pp. 513–526, Sept. 1994.

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

[29] T. Wiegand, M. Lightstone, D. Mukherjee, T. G. Campbell, and S. K. Mitra, "Rate-distortion optimized mode for very low bit rate video coding and emerging H.263 standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 6, pp. 182–190, Apr. 1996.

[30] K. T. Ng, S. C. Chan, and T. S. Ng, "Buffer control algorithm for low bit rate video compression," in *Proc. IEEE Int. Conf. Image Processing*, Sept. 1996.

[31] K. Oehler and J. L. Webb, "Macroblock quantizer selection for {H.263} video coding," in *Proc. IEEE Int. Conf. Image Processing*, vol. 1, Oct. 1997, pp. 365–368.

[32] G. Schuster and A. Katsaggelos, "Fast and efficient mode and quantizer selection in the rate and distortion sense for H.263," in *Proc. SPIE, Vis. Commun. Image Process.*, 1996.

[33] H. Song, J. Kim, and C. -C. J. Kuo, "Real-time encoding frame rate adjustment for H.263+ video over the internet," *Signal Process.: Image Commun.*, vol. 15, no. 1–2, pp. 127–148, Sep. 1999.

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

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

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

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

[38] D. W. Lin and J. J. Chen, "Efficient optimal rate-distortion coding of video sequences under multiple rate constraints," in *Proc. IEEE Int. Conf. Image Processing*, vol. 2, Oct. 1997, pp. 29–32.

[39] "H.263+ Encoder/Decoder," Univ. British Columbia, Image Processing Lab, Vancouver, BC, Canada, TMN (H.263) codec, 1998.

[40] H. Song and C.-C. J. Kuo, "Rate control for low bit rate video via variable encoding frame rates," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 11, no. 4, pp. 512–521, Apr. 2001.

[41] J. Hwang, T. Wu, and C. Lin, "Dynamic frame skipping in video transcoding," in *IEEE Second Workshop on Multimedia Signal Processing*, 1998.

[42] K. Wong, K. Lam, and W. Siu, "An efficient low bit-rate video-coding algorithm focusing on moving regions," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 11, no. 10, pp. 1128–1134, Oct. 2001.

Hwangjun Song received the B.S. and M.S. degrees from the Department of Control and Instrumentation (EE), Seoul National University, Seoul, Korea, in 1990 and 1992, respectively, and the Ph.D. degree in electrical engineering-systems, University of Southern California (USC), Los Angeles, in 1999.

He was a Research Engineer at LG Industrial Lab, Korea, in 1992. From 1995 to 1999, he was a Research Assistant in SIPI (Signal and Image Processing Institute) and IMSC (Integrated Media Systems Center), USC. Since 2000, he has been with the School of Electronic and Electrical Engineering, Hongik University, Seoul. His research interests include multimedia signal processing and communication, image/video compression, digital signal processing, network protocols necessary to implement functional image/video applications, control systems and fuzzy-neural systems.

Dr. Song served as a Technical Organizing Committee member of Packet Video 2001 and the IEEE International Symposium on Industrial Electronics 2001. He served as a Program Committee member of SPIE Visual Communication and Image Processing 2002.

C.-C. Jay Kuo (M'87–SM'92–F'99) received the B.S. degree from the National Taiwan University, Taipei, in 1980 and the M.S. and Ph.D. degrees from the Massachusetts Institute of Technology, Cambridge, in 1985 and 1987, respectively, all in electrical engineering.

He was Computational and Applied Mathematics (CAM) Research Assistant Professor in the Department of Mathematics at the University of California, Los Angeles, from October 1987 to December 1988. Since January 1989, he has been with the Department of Electrical Engineering-Systems and the Signal and Image Processing Institute at the University of Southern California, Los Angeles, where he currently has a joint appointment as Professor of electrical engineering and mathematics. His research interests are in the areas of digital signal and image processing, audio and video coding, media communication technologies and delivery protocols, and network computing. He is Editor-in-Chief for the *Journal of Visual Communication and Image Representation*, and Editor for the *Journal of Information Science and Engineering* and the *EURASIP Journal of Applied Signal Processing*. He has guided 37 students to their Ph.D. degrees. He is a co-author of more than 500 technical publications in international conferences and journals as well as the following books: *Content-based Audio Classification and Retrieval for Audiovisual Data Parsing*, with Tong Zhang (Norwell, MA: Kluwer, 2001), *Semantic Video Object Segmentation for Content-based Multimedia Applications*, with Ju Guo (Norwell, MA: Kluwer, 2001), *Intelligent Systems for Video Analysis and Access over the Internet*, with Wensheng Zhou (Upper Saddle River, NJ: Prentice-Hall, 2002), *Quality of Service Provisioning for Multimedia Applications in Service Differentiation Networks*, in preparation with Jitae Shin and Daniel Lee (Upper Saddle River, NJ: Prentice-Hall, 2003).

Dr. Kuo is a member of SIAM, ACM, and a Fellow of SPIE. He is an Associate Editor for IEEE TRANSACTIONS ON SPEECH AND AUDIO PROCESSING. He served as Associate Editor for IEEE TRANSACTIONS ON IMAGE PROCESSING in 1995–1998 and IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY in 1995–1997. He received the National Science Foundation Young Investigator Award (NYI) and Presidential Faculty Fellow (PFF) Award in 1992 and 1993, respectively.