A Novel Rate Control Algorithm for H.264/AVC Based on Human Visual System

Jiangying Zhu, Mei Yu, Qiaoyan Zheng, Zongju Peng, Feng Shao, Fucui Li, and Gangyi Jiang

Faculty of Information Science and Engineering, Ningbo University
315211 Ningbo, China

Abstract. To improve performance of rate control algorithm for H.264/AVC, and keep a better control accuracy of the output of the compressed video stream, a novel rate control algorithm based on human visual system (HVS) is proposed in this paper. The proposed rate control algorithm consists of two layers: frame level and basic unit (BU) level. In frame level, changed scene is first detected and frame difference ratio is utilized to represent the motion complexity of the frame, and then target bit for frame level are allocated by considering the two factors. In BU level, by analyzing motion information, texture characteristics and the location of the frames, visual sensitivity of a macroblock is first measured, and the bit is allocated for the macroblock based on the sensitivity factor. Experimental results show that the proposed method can provide an improved visual quality and higher PSNR while almost the same control accuracy, compared with traditional rate control method.

Keywords: H.264/AVC, rate control, scene change detection, visual sensitivity.

1 Introduction

In recent years, it is crucial to maintain a good balance between the allocated rate and the video quality under constraint of bandwidth requirement. Rate control has been widely studied in digital video coding standards and applications, such as TM5 for MPEG-2 [1], TMN8 for H.263 [2], VM8 for MPEG-4 [3] and JVT-G012 for H.264/AVC [4]. It dynamically adjusts encoder parameters to achieve a target bit rate and effective rate control algorithm can result in high video quality, low video quality fluctuation, and a low mismatch between the target and the actual encoded bit rates.

In the literatures, many RC algorithms had been proposed to improve the quality in H.264 coding. The JVT-G012 model used a fluid flow traffic model to compute the target bits for the current encoding frame and a linear model to predict mean absolute difference (MAD) to solve the chicken and egg dilemma. To improve performance of rate control algorithm for JVT-G012, Lee et al. used the structural and statistical features of local textures to adaptively set proper initial QP values for versatile video contents and peak signal-to-noise ratio (PSNR) variation-limited bit-allocation [5]. In addition, Sun et al. proposed a new rate-complexity-quantization model and an

Y.-S. Ho (Ed.): PSIVT 2011, Part II, LNCS 7088, pp. 227–239, 2011.

incremental rate control algorithm for H.264/AVC video coding [6], by estimating the picture complexity and rate-quantization modeling in an incremental rate control for P-frames. Xie *et al.* used texture-complexity as an innovative coding characteristic [7], based on the discrete cosine transform (DCT) coefficients distribution and distortion-quantization (DQ) relationship, to propose a general rate-distortion (RD) model for block-based video coding. There are many works aimed at reducing computational complexity and improving prediction accuracy [8-10]. Zhao *et al.* presented an effective macro block layer rate control algorithm [11], analyzing the temporal-spatial correlation and object direction, and calculating QP based on a more accurate header bits prediction model. However, these methods do not take into account human visual system in rate control; therefore, many scholars recently carry out deeply research in this connection. Chen *et al.* proposed a perceptually-friendly H.264/AVC video coding based on foveated just-noticeable-distortion [12]. Tang *et al.* used the loss of human visual sensitivity model to allocation the bits [13]. The common approach is based on region of interest rate control algorithm [14]. Zheng *et al.* proposed a human visual system based rate control algorithm [15], by analyzing the motion complexity of the frames and human visual sensitivity, to control bit allocation. For some sequences with high motion or scene changes, these approaches failed to predict the abrupt changes of MAD. In order to improve these problems, we propose a more accurate MAD model based on human visual system (HVS).

In this paper, we propose a simple and effective model based on HVS, especially for high motions areas or scene changes areas. The frame difference ratio between the current frame and its previous frame is defined to describe frame complexity, while visual sensitivity factor is defined to describe BU complexity, which based on motion information, texture characteristics and the location of the frames. Meanwhile, by thoroughly analyzing the temporal–spatial correlation, a more accurate MAD model is presented to improve the quality of frames. The experimental results show the effectiveness of the proposed scheme.

2 Analysis of the Rate Control Algorithm in JVT-G012

There are three layers in H.264/AVC bits allocation: GOP-layer, frame-layer and BU layer. The rate control algorithm in JVT-G012 is adaptive BU layer rate control. BU can be a frame, a slice, or an MB. The algorithm is described as follows.

1) Compute the target bit for the current frame

The target bits allocated for each frame in the GOP is determined based on buffer occupancy and the number of remaining bits. Meanwhile, in order to keep consistent the visual quality, the relative complexity of the current frame also is considered. T_i is the number of target bits generate by the ith GOP, which is computed by

$$T_i = \beta \times T_i' + (1-\beta) \times T_i'' \tag{1}$$

where β is a constant and its typical value is 0.5 if there is no B frame, T_i' is the number of frame bits which is computed by using the hypothetical reference decode

(HRD) model, and T_i'' is the number of frame bits which is computed by remaining bits in the video sequence and the relative complexity in the frame (W_p,W_b), N_{pr} and N_{br} are the number of remaining bits of P frame and B frame, which is computed by

$$T_i'' = \frac{W_p T_r}{W_p N_{pr} + W_b N_{br}} \tag{2}$$

In the case that there is no B frame, equation (2) can be simplified as

$$T_i'' = \frac{T_r}{N_{pr}} \tag{3}$$

2) Allocate the target bit for the basic unit

In JVT-G012, a linear model is used to predict the MAD of current basic unit in the current frame by that of the co-located basic unit of the previous frame. Suppose that the predicted MAD of current basic unit in the current frame and the actual MAD of the co-located basic unit in the previous frame are denoted by MAD_{cb} and MAD_{pb}, respectively. The linear prediction model is then given by

$$MAD_{cb} = a_1 \times MAD_{pb} + a_2 \tag{4}$$

where a_1 and a_2 are two coefficients of the prediction model. The initial values of a_1 and a_2 are set to 1 and 0, respectively. They are updated after coding each basic unit. If T_{f_r} is the number of remaining bits of previous frame, N is the number of BU in the current frame. T_{f_i} is the target bit for the current BU is computed by

$$T_{f\_i} = T_{f\_r} \times \frac{MAD_i^2}{\sum_{j=i}^{N} MAD_j^2} \tag{5}$$

3) Compute the corresponding parameter by using the quadratic R-D model

The parameter of the used quadratic R-D model is computed by

$$T - H = \left(\frac{C_1}{Q_{step}} + \frac{C_2}{Q_{step}^{\;2}}\right) \times MAD \tag{6}$$

where T is the total number of target bits, H is the header bits, C_1 and C_2 are two coefficients of the model, Q_{step} is quantization step.

If there is no B frame in equation (3), the remaining bits are allocated to all frames equally, while ignoring the image complexity. Therefore, there is a bigger difference if the sequences with high motion or scene changes, resulting in significant fluctuations with the PSNR, especially at low bit rate.

In equation (4), there are two problems for MAD prediction. One is the reference MAD information used in the model, only co-located MB in the previous frame selected to predict the MAD value. For some sequences with high motion or scene changes, this prediction model will fail to predict the abrupt changes of MAD, consequently resulting in a weak rate control. The other is the two coefficients of prediction model updated by linear regress will introduce the computation cost to the already existed computation-demanding H.264/AVC encoder.

3 Improved Rate Control Method Based on HVS

3.1 Bits Allocation at Frame Level

To better allocate the target bits for every frame video that consider scene change or high motion and the complexity of current frame content. More bits is allocate to scene change frames or high complexity frames, and fewer bits for low complexity frames or unimportance frames to achieve high video quality.

There are some types of video scene changes: mutation scene change, melting, fading and so on. There are three algorithms with fast scene change detection: the detection based on gray value, motion and search for edges. The latter two kinds of detection algorithms with good performance, but the algorithm of high complexity greatly limits their applications, especially in the high demand for real-time video communication rate control algorithm. Accordingly, the current frame and reference frame use the three components of the mean absolute difference in the current frames for judging whether a scene change, the difference function is given by

$$D_x(Scur, Sref) = \left| mean(Scur, x) - mean(Sref, x) \right| \tag{7}$$

where *mean(X)* denotes mean function, *Scur* and *Sref* are the reconstructed frames of the current frame and previous frame; *X* is the component in YUV. To determine whether a scene change, equations (8) and (9) are used.

$$\frac{D_Y(S_i, S_{i-1})}{mean(S_i, Y)} + \frac{D_U(S_i, S_{i-1})}{mean(S_i, U)} + \frac{D_V(S_i, S_{i-1})}{mean(S_i, V)} \geq t_{TH1} \tag{8}$$

$$D_Y(S_i, S_{i-1}) + D_U(S_i, S_{i-1}) + D_V(S_i, S_{i-1}) \geq t_{TH2} \tag{9}$$

The current frame is scene change frame when equations (8) and (9) are both true, t_{TH1} and t_{TH2} are the decision threshold, which are computed by

$$t_{TH1} = \frac{\left| mean(D_Y(S_i, S_{i-1})) \right|}{mean(Y)} + \frac{\left| mean(D_U(S_i, S_{i-1})) \right|}{mean(U)} + \frac{\left| mean(D_V(S_i, S_{i-1})) \right|}{mean(V)} \tag{10}$$

$$t_{TH2} = mean(D_Y(S_i, S_{i-1})) + mean(D_U(S_i, S_{i-1})) + mean(D_V(S_i, S_{i-1})) \tag{11}$$

We use frame difference function describe the frame complexity, shown as

$$D_i = \sum_{y=0}^{y=H-1} \sum_{x=0}^{x=W-1} \left| I_C(x,y) - I_P(x,y) \right| \tag{12}$$

where $I_c(x, y)$ and $I_p(x, y)$ are the luminance value of the current frame and previous frame in the pixel (x, y), D_i is the ith frame of the frame difference. In general, if there is strong motion or scene change between two frames, there is larger value of the frame difference. Then, Eq.(3) is further bounded

$$T_i^{''} = \alpha \times \frac{T_r}{N_{p,r}} \tag{13}$$

$$\alpha = \min\left\{ \max\left\{ \frac{D_i}{D_{i-1}}, k_1 \right\}, k_2 \right\} \tag{14}$$

where a is the frame complexity, D_i is the frame difference value of the ith frame that is non-coded, and D_{i-1} is the frame difference value of previous frame. The value of coefficient a has a limit region; the constants k_1 and k_2 come from experiments. Separately, if the current frame detected as scene change, $a=MAX$ (k_1, k_2).To smooth the visual quality, the value of k_1 and k_2 are typical set to 0.5 and 2.

3.2 Bits Allocation at BU Level

To achieve good subject quality, rate control should consider the characters of Human Visual System. As we all known, the regions of interest often focus on the middle position of the image, the movement of objects and complex texture of the object, when we note a video sequence. The same amount of MAD in the video sequence of different regions will have distinct subjective experience: visually sensitive areas of the visual distortion caused significantly greater than in other regions. Therefore, taking into account the human visual systems, in this paper we first define a reasonable visual factor in the BU level for target bit allocation.

1) Visual sensitivity measure model

Visual sensitivity of each MB is corresponded with position, movement and the complexity of the texture. In order to get visual weight conform to human visual sensitivity, if it is sensitivity to human, the value of weight great than 1, otherwise less than 1.In the JVT-G012, the visual weight in BU layer are the same (namely the weighting factor W is set to 1).Firstly, we should compute the value of W, $W \in [W_{MIN}, W_{MAX}]$ ($W_{MIN} <1$, $W_{MAX} >1$), then, using W to allocate target bits at BU layer. While BU is composed by several macro blocks, we should calculate the location of each macro block, and the complexity of texture.

Gaussian function is used to determine the location weight of each MB, due to the objects in the central region of the image more attractive eyes than in its external edge of the area. Suppose that the coordinate of the image center is (*xc, yc*), MB's location weight at (*x, y*) is defined by

$$L_{MB}(x,y) = \exp(-\frac{(x-xc)^2 + (y-yc)^2}{2\sigma^2}) \tag{15}$$

where σ is the scale parameter of Gaussian function, the smaller the value of σ, the faster the speed of reducing. If $\delta = min\ ((W\text{-}1)/2,\ (H\text{-}1)/2)$, equation (15) exactly describes the location weight of the macro block, where *W* and *H* are image's weight and height.

Here, we use mean absolute difference between in the co-located position of the current and previous frames to describe motion degree of MB, in order to measure the activity of the MB, which is shown in Eq. (16)

$$A_{MB}(n,m) = \frac{1}{256}\sum_{i,j}^{16}\left|f_{n,m}(i,j) - f_{n-1,m}(i,j)\right| \tag{16}$$

where $f_{n,m}(i, j)$ and $f_{n\text{-}1,m}(i, j)$ denote the luminance value in the pixel (*i, j*) of the *m*th MB in the *n*th and (*n-1*)th frame. Obviously, the movement area with a large value and the method will not bring too much computational complexity. Therefore, the MB activities measure has a new definition with the location weight of each MB (L_{MB})

$$A_{MB}' = L_{MB} \times A_{MB} \tag{17}$$

Then *mot* (*i, j*) denotes regular of the MB activities, as shown in Eq. (18)

$$mot(i,j) = \frac{2 \times A(i,j) + A_{avg}}{A(i,j) + 2 \times A_{avg}} \tag{18}$$

where $A(i, j)$ is the activities of the *j*th BU in the *i*th frame, A_{avg} is the average activities of the remaining BU in current frame.

Due to take into account the human eyes are more sensitive to the distortion on the edges of objects, we use gradient to get the object edges as the complexity of the MB texture. While the complex dynamic objects in the static area are not attractive eye's attention like in the dynamic region, so we directly detect edge to the difference image Δf between adjacent frames rather than the original image, and then the difference for each pixel separately do convolution with two operators, it will create a corresponding edge vector $\overline{Grad(i,j)} = (gradx(i,j), grady(i,j))$,which are computed by

$$\begin{aligned} gradx(i,j) = {} & \Delta f(i-1,j+1) + 2\Delta f(i,j+1) + \Delta f(i+1,j+1) - \\ & \Delta f(i-1,j) - 2\Delta f(i,j-1) - \Delta f(i+1,j-1) \end{aligned} \tag{19}$$

$$\begin{aligned} grady(i,j) = {} & \Delta f(i+1,j-1) + 2\Delta f(i+1,j) + \Delta f(i+1,j+1) - \\ & \Delta f(i-1,j-1) - 2\Delta f(i-1,j) - \Delta f(i-1,j+1) \end{aligned} \tag{20}$$

where $\Delta f\,(i, j)$ denotes the luminance value of the pixel (i, j) in the difference image Δf. So the edge vector of the pixel (i, j) in the difference image Δf is computed by

$$Grad(i, j) = \sqrt{gradx(i, j)^2 + grady(i, j)^2} \tag{21}$$

The complexity of the MB texture ($text_{MB}$) denotes the edge vector average value of all pixels in MB is computed as follows

$$text_{MB} = \frac{1}{256}\sum_{i,j}^{16} Grad(i, j) \tag{22}$$

The complexity of the MB is weighted with the location weight of each MB (L_{MB})

$$text_{MB} = L_{MB} \times text_{MB} \tag{23}$$

Then $texture\,(i, j)$ denotes the texture complexity of the jth BU in the ith frame

$$texture(i, j) = \frac{2 \times text(i, j) + text_{avg}}{text(i, j) + 2 \times text_{avg}} \tag{24}$$

Let $W\,(i, j)$ denote visual sensitivity of the jth BU in the ith frame, and it is computed by

$$W_{(i,j)} = w_1 \times mot(i, j) + w_2 \times texture(i, j) \tag{25}$$

where w_1 and w_2 are the weighting coefficients, whose typical values must fulfill the conditions: $0 \le w_1, w_2 \le 1$ and $w_1 + w_2 = 1$, which come from experiments.

2) A novel MAD prediction model

Based on the previous analysis, MAD predictions may be not very accurate, so a new MAD prediction model is presented by

$$MAD_{cb} = a_1 \times MAD_{pb}' + a_2 \tag{26}$$

where MAD_{pb}' denotes a new measure for evaluating the difference between the current original frame and previous reconstructed frame. In order to reduce the computational complexity, we use three kinds of MB included co-located MB, left MB and upper MB to predict MAD for current MB, which is computed by

$$MAD_{pb}' = \frac{a \times MAD_{pb} + b \times MAD_L + c \times MAD_U}{a + b + c} \tag{27}$$

where a, b and c represent the weighting coefficients of a MB, and $a+b+c=1$. The MAD_L and MAD_U are the actual MAD values of the left-side MB and upper-side MB.

If the current MB is the first MB in the current Frame, Eq. (27) is equal to

$$MAD1 = MAD_P \tag{28}$$

If the current MB is the first row MB in the current Frame, Eq. (27) is equal to

$$MAD2 = (1-a)\times MAD_P + a\times MAD_L \tag{29}$$

If the current MB is the first column MB in the current Frame, Eq. (27) is equal to

$$MAD3 = (1-a)\times MAD_P + a\times MAD_U \tag{30}$$

3) Improved Bits Allocation Scheme at BU level
W is a simple and accurate measure of frame complexity, therefore, it can provide a mechanism to control estimation of the target bit. If the value of *W* is bigger, it also means that the current BU can better attract human eyes attention, the target bits of BU is computed by

$$T_{(i,j)} = T_{f\_r} \times W_{(i,j)} \times \frac{MAD^2_{(i,j)}}{\sum_{l=j}^{N} MAD^2_{(i,l)}} \tag{31}$$

where $T(i, j)$ denotes the total bits of the *j*th BU in the *i*th frame, T_{f_r} denotes the number of remaining bits in the current frame, $MAD(i, j)$ denotes MAD of the *j*th BU in the *i*th frame, *N* is the number of BU in the current frame.

Comparing the header bits with the texture bits, if the texture bits are less than header bits, $QPcb=QPpb+2$; on the contrary, it will get QP based on R-D model of the H.264/AVC rate control.

4 Experimental Results and Analyses

To evaluate performances of the proposed method, the test platform is JM10.1 [16]; JVT-G012 is used as a benchmark for comparison. Rate control experiments are implemented on seven QCIF video sequences with different activity and motion features. The target bit rate is set to 32, 48 and 64, *a, b, c* in Eq.(27) is set 0.8, 0.1 and 0.1. Respectively, some important test conditions are listed in Table 1. The rate control error is used to measure the accuracy of the bitrate estimation, and defined by

$$E_{rate\_predict} = \frac{\left|R_{acture} - R_{predict}\right|}{R_{predict}} \times 100\% \tag{32}$$

where R_{acture} is the bit rate of the test sequence, and $R_{predict}$ is the target bit rate.

In Table 2, compared with JVT-G012, the proposed algorithm achieves higher average PSNR and more accurate target bit rates. It is clear that JVT-G012 has a rate error range from 0.52%to1.53%, while the proposed method ranges from 0.27% to

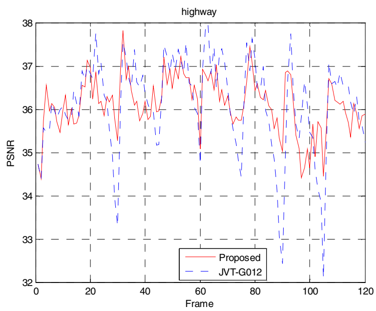
Table 1. Test conditions

Hadamard, RDO	used
Frame rate	30f/s
MV search range	16 (QCIF)
Entropy coding	CABAC
Reference frames	5
Sequence type	IPPP.........
GOP length	15
Number of frame	120

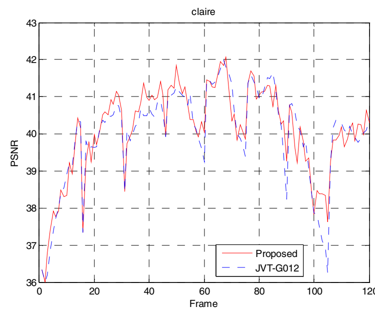
Table 2. Performance comparison between JVT-G012 and the proposed method

Target bit rate (Kbps)	Sequence	Bit rate					PSNR (dB)		
		Actual(Kbps)		Rate error (%)		Rate error gain (%)			
		G012	Pro.	G012	Pro.		G012	Pro.	Gain
32	Highway	32.33	32.13	1.03	0.41	0.62	34.39	34.72	0.33
	Claire	32.2	32.16	0.63	0.50	0.13	37.56	37.8	0.24
	Carphone	32.29	32.12	0.91	0.37	0.54	30.87	31.03	0.16
	foreman	32.24	32.13	0.75	0.41	0.34	27.89	27.97	0.08
	monitor	32.22	32.22	0.69	0.69	0.00	31.85	31.58	-0.27
	mother-daughter	32.15	32.19	0.47	0.59	-0.12	34.12	34.22	0.1
	news	32.18	32.14	0.56	0.44	0.12	29.71	29.78	0.07
48	Highway	48.63	48.29	1.31	0.60	0.71	36.11	36.11	0
	Claire	48.39	48.19	0.81	0.40	0.41	39.99	40.09	0.1
	Carphone	48.44	48.15	0.92	0.31	0.61	32.97	33.19	0.22
	foreman	48.27	48.19	0.56	0.40	0.16	30.46	30.39	-0.07
	monitor	48.38	48.29	0.79	0.60	0.19	34.16	34.26	0.1
	mother-daughter	48.5	48.23	1.04	0.48	0.56	35.87	36.03	0.16
	news	48.31	48.15	0.65	0.31	0.34	31.9	31.98	0.08
64	Highway	64.98	64.23	1.53	0.36	1.17	37.03	37.07	0.04
	Claire	64.34	64.25	0.53	0.39	0.14	41.64	41.74	0.1
	Carphone	64.55	64.21	0.86	0.33	0.53	34.43	34.71	0.28
	foreman	64.56	64.27	0.88	0.42	0.46	32.07	32.05	-0.02
	monitor	64.37	64.31	0.58	0.48	0.10	35.92	36	0.08
	mother-daughter	64.59	64.36	0.92	0.56	0.36	37.35	37.43	0.08
	news	64.33	64.17	0.52	0.27	0.25	33.5	33.75	0.25

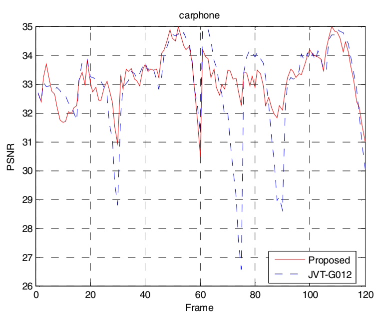
0.69%, the maximum of rate error gain is 1.17%. PSNR of the proposed method is improved up to 0.33dB compared with JVT-G012. In these sequences, the sequence that scene changes intense is Highway sequence, resulting in the most of the error rate decreased, smaller PSNR fluctuation, because of the frame difference in the frame layer. Second is the Carphone, the Claire is the last, another reason is the use of the human visual system in the BU level control. While in the lower bit rates, the Monitor sequence, which the texture of the around regions are more complex than the middle, each BU requires a lot of bit rates. The foreman sequence which motion is more violent, each frame requires a lot of bit rates.



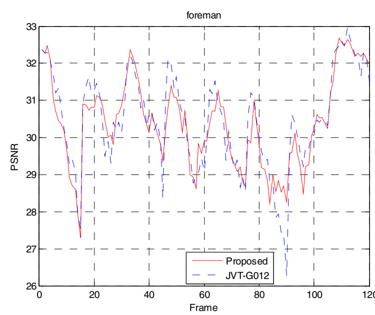
(a) Highway: Pro. =0.91, G012=1.10



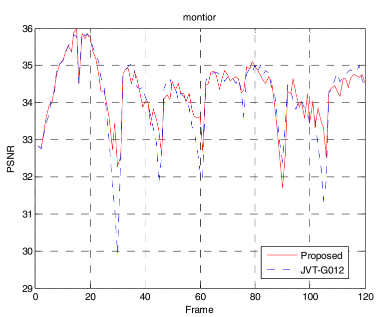
(b) Claire: Pro. =1.23, G012=1.33



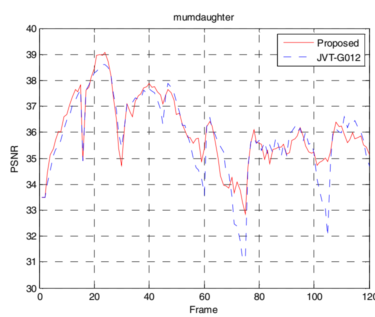
(c) car phone : Pro.=0.91, G012=1.60



(d) foreman: Pro.=1.21, G012=1.19



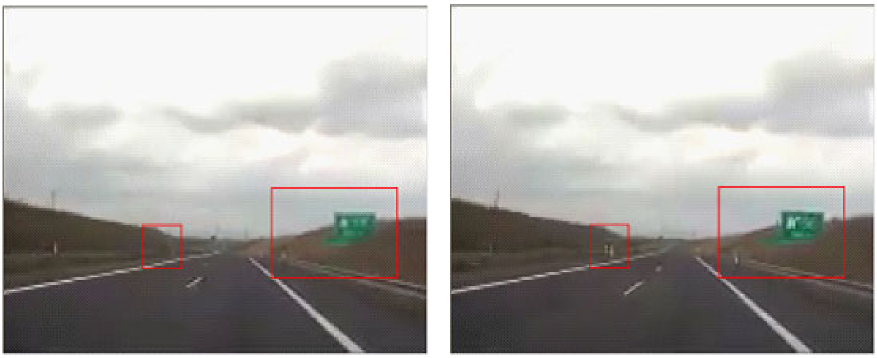
(e) Monitor: Pro. =0.81, G012=1.08



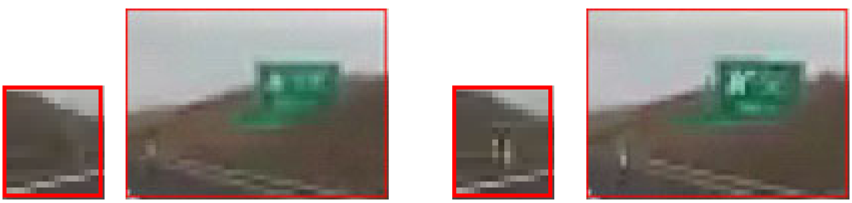
(f) mother-daughter : Pro.=1.33,G012=1.56

Fig. 1. PSNR fluctuation comparison between the two RC algorithms

In order to provide a specific comparison between JVT-G012 and the proposed rate control scheme, some experimental results are shown in Figs.1-3. Fig.1 illustrated that the proposed method can avoid drastic visual quality variation caused by scenes. Smaller PSNR fluctuation implies more stable visual quality which is highly desired in video coding. Figs.2-3 shows that the actual video sequences demonstrate that our algorithm suits real applications. From visual comparison, our proposed algorithm can select a more precise method to provide an acceptable visual image quality. It is clear that our proposed algorithm can achieve a higher PSNR than the existing algorithms and reduces the bit rate mismatch ratio.



(a) the reconstructed image of G012 (b) the reconstructed image of proposed

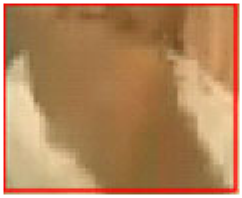


(a) the local image of G012 (b) the local image of proposed

Fig. 2. Subjective visual comparison of two RC algorithms, in the 87th frame of "Highway"

(a) the reconstructed image of G012 (b) the reconstructed image of proposed

Fig. 3. Subjective visual comparison of two RC algorithms at the 59th frame of "mother-daughter"



(a) the local image of G012

(b) the local image of proposed

Fig. 3. (*Continued*)

5 Conclusion

This paper has presented an efficient rate control algorithm based on human visual system for a H.264/AVC video coding. By analyzing the model of JVT-G012 rate control, it is clear that scene's characteristics and its bit allocation is more reasonable so that it can maintain a video stream with a smoother PSNR variation which is highly desirable in real-time video coding and transmission. Experimental results show that compared with JVT-G012 in the low bit rate, the proposed algorithm achieves higher average PSNR and more accurate rate control.

In future work, the proposed algorithm will be extended to rate control in stereo video communication, such as mobile 3DTV. In stereo video algorithm, the bit rate budget is first allocated to each pair of stereo image frame adaptively updated according to bandwidth and buffer status, combined with human visual system of stereo image frame. Then it should also allocate bits among views according to frame complexity to keep the quality of each view. It may be helpful to get a more consistent visual quality when scene switching occurs in stereo video.

References

1. MPEG-2 Test Model 5, Doc. ISO/IEC JTC1/SC29 WG11/93-400 (April 1993)
2. Corbera, J.R., Lei, S.: Rate control in DCT video coding for low delay communication. IEEE Transactions on Circuits and Systems for Video Technology 9(1), 172–185 (1999)
3. Lee, H.J., Chiang, T.H., Zhang, Y.Q.: Scalable rate control for MPEG-4 video. IEEE Transactions on Circuits and Systems for Video Technology 10(6), 878–894 (2000)
4. Li, Z.G., Pan, F., Lim, K.P., et al.: Adaptive basic unit layer rate control for JVT. In: The 7th JVT Meeting, JVT-G012-rl, Thailand (2003)
5. Lee, G.G., Lin, H.Y., Wang, M.J.: Rate control algorithm based on intra-picture complexity for H.264/AVC. IET Image Process. 3(1), 26–39 (2009)
6. Sun, Y., Zhou, Y., Feng, Z., He, Z., Sun, S.: Incremental rate control for H.264/AVC video compression. IET Image Processing 3(5), 286–298 (2009)
7. Xie, Z., Bao, Z., Xu, C., Zhang, G.: Optimal bit allocation and efficient rate control for H.264/AVC based on general rate-distortion model and enhanced coding complexity measure. IET Image Processing 4(3), 172–183 (2010)
8. Tiago, A.F., Ricardo, L.Q.: Complexity-scalable H.264/AVC in an IPP-based video encoder. In: IEEE International Conference on Image Processing (2010)

9. SanzRodríguez, S., del Ama-Esteban, O., de Frutos-López, M., de María, F.D.: Cauchy-density-based basic unit layer rate controller for H.264/AVC. IEEE Transactions on Circuits and Systems for Video Technology 20(8), 1139–1143 (2010)
10. Chen, F.C., Hsu, Y.P.: An adaptive content based H.264/AVC rate control in low bit rate video. International Journal of Electronics and Communications 65(6), 516–522 (2011)
11. Zhao, D.D., Zhou, Y.J., Wang, D.Y, Mao, J.F.: Effective macro block layer rate control algorithm for H.264/AVC. Computers and Electrical Engineering (2011)
12. Chen, Z.Z., Guillemot, C.: Perceptually-friendly H.264AVC video coding based on foveated just-noticeable-distortion model. IEEE Transactions on Circuits and Systems for Video Technology 20(6), 806–819 (2010)
13. Tang, C.W.: Spatiotemporal visual considerations for video coding. IEEE Transactions on Multimedia 9(2), 231–238 (2007)
14. Liu, Y., Li, Z.G., Soh, Y.C.: Region-of-interest based resource allocation for conversational video communication of H.264/AVC. IEEE Transactions on Circuits and Systems for Video Technology 18(1), 134–139 (2008)
15. Zheng, Q.Y., Yu, M., Peng, Z.J., Shao, F., Li, F.C., Jiang, G.Y.: Human Visual System Based Rate Control Algorithm for H.264/AVC. Journal of Optoelectronics · Laser 22(3) (2011)
16. JM Reference Software Version 10.1., `http://iphome.hhi.de/suehring/tml/download/`