

## Research Article

# On Optimizing H. 264/AVC Rate Control by Improving R-D Model and Incorporating HVS Characteristics

Zhongjie Zhu,<sup>1</sup> Yuer Wang,<sup>1</sup> Yongqiang Bai,<sup>1</sup> and Gangyi Jiang<sup>2</sup>

<sup>1</sup>Ningbo Key Laboratory of DSP, Zhejiang Wanli University, Ningbo 31 51 00, China

<sup>2</sup>Institute of Circuits and Systems, Ningbo University, Ningbo 315211, China

Correspondence should be addressed to Zhongjie Zhu, zhongjiezhu@hotmail.com

Received 18 March 2010; Revised 4 July 2010; Accepted 24 August 2010

Academic Editor: Lisimachos P. Kondi

Copyright © 2010 Zhongjie Zhu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The state-of-the-art JVT-G012 rate control algorithm of H.264 is improved from two aspects. First, the quadratic rate-distortion (R-D) model is modified based on both empirical observations and theoretical analysis. Second, based on the existing physiological and psychological research findings of human vision, the rate control algorithm is optimized by incorporating the main characteristics of the human visual system (HVS) such as contrast sensitivity, multichannel theory, and masking effect. Experiments are conducted, and experimental results show that the improved algorithm can simultaneously enhance the overall subjective visual quality and improve the rate control precision effectively.

## 1. Introduction

Rate control plays an important role in the field of video coding and transmission, which has been extensively studied in the literature. Strictly speaking, rate control technology is not a normative part of video coding standards. However, due to its importance for video coding and transmission, rate control has been widely studied, and several rate control algorithms have been proposed for inclusion in the reference software implementations of the existing video coding standards such as the TM5 for MPEG-2, the TMN8 for H.263, the VM8 for MPEG-4, and the F086 and G012 for H.264 [1–4]. Of all those algorithms the JVT-G012 has attracted great attention over the last few years and is being widely used.

R-D models describe the relationships between the bitrates and the distortions in the reconstructed video, which can enable an encoder to determine the required bit rate to achieve a target quality. For a rate control algorithm, the R-D model is a key part, and its accuracy greatly affects the rate control performance. Hence, how to improve the prediction accuracy of R-D model is the key to enhance the performance of rate control algorithm. Quite a few R-D models have been proposed such as the logarithm model, exponential model, and quadratic model [5]. Of all these models the quadratic

model is adopted by JVT-G012 which is derived based on the assumption that video source follows Gaussian distribution. However, as analyzed in Section 2, it is not accurate enough in practical applications. There also exist several works where the quadratic models have been improved [6, 7]. However, almost all of them have just emphasized on how to improve the accuracy of frame complexity prediction, and their R-D models' structure is very similar to that of the original quadratic model. At the same time, studies on visual physiology and visual psychology indicate that observers usually have different sensitivities and interests to different video contents in video sequence, and the contents with more attention-attraction and higher visual sensitivity will be more sensitive to coding error. Hence, during video coding process, the regions with high sensitivities can be allocated more bits to acquire a higher overall subjective quality. Many scholars have been working on HVS-based video coding technologies, and many achievements have been made. For example, a novel subjective criteria for video quality evaluation based on the foveal characteristics has been developed in [8]. Nguyen and Hwang have presented a rate control approach for low bit-rate streaming video by incorporating the HVS characteristic of smooth pursuit eye movement, which can improve the quality of moving scenes in a video sequence [9]. Tang et al. have also proposed

a novel rate control algorithm by considering the motion and texture structures of video contents [10]. However, since HVS is extremely complex, it is very difficult to obtain perfect perceptually consistent result during rate control and video coding. Thus, this topic should be further studied [11–13].

Based on above observations, the G012 algorithm is optimized from two aspects in this paper. First, the traditional quadratic R-D model is improved based on both empirical observations and theoretical analysis. Second, based on current physiological and psychological research findings of HVS, the main HVS characteristics such as contrast sensitivity, multichannel theory, and masking effect are analyzed, and then the G012 algorithm is optimized by incorporating these characteristics.

This paper is organized as follows. In Section 2, the quadratic model is analyzed and improved. In Section 3, some main HVS characteristics and the HVS-based rate control optimization are introduced. Experiments are conducted in Section 4, and Section 5 concludes the paper.

## 2. Improved R-D Model

The classical quadratic model is derived based on the assumption that video source follows Gaussian distribution [14]. Its R-D function can be formulated as follows:

$$R(D) = \ln\left(\frac{1}{\alpha D}\right), \quad 0 < D < \frac{1}{\alpha}. \quad (1)$$

After expanded with Taylor series, the above formula can be expressed as

$$\begin{aligned} R(D) &= \left(\frac{1}{\alpha D} - 1\right) - \frac{1}{2}\left(\frac{1}{\alpha D} - 1\right)^2 + R(o) \\ &= -\frac{3}{2} + \frac{2}{\alpha}D^{-1} - \frac{1}{2\alpha^2}D^{-2} + R(o). \end{aligned} \quad (2)$$

Let  $Q = D$ , then the quadratic R-Q model can be acquired

$$R(D) = \frac{\alpha}{Q} + \frac{\beta}{Q^2}, \quad (3)$$

where  $\alpha$ ,  $\beta$  are model parameters,  $Q$  denotes quantization step, and  $R(o)$  is the residual.

The above derivation is not very precise for the following several reasons. Firstly, the assumption that the DCT coefficients follow Gaussian distribution is not always true in practical applications. Secondly, only the first and the second items in expanded Taylor series are retained in (2). The constant and high-order items are totally discarded, which will introduce errors. In addition, it is not accurate to just let  $Q = D$ .

To further investigate the accuracy of the quadratic model, experiments have been performed. The actual relationship between the bitrates and the quantization steps has been analyzed. Nearly twenty video sequences have been tested, including *Alex*, *Claire*, *Train*, *Discuss*, *Mom*, *Grandmom*, *Mom\_daughter*, *Trevor*, *Tennis*, *Foreman*, *Sergio*, *Susan*, *Miss*, *Sflowg*, *Janine*, *Pascal*, and *Herve* sequences. For each sequence, more than 50 frames are tested. Our experiments

TABLE 1: Test parameters.

Parameter	Value
Frame Rate	30
Number Reference Frames	5
Search Range	16
Use Hadamard	1
RD Optimization	1
Rate Control Enable	1
Frame Skip	2
Number B Frames	2

TABLE 2: MOS metric.

Quality	Score
very good	9, 10
good	7, 8
average	4, 5, 6
bad	2, 3
very bad	0, 1

TABLE 3: The average intensities and spatial frequency of test sequences (unit: c/d).

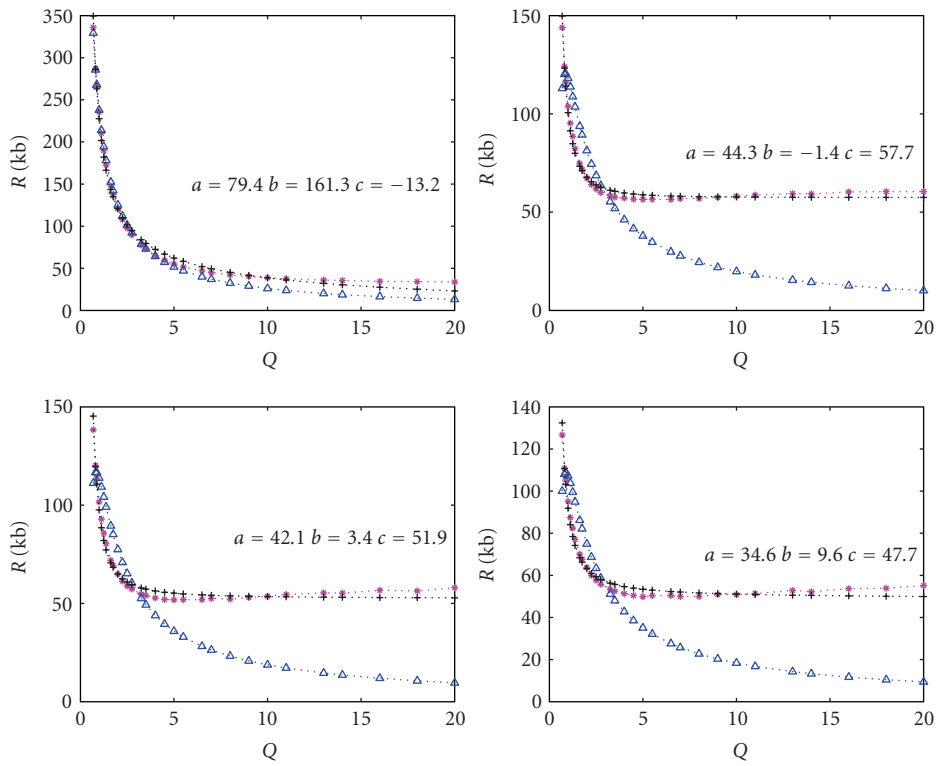
Testing sequence	Alex	Claire	Train
Intensity	79.50	96.79	72.57
Spatial frequency	5.39	3.99	14.53

are performed on our improved JM 10.2 platform, and the IBBP structure is used, where  $I$  denotes “Intracoded frame,”  $P$  denotes “Predicted frame,” and  $B$  denotes “Bipredictive frame.” For H.264/AVC, quantization parameter’s range is within [0, 51]. The corresponding quantization step’s range is within [0.625, 224]. In our experiments, 30 quantization parameters from 1 to 30 are used. Each sequence is firstly encoded with different quantization steps. Then the acquired R-Q data are fitted by the quadratic R-Q model, and the fit errors are analyzed. From experimental results of nearly one thousand frames, we have found that, as shown in Figure 1, the actual bit-rate does not always decrease obviously with the increment of  $Q$ , while the fitted curve of the quadratic R-Q model almost decreases obviously with the increment of  $Q$ . Hence, there exists considerable gap between the actual R-Q curve and the fitted curve of the quadratic R-Q model. As a result, the quadratic model is not accurate enough in practical rate control applications.

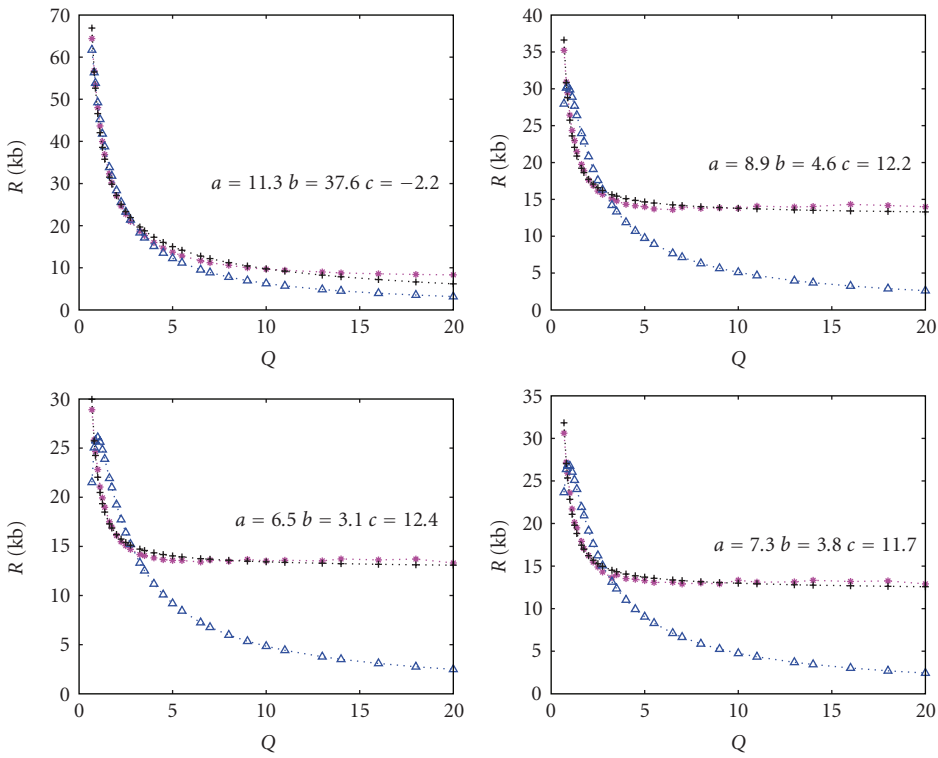
Based on above observations, the quadratic model is adapted to be

$$R = \frac{a}{\sqrt{Q}} + \frac{b}{Q^2} + c, \quad (4)$$

where  $a$ ,  $b$ , and  $c$  are model parameters and they can be computed based on linear regression method, which is briefly introduced as follows.

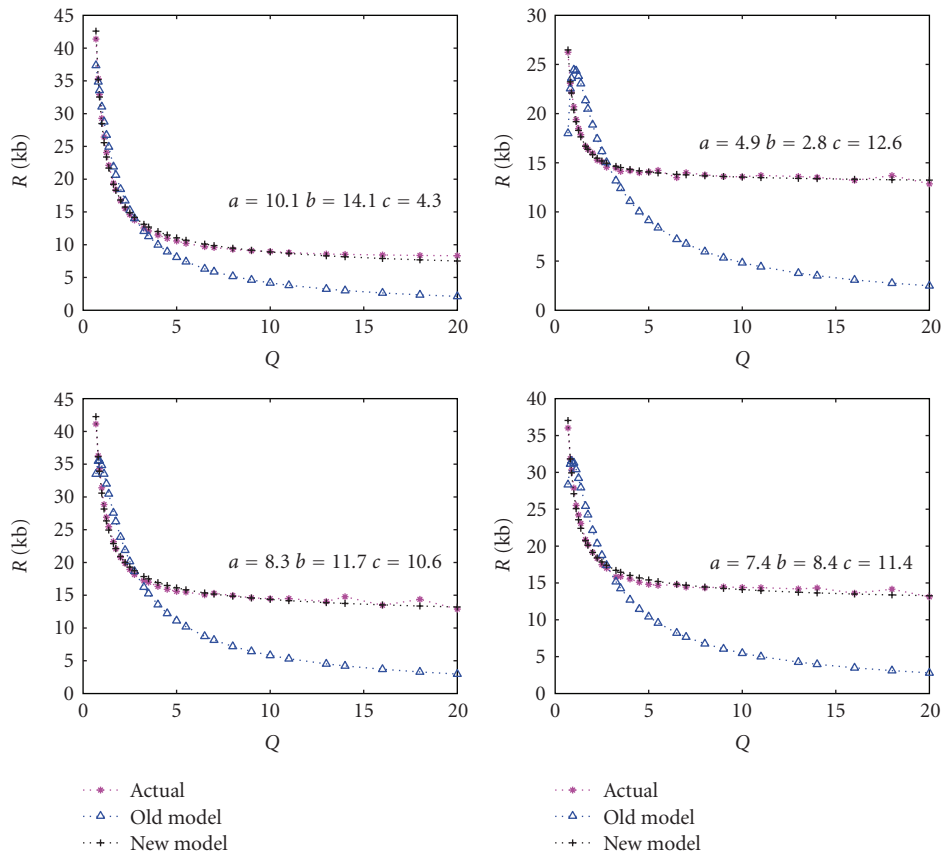


(a) Results of four frames of *Train* sequence (the top-left is an I-frame; others are P-frames)



(b) Results of four frames of *Discus* sequence (the top-left is an I-frame; others are P-frames)

FIGURE 1: Continued.



(c) Results of four frames of *Trevor* sequence (the top-left is an I-frame; others are P-frames)

FIGURE 1: Some experimental results on R-Q models.

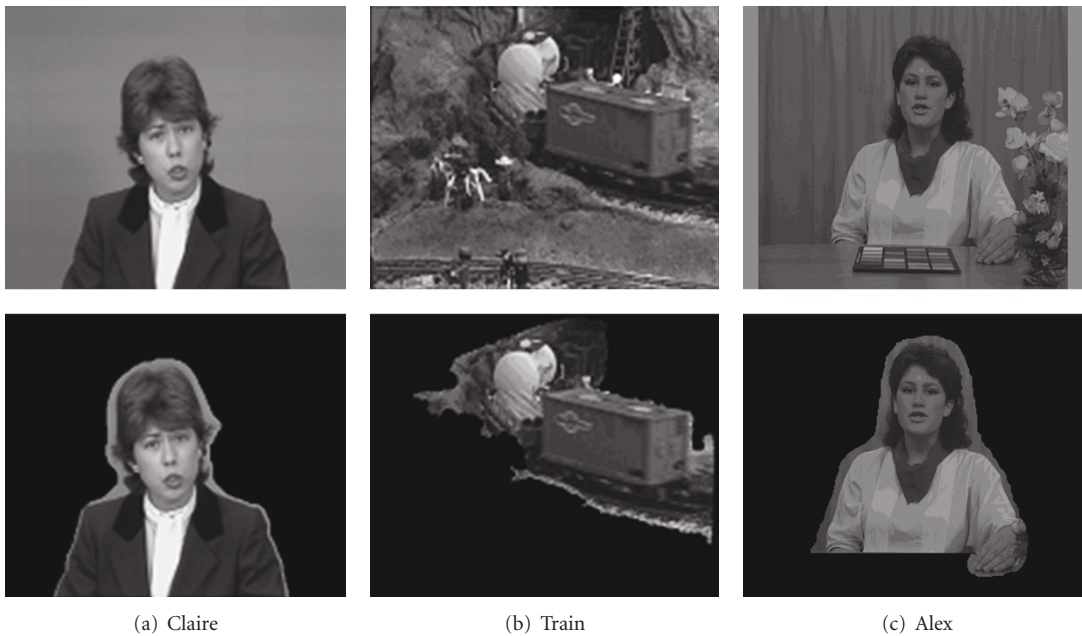


FIGURE 2: Moving segmentation results.

TABLE 4: Rate-control results of *Claire* sequence.

$R_T$	160 Kbps		150 Kbps		140 Kbps		130 Kbps	
Algorithm	Original	New	Original	New	Original	New	Original	New
$\delta_d$ (%)	0.4394	0.2556	0.1857	0.1243	0.5466	0.6897	0.7058	0.6326
$R_a$	160.703	159.591	150.279	149.814	139.235	139.035	129.083	129.178
$\bar{P}$	38.920	37.874	38.511	37.432	37.969	36.982	37.536	36.564

TABLE 5: Rate-control results of *Alex* sequence.

$R_T$	160 Kbps		150 Kbps		140 Kbps		130 Kbps	
Algorithm	Original	New	Original	New	Original	New	Original	New
$\delta_d$ (%)	1.4250	0.5553	1.2844	0.9736	1.1171	0.9176	1.2379	0.9115
$R_a$	162.280	160.888	151.927	151.461	141.564	141.285	131.609	131.185
$\bar{P}$	36.801	34.285	36.099	33.616	35.668	33.238	34.771	32.640

TABLE 6: Rate-control results of *Train* sequence (1).

$R_T$	1300 Kbps		1200 Kbps		1100 Kbps		1000 Kbps	
Algorithm	Original	New	Original	New	Original	New	Original	New
$\delta_d$ (%)	1.1469	5.3759	0.9871	1.6468	0.8959	0.2020	0.4331	0.1977
$R_a$	1.3149	1.3699	1.2118	1.2198	1.1099	1.1022	0.9957	0.9980
$\bar{P}$	32.662	32.124	32.411	31.906	32.148	31.646	31.665	31.160

TABLE 7: Rate-control results of *Train* sequence (2).

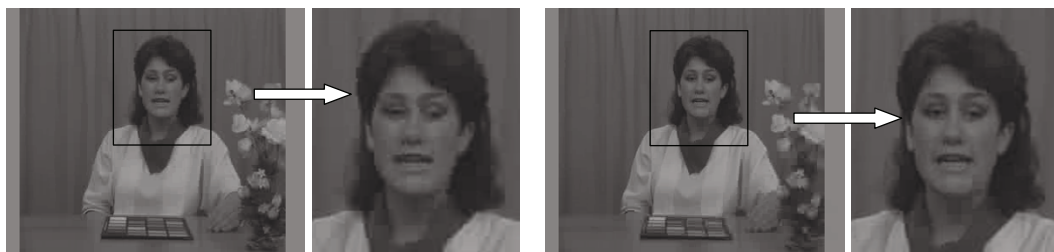
$R_T$	900 Kbps		800 Kbps		700 Kbps		64 Kbps	
Algorithm	Original	New	Original	New	Original	New	Original	New
$\delta_d$ (%)	0.9894	0.2003	1.3284	0.6510	0.7353	0.0582	0.2461	0.0124
$R_a$	0.8912	0.8982	0.7894	0.7948	0.6948	0.6996	0.6384	0.6399
$\bar{P}$	31.217	30.697	30.792	30.268	29.294	28.729	28.602	28.038



(a) Original algorithm

(b) Improved algorithm

FIGURE 3: Image quality of the decoded 34th frame of *Claire* sequence at bit rate 150 Kbps.



(a) Original algorithm

(b) Improved algorithm

FIGURE 4: Image quality of the decoded 34th frame of *Alex* sequence at bit rate 150 Kbps.

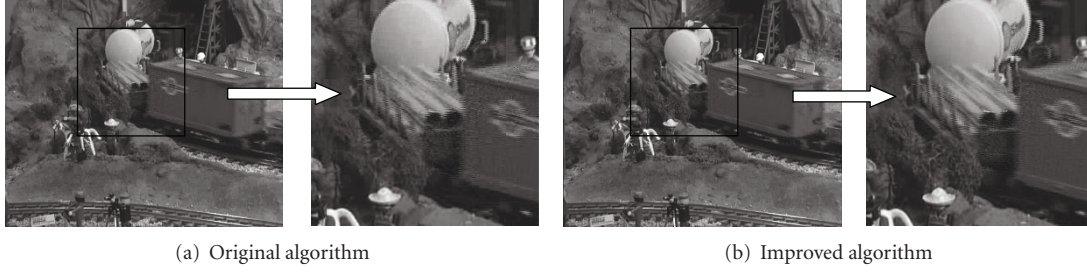
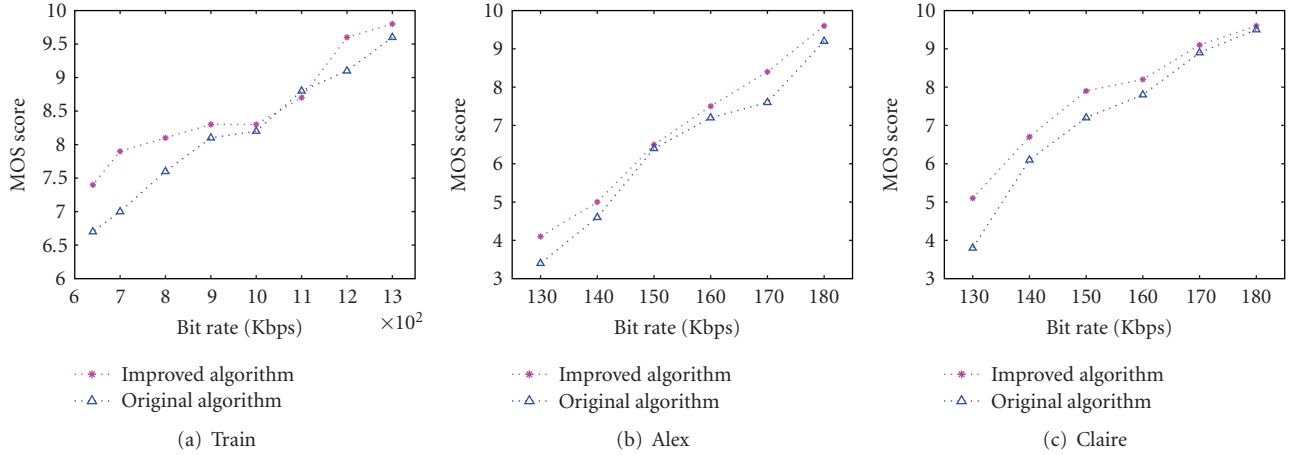
FIGURE 5: Image quality of the decoded 19th frame of *Train* sequence at the bit rate 1500 Kbps.

FIGURE 6: Subjective test results.

- (1) Firstly, let  $Q_i$  denote the quantization step of the  $i$ th frame, and define

$$x_{1i} = \frac{1}{Q_i^2}, \quad x_{2i} = \frac{1}{\sqrt{Q_i}}. \quad (5)$$

Suppose that  $(x_{1i}, x_{2i}, R_i)$  ( $i = 1, \dots, n$ ) are existing samples from previously coded frames, where  $n$  is the number of the previously coded frames and  $R_i$  is the corresponding bit rate of the  $i$ th frame.

- (2) Secondly, define

$$M = \begin{pmatrix} 1 & x_{11} & x_{21} \\ 1 & x_{12} & x_{22} \\ \vdots & \vdots & \vdots \\ 1 & x_{1n} & x_{2n} \end{pmatrix}, \quad R = \begin{pmatrix} R_1 \\ R_2 \\ \vdots \\ R_n \end{pmatrix}, \quad C = \begin{pmatrix} c \\ b \\ a \end{pmatrix}. \quad (6)$$

Then (4) can be rewritten in matrix form as  $MC = R$  using (6).

- (3) Finally, based on linear regression method, the parameters in  $C$  can be estimated by

$$C = (M^T M)^{-1} M^T R, \quad (7)$$

where  $M^T$  is the transpose of  $M$  and  $(M^T M)^{-1}$  is the inverse matrix of  $M^T M$ .

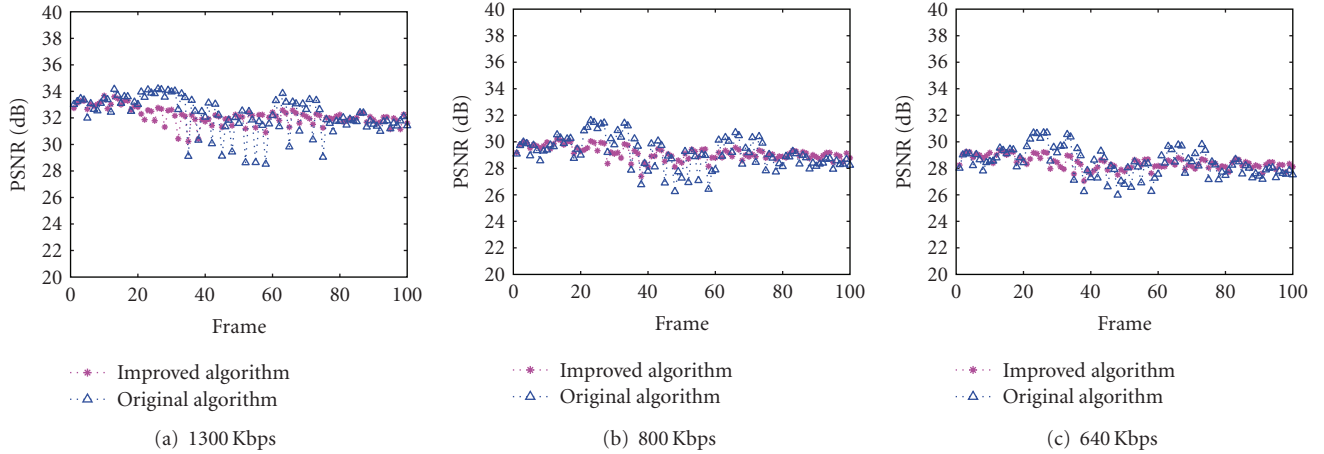
In order to evaluate the performance of the new proposed R-D model, the actual R-Q data are also fitted by the proposed R-D model, and the fitted result is compared with that of the quadratic model. Some simulation results are shown in Figure 1, where the actual R-Q curve and the fitted R-Q curve of each frame with two models are drawn on the same figure. From the simulation results, it can be found that, compared with the old model, the new one can more accurately approach the actual R-Q relation.

### 3. HVS-Based Rate Control

G012 can be viewed as performing rate control at three different layers: GOP layer, frame layer, and basic unit (BU) layer. Within a BU, all the macroblocks (MB) are encoded with the same quantization parameter, which does not consider the visual sensitivity difference between MBs. Thus, it is hard to obtain perceptually consistent results in practical applications.

HVS is so complex that it is impossible to incorporate the whole HVS characteristics during rate control and video coding. This paper only discusses several important HVS characteristics including the contrast sensitivity, multichannel theory, and masking effect [15–17]. HVS-related research findings contend that there exists a threshold contrast for an observer to detect the change in intensity. The inverse of the contrast threshold is usually defined as contrast




 FIGURE 7: Frame by frame image quality of *Train* sequence at several bitrates.

sensitivity, which is a function of spatial frequency with a bandpass nature [18]. The multichannel theory of human vision states that different stimuli are processed in different channels in the human visual system. Both physiological and psychophysical experiments carried out on perception have given the evidence of the bandpass nature of the cortical cells' response in the spectral domain, and the human brain seems therefore to possess a collection of separate mechanisms, each being more sensitive to a portion of the frequency domain. Masking effect can be explained as the interaction among stimuli. The detection threshold of a stimulus may vary due to the existence of another. This kind of variation can be either positive or negative. Due to masking effect, similar artifacts may be disturbing in certain regions of an image while they are hardly noticeable elsewhere. Hence, during rate control the masking effect can be used to deal with different scenarios with different tips to acquire perceptually consistent result [19].

In the proposed algorithm, the quantization step for each MB will be adjusted by its visual sensitivity. MBs with high visual sensitivity will be encoded with small quantization parameters, and MBs with low visual sensitivities are encoded with large quantization parameters so that better subjective visual quality could be obtained under the same given bit rate constraint. In order to incorporate HVS characteristics, three new modules are added to the original G012 framework, which are preprocessing module, visual sensitivity calculation module, and quantization step modification module. The three new added modules are briefly introduced as follows.

**3.1. Preprocessing.** In this module, the background intensity and the average spatial frequency of each image are calculated. For an image, its background intensity,  $I_B$ , is calculated by averaging the values of all pixels

$$I_B = \frac{1}{N_R \times N_C} \sum_{i=0}^{N_R \times N_C - 1} X_i, \quad (8)$$

where  $X_i$  is the value of the  $i$ th pixel,  $N_R$  denotes the image width, and  $N_C$  denotes the image height.

The average spatial frequency,  $f_{av}$ , is defined as the average spatial frequency of all the MBs, that is,

$$f_{av} = \frac{1}{\text{NUM}} \sum_{i=0}^{\text{NUM}-1} f_i, \quad (9)$$

where

$$f_i = \sqrt{f_{xi}^2 + f_{yi}^2},$$

$$f_{xi} = \sqrt{\frac{1}{256} \sum_{j=0}^{15} \sum_{k=0}^{15} [X_i(j, k) - X_i(j, k-1)]^2}, \quad (10)$$

$$f_{yi} = \sqrt{\frac{1}{256} \sum_{j=0}^{15} \sum_{k=0}^{15} [X_i(j, k) - X_i(j-1, k)]^2},$$

where NUM denotes the number of MBs in the image,  $f_i$  denotes the spatial frequency of the  $i$ th MB,  $f_{xi}$  and  $f_{yi}$  denote the horizontal and vertical frequencies, respectively, and  $X(j, k)$  denotes the  $(j, k)$ th pixel value.

**3.2. Calculation of Visual Sensitivity.** In the proposed algorithm, the quantization step for an MB will be adjusted according to its visual sensitivity. For each MB, we consider its visual sensitivity from four aspects: motion sensitivity, brightness sensitivity, contrast sensitivity, and position sensitivity.

**(1) Motion Sensitivity.** Studies on human perception indicate that motion has a strong influence on visual attention, and people often pay more attention to fast moving objects than to low or still background regions [20, 21]. Hence, compared with still background regions, the moving regions are considered to have high visual sensitivities. In the proposed algorithm, each frame is divided into moving

regions and background regions, and different sensitivities are defined for each region

$$S_M = \begin{cases} a_1, & \text{background,} \\ a_2, & \text{moving object,} \end{cases} \quad (11)$$

where  $S_M$  denotes the motion sensitivity,  $a_1$  and  $a_2$  denote the sensitivities of background region and moving object, respectively. In this paper,  $a_1$  is set to 0.5, and  $a_2$  is set to 1.5. For moving objects detection, our earlier proposed object extraction and tracking algorithm based on spatio-temporal information is employed [22].

(2) *Brightness Sensitivity.* For a digital image, its pixels' values are represented by a finite number of intensity levels. For example, a gray image with 8 bits per pixel is represented by 256 intensity levels, where 0 usually corresponds to the darkest level and 255 corresponds to the brightest level. According to the Weber's law, the perceived brightness of HVS is a nonlinear function of intensity, which can be modeled by nonlinear monotonically increasing functions. Various models have been proposed such as the logarithmic model, and cube root model. In this paper, the following nonlinear model is used:

$$B = \begin{cases} 0, & 0 \leq I \leq I_{TR}, \\ \frac{B_{\max}}{2} \left( \frac{2(I - I_{TR})}{I_{\max} - I_{TR}} \right)^2, & I_{TR} \leq I < \frac{(I_{\max} + I_{TR})}{2}, \\ B_{\max} - \frac{B_{\max}}{2} \left( \frac{2(I - I_{TR})}{I_{\max} - I_{TR}} \right)^2, & \frac{(I_{\max} + I_{TR})}{2} < I \leq I_{\max}, \end{cases} \quad (12)$$

where  $I$  denotes image intensity,  $B$  denotes the corresponding perceived brightness,  $I_{TR} = 20$  is the threshold for intensity,  $I_{\max} = 255$  is the maximum value of intensity, and  $B_{\max} = 100$  is the maximum value of perceived brightness [23]. Suppose that  $B_{\text{avg}}$  is the perceived global brightness of an image and  $B_m$  is the perceived brightness of the current MB, then the brightness sensitivity of the MB,  $S_B$ , is defined by

$$S_B = \frac{B_m}{B_{\text{avg}}}. \quad (13)$$

(3) *Contrast Sensitivity.* HVS-related researches have shown that HVS is not equally sensitive to all spatial frequencies and CSFs are defined to describe the frequency response of HVS. Various CSFs have been proposed, and most of them are bandpass in nature, but maximum at different frequencies. In practical applications, due to the difference of video content, the peak frequencies are also various for different video

sequences. Hence, the following improved CSF function is used in this paper:

$$S_C = \begin{cases} (0.0512 + 0.8512f)e^{-0.3192f}, & f \leq 3, \\ 1, & 3 < f < f_0, \\ e^{-0.1(f-f_0)^{1.1}}, & f \geq f_0, \end{cases} \quad (14)$$

where  $f_0$  denotes the peak frequency,  $S_C$  denotes the contrast sensitivity, and  $f$  denotes the spatial frequency in cycles per degree.

(4) *Position Sensitivity.* Many HVS-related research experiments have indicated that people usually pay more attention to the central region than to the side regions, and the visual sensitivity decreases progressively from central region to side regions. Based on this observation, we define a new sensitivity, position sensitivity termed  $S_P$ , to model this effect

$$S_P = \frac{\sqrt{(x - x_c)^2 + (y - y_c)^2}}{\max}, \quad (15)$$

where  $(x, y)$  denotes the center of the current MB,  $(x_c, y_c)$  denotes the center of the image, and  $\max$  represents the maximum distance from the side to the center of the image.

3.3. *Quantization Parameter Modification Based on Visual Sensitivity.* In our proposed algorithm, the quantization step for each MB will be adjusted by its visual sensitivity. For each MB, we compute a quantization adjustment according to each kind of sensitivity. The finally overall adjustment for each MB is the sum of them. The quantification adjustment induced by each kind of sensitivity is calculated as follows.

(1) The motion sensitivity-induced quantization adjustment,  $Q_M(S_M)$ , is defined by

$$Q_M(S_M) = \begin{cases} 1, & S_M < 1, \\ -1, & S_M > 1. \end{cases} \quad (16)$$

(2) The brightness sensitivity-induced quantization adjustment,  $Q_B(S_B)$ , is defined by

$$Q_B(S_B) = \begin{cases} 1 + \Delta L, & S_B \leq K_1, \\ 0 + \Delta L, & K_1 < S_B < K_2, \\ -1 + \Delta L, & S_B \geq K_2, \end{cases} \quad (17)$$

where  $\Delta L$  is a weighting factor of background intensity,  $K_1$  and  $K_2$  are thresholds for intensity.

(3) The contrast sensitivity-induced quantization adjustment,  $Q_C(S_C)$ , is defined by

$$Q_C(S_C, f, f_0) = \begin{cases} 2, & S_C < 0.6, \quad f \leq 3, \\ -1, & 0.6 \leq S_C < 1, \quad f \leq 3, \\ -2, & S_C = 1, \\ -1, & 0.6 \leq S_C < 1, \quad f \geq f_0, \\ 2, & S_C < 0.6, \quad f \geq f_0. \end{cases} \quad (18)$$



- (4) The position sensitivity-induced quantization adjustment,  $Q_P(S_P)$ , is defined by

$$Q_P(S_P) = \begin{cases} -2, & S_P < 0.25, \\ -1, & 0.25 \leq S_P < 0.5, \\ 0, & \text{otherwise.} \end{cases} \quad (19)$$

Finally, the overall quantization adjustment for each MB  $Q_{\text{adjust}}$  is given by

$$Q_{\text{adjust}} = Q_M(S_M) + Q_B(S_B) + Q_C(S_C) + Q_P(S_P). \quad (20)$$

And during rate control and video coding, the finally used quantization step,  $Q_f$ , is determined by

$$Q_f = Q + Q_{\text{adjust}}, \quad (21)$$

where  $Q$  is the quantization step for the whole BU acquired during BU layer rate control. For practical rate control, MB-based quantization adjustment may lead to visual discomfort such as artifact. Hence, in our actual implementations, the quantization adjustment  $Q_{\text{adjust}}$  for each MB is confined to  $[-Q/2, Q/2]$ .

## 4. Experimental Results

In order to evaluate the performance of the improved rate-control algorithm, experiments are performed based on our improved JM 10.2 platform. Some test parameters are given in Table 1. Several test sequences are used including the *Alex*, *Claire*, and *Train* sequences. For each sequence, over 100 frames are encoded. We mainly investigate the rate-control precision and the image quality of decoded videos. Therefore, two metrics are defined for algorithm evaluation, rate-control deviation  $\delta_d$  and average peak signal-to-noise ratio (PSNR)  $\bar{P}$ .  $\delta_d$  is defined to reflect the rate-control precision. The smaller the  $\delta_d$  is, the better the rate-control precision will be. The two metrics are defined as

$$\delta_d = \frac{|R_a - R_T|}{R_T}, \quad (22)$$

$$\bar{P} = \frac{1}{N_{\text{frame}}} \sum_{i=1}^{N_{\text{frame}}} P_i,$$

where  $R_T$  and  $R_a$  are the target and actual bitrates, respectively,  $N_{\text{frame}}$  is the number of frames encoded, and  $P_i$  denotes the PSNR of the  $i$ th frame.

To evaluate the proposed algorithm objectively, subjective tests are also performed. To implement subjective evaluation, our test procedure is set up similar to the double stimulus impairment scale (DSIS) method formalized in ITU-R Recommendation BT.500-10 [24]. In our experiments, for every test sequence, both the decoded versions with the original and the proposed algorithms are presented to the observers under the same lighting and viewing conditions. Fifteen observers participated in the test, and they are asked to rate the perceptual quality of the decoded

versions and make a mean opinion score (MOS) between 0 and 10 for each version. The mean opinion scores are made according to the principle given in Table 2.

Partial experimental results are shown in Figures 2, 3, 4, 5, 6, and 7 and Tables 3, 4, 5, 6, and 7. Table 3 shows the average intensities and the average spatial frequencies of the test sequences. Figure 2 gives the moving segmentation results. Tables 4, 5, 6, and 7 show partial comparison results of the two algorithms with respect to rate control precision, average PSNR, while Figures 3–6 show partial results with respect to subjective quality of the decoded image sequences. Figure 7 shows some frame by frame results in terms of PSNR of *Train* sequence at some bitrates. From Tables 4–7, it can be observed that, compared with the original algorithm, the rate control precision of the proposed algorithm is improved obviously at most bit-rates, which is mainly attributed to the improved R-Q model. From Figures 3–7, it can be seen that, though the average PSNR of reconstructed video of the proposed algorithm decreases slightly, the overall subjective visual quality is enhanced due to the incorporation of main HVS characteristics. From the results in Figure 7, one can also tell that, compared with the original one, the image quality of the successive frames changes more smoothly in the proposed algorithm.

However, rate control is a comprehensive technology whose performance is affected by many factors in practical applications. There also exist some limitations for our algorithm. For example, we also used the same temporal Mean Absolute Difference (MAD) prediction method as JVT-G012, which has the following drawbacks: (a) inaccurate MAD prediction for sudden changes; (b) inaccurate bit allocation since the header bits of the current basic unit are predicted from previous encoded basic units. Another limitation is the algorithm complexity. Compared with the original one, the proposed algorithm has higher computational load. Our future work will be on how to handle these limitations.

## 5. Conclusion

Rate control plays an important role in the field of video coding and transmission, and its performance is of great importance to the video coding effectiveness. In this paper, by adapting quadratic R-D model and incorporating the main HVS characteristics, the classic JVT-G012 rate control algorithm has been improved, which can both improve the rate control precision and enhance the overall subjective quality of the reconstructed video images. However, rate control is a comprehensive technology whose performance is affected by many factors in practical applications, and there still exist some limitations for the proposed algorithm. Hence, it should be further optimized if only better performance is expected.

## Acknowledgments

This paper was supported in part by the National Natural Science Foundation of China (no. 60902066, 60872094,

60832003), the Zhejiang Provincial Natural Science Foundation (no. Y107740), the Projects of Chinese Ministry of Education (no. 200816460003), the Open Project Foundation of Ningbo Key Laboratory of DSP (no. 2007A22002), and the Scientific Research Fund of Zhejiang Provincial Education Department (Grant no. Z200909361).

## References

- [1] L. Merritt and R. Vanam, "Improved rate control and motion estimation for H.264 encoder," in *Proceedings of the 14th IEEE International Conference on Image Processing (ICIP '07)*, pp. V309–V312, September 2007.
- [2] N. Kamaci, Y. Altunbasak, and R. M. Mersereau, "Frame bit allocation for the H.264/AVC video coder via cauchy-density-based rate and distortion models," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 15, no. 8, pp. 994–1006, 2005.
- [3] Z. Chen and K. N. Ngan, "Recent advances in rate control for video coding," *Signal Processing: Image Communication*, vol. 22, no. 1, pp. 19–38, 2007.
- [4] Z. G. Li, F. Pan, K. P. Lin, G. Feng, X. Lin, and S. Rahardja, "Adaptive basic unit layer rate control for JVT," in *Proceedings of the 7th JVT Meeting*, Pattaya II, Thailand, 2003.
- [5] W. Yuan, *Research for the rate control algorithm in H.264*, M.S. thesis, Hefei University of Technology, China, 2006.
- [6] X. Yi and N. Ling, "Improved H.264 rate control by enhanced MAD-based frame complexity prediction," *Journal of Visual Communication and Image Representation*, vol. 17, no. 2, pp. 407–424, 2006.
- [7] J.-Y. Kim, S.-H. Kim, and Y.-S. Ho, "A frame-layer rate control algorithm for H.264 using rate-dependent mode selection," in *Proceedings of the 6th Pacific Rim Conference on Multimedia (PCM '05)*, vol. 3768 of *Lecture Notes in Computer Science*, pp. 477–488, 2005.
- [8] S. Lee, M. S. Pattichis, and A. C. Bovik, "Foveated video compression with optimal rate control," *IEEE Transactions on Image Processing*, vol. 10, no. 7, pp. 977–992, 2001.
- [9] A. G. Nguyen and J.-N. Hwang, "A novel hybrid HVPC/mathematical model rate control for low bit-rate streaming video," *Signal Processing: Image Communication*, vol. 17, no. 5, pp. 423–440, 2002.
- [10] C.-W. Tang, C.-H. Chen, Y.-H. Yu, and C.-J. Tsai, "Visual sensitivity guided bit allocation for video coding," *IEEE Transactions on Multimedia*, vol. 8, no. 1, pp. 11–18, 2006.
- [11] Z. Z. Chen and K. N. Ngan, "A unified framework of unsupervised subjective optimized bit allocation for video coding using visual attention model," in *Multimedia Systems and Applications VIII*, Proceedings of SPIE, Boston, Mass, USA, 2005.
- [12] Z. Liu, L. J. Karam, and A. B. Watson, "JPEG2000 encoding with perceptual distortion control," *IEEE Transactions on Image Processing*, vol. 15, no. 7, pp. 1763–1778, 2006.
- [13] M. Jiang and N. Ling, "On lagrange multiplier and quantizer adjustment for H.264 frame-layer video rate control," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 16, no. 5, pp. 663–669, 2006.
- [14] T. Chiang and Y.-Q. Zhang, "A new rate control scheme using quadratic rate distortion model," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 7, no. 1, pp. 246–250, 1997.
- [15] Z. Wang, R. H. Sheikh, and C. A. Bovik, "Objective video quality assessment," in *The Handbook of Video Databases: Design and Applications*, B. Furht and O. Marqure, Eds., pp. 1041–1078, CRC Press, Boca Raton, Fla, USA, 2003.
- [16] E. P. Simoncelli and B. A. Olshausen, "Natural image statistics and neural representation," *Annual Review of Neuroscience*, vol. 24, pp. 1193–1216, 2001.
- [17] H. R. Sheikh and A. C. Bovik, "Image information and visual quality," *IEEE Transactions on Image Processing*, vol. 15, no. 2, pp. 430–444, 2006.
- [18] C. K. Wei, *Image quality assessment model via HVS*, M.S. thesis, National University of Defense Technology, China, 2003.
- [19] J. L. Mannos and D. J. Sakrison, "The effects of a visual fidelity criterion on the encoding of images," *IEEE Transactions on Information Theory*, vol. 20, no. 4, pp. 525–536, 1974.
- [20] W.-H. Cheng, W.-T. Chu, J.-H. Kuo, and J.-L. Wu, "Automatic video region-of-interest determination based on user attention model," in *Proceedings of IEEE International Symposium on Circuits and Systems (ISCAS '05)*, pp. 3219–3222, May 2005.
- [21] W. Osberger, A. Maeder, and N. Bergmann, "A technique for image quality assessment based on a human visual system model," in *Proceedings of the European Signal Processing Conference*, pp. 1049–1052, 1998.
- [22] Z. J. Zhu, G. Y. Jiang, M. Yu, and X. W. Wu, "New algorithm for extracting moving object based on spatio-temporal information," *Chinese Journal of Image and Graphics*, vol. 8, no. 4, pp. 422–425, 2003.
- [23] G. Ivkovic and R. Sankar, "An algorithm for image quality assessment," in *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '04)*, pp. 713–716, May 2004.
- [24] ITU-R, *Methodology for the subjective assessment of the quality of television pictures*, ITU-R Recommendation BT.500-10, Geneva, Switzerland, 2000.