Journal of Theoretical and Applied Information Technology
31st March 2013. Vol. 49 No.3

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

A NEW RATE CONTROL ALGORITHM FOR VIDEO CODING BASED ON JUST NOTICEABLE DISTORTION

[1]YI LIAO, [1,2]MEI YU, [1]XIAODONG WANG, [1,2]GANGYI JIANG, [1]WEIYUE LIU
[1] Faculty of Information Science and Engineering, Ningbo University, Ningbo, China
[2]National Key Lab of Software New Technology, Nanjing University, Nanjing, China
E-mail: [1]liaoyi518@163.com , [2]jianggangyi@126.com

ABSTRACT

To improve the performance of rate control algorithm for H.264/AVC and keep a good quality for video, a new rate control algorithm is proposed based on the just noticeable distortion (JND). Firstly, the motion characteristics of inter-frames were used to measure the frame complexity as to allocate the target bits rationally in frame level. Secondly, just noticeable distortion for measuring visual sensitivity of macro-block was introduced, and the bits allocation of macro-block is based on the JND value. The experimental results show that the proposed algorithm can control the bit-rate more accurately and improve the subjective quality of the visual sensitive areas without any rate-distortion performance decline compared with JVT-G012 algorithm.

Keywords: *Video coding, Rate Control (RC), Just Noticeable Distortion (JND), Frame complexity*

1. INTRODUCTION

Rate control (RC) plays an essential role in video communication which can adjust the video quality to satisfy the channel bandwidth constraint and the buffer constraint. So far, several rate control algorithms have been proposed, such as TM5 for MPEG-2, TMN8 for H.263, VM8 for MPEG-4 and JVT-G012 [1] for H.264/AVC. JVT-G012 is the most classical rate control algorithm in which the leaky bucket model is proposed to keep the bit stream stable, the linear prediction model is used to predict the complexity of each frame and each basic unit (BU) and the quadratic rate-distortion model is used to compute the quantization parameter (QP).

However, the algorithm of JVT-G012 still exist some shortcomings. Firstly, in the frame level, it takes the mean absolute difference (MAD) as the complexity which will leads poor visual quality in scene of vigorous movement for video. Secondly, the complexity of BU is used to guide bit allocation in macro-block level which is not consistent with the visual perception. To solve the former problem, Yan et.al [2] established a linear relationship between the frame complexity and the generated bite rate at the referential base QP, and controlled I-frame and P-frame jointly to allocate bit rate. Jiang et.al [3] analyzed the relationship between the video scene complexity and the gradient value, and then set a RCQ model to estimate the scene complexity. To solve the latter one, Hu et.al [4] used the gray projection method to extract the moving region and divided a frame into multiple regions based on the human visual system. Wang et.al [5] optimized the rate distortion model by the structural similarity index in place of mean squared error which can be integrated with human perception.

In this paper, we propose a rate control algorithm based on the just noticeable distortion (JND). In the frame level, the total number of motion macro-blocks and the summation of motion vector (MV) of motion macro-blocks are used to measure the motion complexity of each frame. In the macro-block level, JND values for macro-blocks are calculated to allocate bitrates according to human visual characteristic. The experimental results show that the proposed rate control method can accurately control the bitrates and provide a good visual quality for video.

The rest of the paper is organized as follows. In Section 2, related works is briefly described. In Section 3, we introduce the spatio-temporal JND model in detail. A rate control algorithm based on JND is proposed in Section 4. Then, the experimental results are analyzed in Section 5. Finally, the conclusions are given in Section 6.

Journal of Theoretical and Applied Information Technology
31st March 2013. Vol. 49 No.3

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

2. RELATED WORKS

JND simulates luminance contrast and spatial-temporal masking in human visual system (HVS), the human eyes cannot sense any changes below the JND threshold around a pixel. JND was always used to improve the perceptual quality and encoding efficiency [6, 7]. In [8], a perceptual video encoder architecture based on JND model under energy resource constraint was bailed, the JND was extended to temporal domain so as to determine perceptual cue in unit of MB and provided the guideline of resource allocation in MBs. In [9], a foveate JND (FJND) model which incorporated both spatial and temporal masking and foveation characteristics was first established, and the proposed FJND model was then used for macro-block (MB) quantization adjustment in video coding. In [10], a stereoscopic JND model was established for block-adaptive residue pre-processing which can improve the stereoscopic video coding efficiency.

Obviously, it is imperative for us to establish an appropriate JND model according to the human perceptual, then we can allocate more bits for the regions of small JND thresholds where are sensitive to any distortion for human visual.

3. JUST NOTICEABLE DISTORTION

We first establish the spatial JND model and the temporal JND model on pixel domain, and then combined them into a spatio-temporal JND model.

3.1 Spatial JND

Chou et.al [11] simulated the impact on JND by luminance contrast and spatial masking, and then obtained a spatial JND

$$SJND(x,y)=\max\{f_1(bg(x,y),mg(x,y)),\ f_2(bg(x,y))\} \quad (1)$$

where $bg(x,y)$ and $mg(x,y)$ denote the average background luminance and the maximum weighted average of luminance differences around the pixel (x,y), respectively. $f_1(bg(x,y),mg(x,y))$ denotes the spatial masking and $f_2(bg(x,y))$ denotes the luminance contrast. $f_1(bg(x,y),mg(x,y))$ is computed by

$$f_1(bg(x,y),mg(x,y))=mg(x,y)\times\alpha(bg(x,y))+\beta(bg(x,y)) \quad (2)$$

where $mg(x, y)$ is calculated as

$$mg(x,y)=\max_{k=1,2,3,4}\{|grad_k(x,y)|\} \quad \text{(3-a)}$$

$$grad_k(x,y)=\frac{1}{16}\sum_{i=1}^{5}\sum_{j=1}^{5}p(x-3+i,y-3+j)\times g_k(i,j) \quad \text{(3-b)}$$

where $g_k(i,j)$ denotes the operators for calculating the weighted average of luminance changes in four directions.

The $\alpha(bg(x,y))$ and $\beta(bg(x,y))$ in formula (2) are both determined by background luminance

$$\alpha(bg(x,y))=bg(x,y)\times 0.0001+0.115 \quad (4)$$

$$\beta(bg(x,y))=\mu-bg(x,y)\times 0.01 \quad (5)$$

The $f_2(bg(x,y))$ in formula (1) is calculated as

$$f_2(bg(x,y))=\begin{cases}17(1-\sqrt{\dfrac{bg(x,y)}{127}})+3, & bg(x,y)\le 127\\ \dfrac{3}{128}(bg(x,y)-127)+3, & bg(x,y)>127\end{cases} \quad (6)$$

3.2 Temporal JND

The temporal masking is determined by the frame difference of two adjacent frames and the background luminance [12]. Let $TJND(x,y,j)$ be the temporal JND

$$TJND(x,y,j)=\begin{cases}\max(\tau,4\cdot\exp(\dfrac{-0.15}{2\pi}(\Delta(x,y,j)+255))\\ +\tau), \quad \Delta(x,y,j)\le 0\\ \max(\tau,1.6\cdot\exp(\dfrac{-0.15}{2\pi}(255-\Delta(x,y,j)))\\ +\tau), \quad \Delta(x,y,j)>0\end{cases} \quad (7)$$

where τ is a constant and its value is 0.8, and $\Delta(x,y,j)$ is calculated by

$$\Delta(x,y,j)=\frac{I(x,y,j)-I(x,y,j-1)+\bar{I}(j)-\bar{I}(j-1)}{2} \quad (8)$$

where $I(x,y,j)$ denotes the luminance value of pixel at (x,y) in the j-th frame, $\bar{I}(j)$ denotes the average luminance value in the j-th frame.

(a)Color image *(b)JND image*

Figure 1. Color image and its corresponding JND image for Breakdancers

3.3 Spatio-temporal JND model

After the above discussion about spatial JND model and temporal JND model, we define the spatio-temporal JND model as

$$JND(x,y,j)=SJND(x,y)\times TJND(x,y,j) \quad (9)$$

Journal of Theoretical and Applied Information Technology
31st March 2013. Vol. 49 No.3

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

The JND map is shown in Figure 1. It can be seen from Figure 1(b) that the value of JND is smaller when the area is darker and human eyes are more sensitive to the distortion that happens on this area.

4. THE PROPOSED RATE CONTROL ALGORITHM

By introducing motion complexity into frame level and spatio-temporal JND model described in section 3, we proposed a rate control algorithm which is consistent with visual perception. The algorithm is performed on three levels, namely group of picture (GOP) level, frame level, macro-block level. In the GOP level, the initial quantization parameter and the target bits are calculated for each GOP. In the frame level, bits allocation is decided by the predicted complexity of the frame. In the macro-block level, the JND value is taken as a sensitive factor to guide the bits allocation.

4.1 GOP Level Rate Allocation

In the GOP level, the total number of bits for each GOP and the QP of the key frame are calculated. Before encoding the *i*-th GOP, the bits allocated for the GOP are computed by

$$T_r(i,0)=\frac{u(i,1)}{F_r}*N_{gop}-(\frac{B_s}{8}-B_c(i-1,N_{gop})) \qquad (10)$$

where $u(i,1)$ denotes the bandwidth, F_r denotes the frame rate of the video, N_{gop} denotes the total number of frames in each GOP, B_s is the capacity of the buffer and $B_c(i\text{-}1,N_{gop})$ is the buffer fullness after coding the (*i*-1)-th GOP. The buffer occupancy should be kept at $B_s/8$ after coding each GOP as to ensure that all GOPs accordance with each other. In the case of constant bandwidth, T_r (i,j) is updated frame by frame as

$$T_r(i,j)=T_r(i,j-1)-A(i,j-1) \qquad (11)$$

where $A(i,j)$ denotes the actual bits of the (*j*-1)-th frame.

4.2 Frame Level Rate Allocation

The rate control algorithm of JVT-G012 allocates the bits for a frame based on the buffer constraints, complexity of each frame and the remained bits, but the scene of vigorous movement is not considered which may lead poor visual quality. Therefore, the total number of motion macro-blocks and the summation of *MV* of motion macro-blocks are used to measure the motion complexity of each frame in this paper.

The absolute value of motion vector for a macro-block is computed by

$$MV=\sqrt{MV_h^2+MV_v^2} \qquad (12)$$

where MV_h and MV_v denote the motion vector value in the horizontal and vertical direction, respectively. When the value of *MV* of a macro-block is larger than 5, it is defined as a motion macro-block.

The relative motion complexity is calculated by three steps as follows.

Firstly, the total number of motion macro-blocks in the *j*-th frame is calculated and expressed by $C(j)$ which denotes the size of motion region. Let $Area(j)$ be the size of motion region in *j*-th frame relative to encoded frames

$$Area(j)=\frac{C(j)}{C_{avg}(j-1)} \qquad (13)$$

where $C_{avg}(j\text{-}1)$ denotes the total number of motion macro-blocks of the encoded frames, and $C(j)$ is predicted by the previous frame.

Secondly, the summation of the MVs of all the macro-blocks in the *j*-th frame is calculated and expressed by $S(j)$ which denotes the motion intensity of the current frame. Let $Amp(j)$ be the motion intensity in *j*-th frame relative to encoded frames

$$Amp(j)=\frac{S(j)}{S_{avg}(j-1)} \qquad (14)$$

where $S_{avg}(j\text{-}1)$ denotes the summation of the MVs of the motion macro-blocks of the encoded frames, and $S(j)$ is predicted by the previous frame.

Lastly, let $M(j)$ be the motion complexity of the *j*-th frame relative to encoded frames

$$M(j)=\lambda\times Area(j)+(1-\lambda)\times Amp(j) \qquad (15)$$

where λ is a constant and its value is 0.5. The larger $M(j)$ is, the more bits should be allocated to the *j*-th frame.

In order to be rationally used for bits allocation, the $M(j)$ is nonlinear quantified by

$$H(j)=\begin{cases}0.5 & M(j)\le 0.7\\ 0.8M(j) & 0.7<M(j)\le 1\\ 1.1M(j) & 1<M(j)\le 1.5\\ 1.7 & M(j)>1.5\end{cases} \qquad (16)$$

By considering the remained bits, the target bits for the *j*-th frame in the *i*-th GOP are calculated by

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

$$f(i,j)=H(j)\times\frac{W_pT_r(i,j)}{W_pN_{pr}(j-1)+W_bN_{br}(j-1)} \quad (17)$$

where W_p and W_b are the complexity weight of P- and B-frames, respectively, and $N_{pr}(j\text{-}1)$ and $N_{br}(j\text{-}1)$ are the numbers of the remaining P-frames and B-frames in the *i*-th GOP, respectively[1].

4.3 Macro-block Level Rate Allocation

In the video coding, we should improve the subjective quality for the region where human is sensitive to the distortion. Consequently, the smaller JND value of a macro-block is, the more bits should be allocated.

The macro-block JND is calculated by

$$JND(j,n)=\frac{1}{256}\sum_{y=0}^{15}\sum_{x=0}^{15}JND_{(j,n)}(x,y) \quad (18)$$

where $JND_{(j,n)}(x,y)$ denotes the JND value at pixel (x,y) in the *n*-th macro-block in the *j*-th frame.

The visual sensitive coefficient of the *n*-th macro-block in the *j*-th frame is expressed by $W(j,n)$

$$W(j,n)=\frac{\frac{1}{N}\sum_{k=0}^{N-1}JND(j,k)}{JND(j,n)} \quad (19)$$

where N denotes the total number of macro-blocks in a frame.

The target bits for the *n*-th macro-block in the *j*-th frame is calculated by

$$T(j,n)=T_r(j)\times W(j,n)\times\frac{MAD^2(j,n)}{\sum_{t=n}^{N}MAD^2(j,t)} \quad (20)$$

where $T_r(j)$ denotes the remaining bits in the *j*-th frame, $MAD(j,n)$ denotes the mean absolute difference of the *n*-th macro-block in the *j*-th frame [1].

5. EXPERIMENTAL RESULTS

To evaluate the performances of the proposed method, JVT-G012 is used as a benchmark for comparison and the test platform is JM18.2 [13]. Several experiments were performed with video sequences of "Ballet", "Breakdancers" and "BookArrival" with the size of 1024×768. The three test sequences are shown in Figure 2.

Some important test conditions are listed in Table 1.

Table 1. Experiment conditions

RD optimization	used
Frame rate	30f/s
Entropy coding method	CABAC
Sequence type	IPPP…
GOP length	8
Frames to be encoded	81

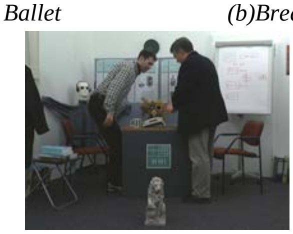
The rate control error is used to measure the accuracy of the bitrates controlling and calculated by

$$E=\left|\frac{R_{acture}-R_{target}}{R_{target}}\right|\times 100\% \quad (21)$$

where R_{acture} is the bitrates generated by the test sequence, and R_{target} is the target bitrates which is generated by coding the test sequences with fixed QP, and the fixed QP is set to 22, 27, 32 and 37, respectively.

(a) Ballet

(b)Breakdancers



(c) BookArrival

Figure 2. Test sequences

Table 2 summarizes the control accuracy and the average PSNR. It can be seem from the table that the accuracy of the G012 is with 0.04% to 0.87% while that of the proposed method is with 0.11% to 0.76%. The proposed method can achieve good control accuracy because the complexity prediction is adopted into the frame level and the frames of vigorous movement are allocated for more bits.

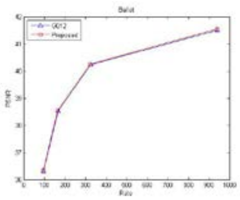
The rate-distortion (RD) performance comparison results between the proposed method and the G012 method are shown in Figure 3. The two methods are of the similarly RD performance.

PSNR cannot accurately reflect the subjective feeling of the human visual, so we compare the subjective image quality which shown in the Figure 4-6. In the Figure 4, it is clear that the hand of the ballet dancer is more distinct in the proposed method. Figure 5 shows that the edge of the arm is

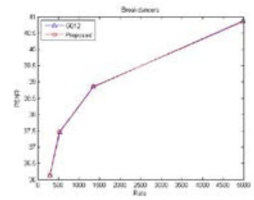
Journal of Theoretical and Applied Information Technology
31st March 2013. Vol. 49 No.3

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

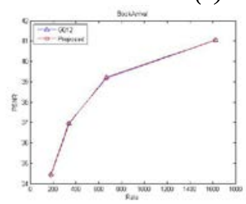
more clearly in the proposed method. Figure 6 shows that the proposed method can achieve better texture of the fingers of the man compared with the G012. Therefore, the proposed algorithm can provide more comfortable visual image quality for visual sensitive areas.



(a) Ballet



(b) Breakdancers



(c) BookArrival

Figure 3. The RD performance comparisons between the two methods

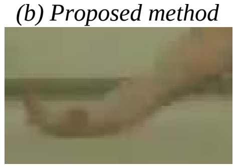
Table 2. Simulation results of the proposed method and the JVT-G012 method

Sequence	Target bitrates (kbps)	Actual bitrates (kbps)		Inaccuracy(%)			PSNR(dB)		
		G012	Proposed	G012	Proposed	Gain	G012	Proposed	Gain
Ballet	936.29	940.20	940.17	0.42	0.41	0.01	41.50	41.55	0.05
	323.44	325.57	323.94	0.66	0.15	0.51	40.23	40.26	0.03
	164.82	165.90	165.24	0.66	0.25	0.41	38.51	38.56	0.05
	93.80	94.60	94.51	0.85	0.76	0.09	36.31	36.34	0.03
Breakdancers	4985.04	4982.89	4984.34	0.04	0.01	0.03	40.88	40.84	-0.04
	1350.83	1352.98	1351.68	0.16	0.06	0.10	38.85	38.86	0.01
	525.43	529.34	527.29	0.74	0.35	0.39	37.46	37.47	0.01
	293.16	295.71	294.11	0.87	0.32	0.55	36.12	36.14	0.02
BookArrival	1621.96	1624.47	1623.06	0.15	0.07	0.08	41.05	41.06	0.01
	666.71	669.30	667.03	0.39	0.05	0.34	39.22	39.18	-0.04
	338.27	339.66	338.56	0.41	0.09	0.32	36.97	36.97	0.00
	181.79	182.72	182.01	0.51	0.12	0.39	34.45	34.43	-0.02

(a) G012 method

(b) Proposed method

(c) Local area in Figure 4 (a)



(d) Local area in Figure 4 (b)

Figure 4. Subjective visual comparison of the two methods for Ballet

(a) G012 method

(b) Proposed method

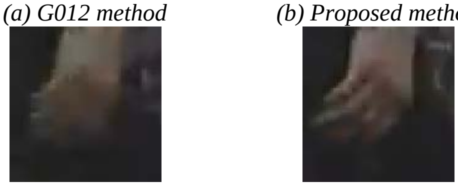
(c) Local area in Figure 5 (a)

(d) Local area in Figure 5 (b)

Figure 5. Subjective visual comparison of the two methods for Breakdancers

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

(a) G012 method *(b) Proposed method*



(c) Local area in Figure 6 (a) *(d) Local area in Figure 6 (b)*

Figure 6. Subjective visual comparison of the two methods for BookArrival

6. CONCLUSIONS

In this paper, a rate control algorithm for video coding based on the just noticeable distortion is proposed. The proposed algorithm is performed on three levels, namely GOP level, frame level and macro-block level. In the GOP level, the initial quantization parameter and the target bits are calculated for each GOP. In the frame level, bits allocation is decided by the predicted complexity of the frame. In the macro-block level, the JND value is taken as a sensitive factor to guide the bits allocation. Experimental results show that the proposed rate control method can accurately control the bitrates and provide a good visual effect. In future probe, more efforts will be focused on consideration of PSNR fluctuation to improve the fluency of the video.

ACKNOWLEDGEMENTS

This work is supported by Natural Science Foundation of China (61271270, 61271021, 61111140392), Natural Science Foundation of Zhejiang Province (Y1101240), and The projects of Ningbo Government (2011A610197, 2011B81002).

REFRENCES:

[1] ISO/IEC JTC1/SC29/WG11 and ITU-T SG16/Q.6, "Adaptive Basic Unit Layer Rate Control for JVT", *Doc. JVT-G012r1*, Pattaya, Thailand, March 7-15, 2003, pp.23-27.

[2] B. Yan and K. Sun, "Joint Complexity Estimation of I-frame and P-frame for H.264/AVC Rate Control", *IEEE Transactions on Circuits and Systems for Video Technology*, Vol.22, No.5, May 2012, pp.790-798.

[3] Z. Jiang, Q. Dai and Y. Ai, "A Scene Complexity Adaptive Intra-Frame Rate Control for Video Coding", *2010 International Conference on Multimedia Communications*, Leshan, China, August 7-8, 2010, pp.108-111.

[4] H. Hu, B. Li, W. Lin, W. Li and M. Sun, "Region-Based Rate Control for H.264/AVC for Low Bit-Rate Applications", *IEEE Transactions on Circuits and Systems for Video Technology*, Vol.22, No.11, November 2012, pp.1564-1576.

[5] S. Wang, Z. Wang, S. Ma and W. Gao, "SSIM-Motivated Rate-Distortion Optimization for Video Coding", *IEEE Transactions on Circuits and Systems for Video Technology*, Vol.22, No.4, April 2012, pp.516-529.

[6] Z. Peng and Z. Chen, "Perceptual Quality Measure Using JND Model of The Human Visual System", *2011 International Conference on Electric Information and Control Engineering (ICEICE)*, April 15-17, 2011, pp.2454-2457.

[7] M. Naccari and F. Pereira, "Advanced H.264/AVC-Based Perceptual Video Coding: Architecture, Tools, and Assessment", *IEEE Transactions on Circuits and Systems for Video Technology*, Vol.21, No.6, June 2011, pp.766-782.

[8] W. Ji, M. Chen, X. Ge, P. Li and Y. Chen, "A Perceptual Macro-block Layer Power Control for Energy Scalable Video Encoder Based on Just Noticeable Distortion Principle", *Journal of Network and Computer Applications*, Vol.34, No.5, September 2011, pp.1489-1497.

[9] Z. Chen and C. Guillemot, "Perceptually-Friendly H.264/AVC Video Coding Based on Foveated Just-Noticeable-Distortion Model", *IEEE Transactions on Circuits and Systems for Video Technology*, Vol.20, No.6, June 2010, pp. 806-819.

[10] L. Zhang, Q. Peng, Q. Wang and X. Wu, "Stereoscopic Perceptual Video Coding Based on Just-Noticeable-Distortion Profile", *IEEE Transactions on Broadcasting*, Vol.57, No.2, June 2011, pp. 572-581.

[11] C. Chou and Y. Li, "A Perceptually Tuned Subband Image Coder Based on The Measure of Just-Noticeable-Distortion Profile", *IEEE Transactions on Circuits and Systems for Video Technology*, Vol.5, No.6, December 1995, pp. 467-476.

[12] Z. Wei and K. Ngan, "A Temporal Just-Noticeble Distortion Profile for Video in DCT Domain", *15th IEEE International Conference on Image Processing (ICIP)*, Hong Kong, October 12-15, 2008, pp.1336-1339.

[13] JM Reference Software Version 18.2 *http://iphome.hhi.de/suehring/tml/download/old_jm*.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195