A PERCEPTUAL BIT ALLOCATION SCHEME FOR H.264

Hongtao Yu[1], *Feng Pan*[2], *Zhiping Lin*[1], *Yin Sun*[3]

[1]Nanyang Technological University, School of EEE, S2, Nanyang Avenue, Singapore 639798
[2]Institute for Infocomm Research, Media Division, 21 Heng Mui Keng Terrace, Singapore 119613
[3]National University of Singapore, Department of ECE, 10 Kent Ridge Crescent, Singapore 119260

ABSTRACT

This paper aims at improving perceptual quality of encoded video sequences, and proposes a new perceptual bit allocation scheme for H.264. Firstly, a new motion complexity measure is defined to represent the amount of motion contents between two consecutive frames, and is used to estimate the target bit at frame level. Secondly, a segmentation method exploiting the perceptual characteristics of the video contents is presented, and is used to adaptively update the Lagrangian multiplier in the coding mode selection at macroblock level. Thirdly, based on the motion complexity and Lagrangian multiplier updating, a rate control scheme for H.264 is proposed. Experimental results show that our scheme can effectively improve the perceptual video quality as compared with the H.264 adopted rate control algorithm [1]. Moreover, our scheme also achieves an average peak signal-to-noise ratio gain of up to 0.138dB for the test sequences.

1. INTRODUCTION

H.264 [2] is the newest international video coding standard developed by the Joint Video Team (JVT), which consists of experts from VCEG and MPEG. It has achieved a significant improvement in coding efficiency compared to all the existing standards. As in other video coding standards, rate control is a necessary part of the encoder in H.264.

In the H.264 adopted rate control algorithm [1], Li *et al.* have used a linear model to predict the mean absolute difference (MAD) of current basic unit in the current frame by using a co-sited basic unit in the previous frame. It is noted that in this approach, the target bit is estimated solely based on the buffer fullness, regardless of the frame's content. This may lead to drastic drops in peak signal-to-noise ratio (PSNR), especially in the case of high motion scenes or scene changes. To improve the video quality at scene changes, Jiang *et al.* [3] have introduced MAD ratio as a measure of motion complexity. In their approach, bit budget is allocated to frames according to their MAD ratio. However, MAD ratio is not a good way of representing the motion contents, as it can only represent the similarity between the current frame and its reference frame.

In the above existing H.264 rate control algorithms, MAD is selected to estimate quantization parameter (QP) and therefore to decide coding modes. However, it is well understood that the minimum MAD does not translate into minimum perceptual distortion [4]. A better approach is to use the perceptual characteristics of the video contents. In [5], a video coding approach that performs adaptive rate-distortion optimization (RDO) guided by perceptual hints is proposed. The key idea is to adaptively adjust the Lagrange multipliers of the RDO coder control module based on visual attention analysis. In [6], a Lagrangian optimized rate control algorithm has been proposed for the H.264 video encoder. The algorithm controls the bit rate by adjusting the Lagrangian multiplier adaptively for every picture and specifying the QP for every macroblock (MB). The success of the above perceptual algorithms depends largely on the good estimation of visual features and thus a more accurate visual important map needs to be established.

In this paper, we focus on accurately estimating target bit at scene changes and high motions, and updating the Lagrangian multiplier according to the perceptual characteristics of the video contents. We propose a rate control scheme including frame level and MB level. At frame level, we estimate the target bits by using the motion complexity measure which represents the amount of motion contents between two consecutive frames. At MB level, we allocate bits perceptually by updating the Lagrangian multiplier according to the MB patterns.

The rest of this paper is organized as follows. The next section discusses the relationship between H.264 bit allocation and motion complexity, and defines a new motion complexity measure. Section 3 describes mode decision and RDO in H.264, and proposes a perceptual mode decision scheme by updating Lagrangian multiplier adaptively. Section 4 proposes our rate control scheme for H.264. Section 5 shows our experimental results. This paper concludes with Section 6.

2. BIT ALLOCATION AND MOTION COMPLEXITY

The complexity of motion contents refers to the moving picture contents of two consecutive frames in a video sequence. Generally bit allocation obeys the following rule: more bits are allocated to high motion frames, while fewer bits are allocated to low motion frames.

Figure 1 shows the bit allocation of the sequence "Foreman" in the case of disabling rate control. From this figure, we can find that the bit allocation is in accordance with the complexity of frame's motion contents. For example, from the 253rd frame to the 254th frame, high motion happens at the man's eyes and hand. Bits which are allocated to the 254th frame is 39% more than those are allocated to the 253rd frame. This shows that the allocated bits can be a good measure to represent the complexity of motion contents.

0-7803-9332-5/05/$20.00 ©2005 IEEE

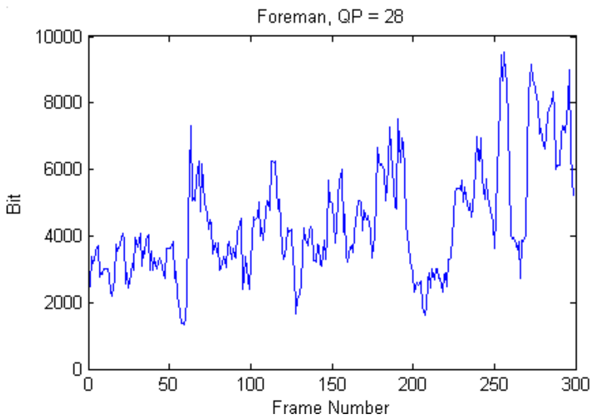

Figure 1 Bit Allocation of Sequence "Foreman"

However, when we enable rate control in H.264, we cannot obtain the actual bits before entropy encoding. What we can do is only to estimate the target bits allocated to the current frame. Usually there is a difference between the estimated bits and finally allocated bits, so we must estimate the target bits accurately. The estimated target bits should represent the complexity of the frame's motion contents. Since there is a temporal correlation between consecutive frames, we can use the previous coded frames to measure current frame's complexity of motion contents. Here we propose a new measure, namely motion complexity, to represent the complexity of a frame's motion contents. Motion complexity depends on the bits that are allocated to the encoded frames. Let $B_{P,i}$ be the predicted bits at frame i, B_j (j = 1, 2,…, i) be the actual allocated bits to the previously encoded frames, then motion complexity at frame i is:

$$C_i = \frac{B_{P,i}}{\frac{1}{i-1}\sum_{j=1}^{i-1} B_j} \tag{1}$$

where $B_{P,i} = \alpha_i B_{i-1}$ is the linear prediction of the actual previous frame's bits. In real implementation, α_i should be updated after bits have been actually finished allocation. It depends upon current frame's bits and previous frame's bits. That is:

$$\alpha_{i+1} = \frac{B_i}{B_{i-1}} \times \alpha_i \tag{2}$$

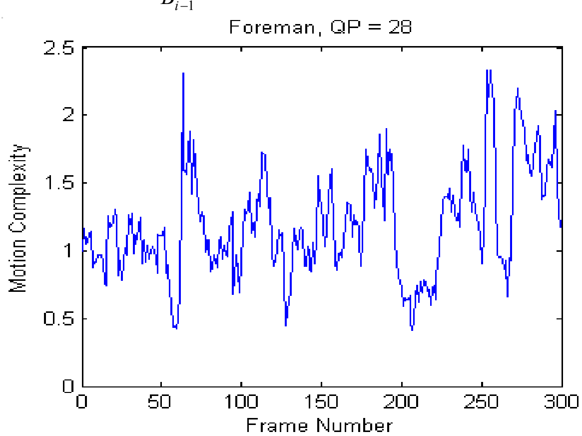

Figure 2 Motion Complexity of Sequence "Foreman"

We illustrate the motion complexity curve in Figure 2. Comparing Figure 2 with Figure 1, we can find that the shape of motion complexity curve is almost the same as the shape of bit curve. This means that motion complexity is also a good measure for the complexity of motion contents. From Equation (1), we can see that the computation of motion complexity is simple. It is only related to the actual bits of previous frames. By employing motion complexity, we can make clear the complexity of current frame's motion contents. This will help us accurately estimate the target bits in rate control.

3. PERCEPTUAL BIT ALLOCATION

3.1. Mode Decision and Rate-Distortion in H.264

One of the novel features of H.264 video coding is the use of 7 different MB coding modes so that the temporal and spatial details in an MB are best presented. To select the best mode, RDO is employed such that for each MB, all the MB modes are tried and the one that leads to the least rate-distortion (RD) cost is selected. This is to achieve the best trade-off of the rate and distortion performance.

In the JVT reference model software, a Lagrangian multiplier method is used to achieve RDO [7]. The Lagrangian multiplier method is based on converting a constraint optimization problem to an unconstraint one. The constraint optimization problem is to minimize the distortion D at the constraint that the rate R should be less than R_C, which can be expressed as follows:

$$\min D: \quad R < R_C \tag{3}$$

Using the Lagrangian multiplier method, the above problem is converted to minimize the RD cost J, with λ being the Lagrangian multiplier:

$$\min J: \quad J = D + \lambda \times R \tag{4}$$

R is generated from every MB i's rate R_i, and D is generated from every MB i's distortion D_i. Assuming that the rate and distortion of each MB are only dependent on the choice of the encoding parameters as described above, the optimization of (4) is simplified to minimizing the cost of each MB separately:

$$\min J_i: \quad J_i = D_i + \lambda_i \times R_i \tag{5}$$

The mode that has the minimum MB cost is selected as the optimum coding mode for this MB.

3.2. Perceptual Mode Decision and Bit Allocation

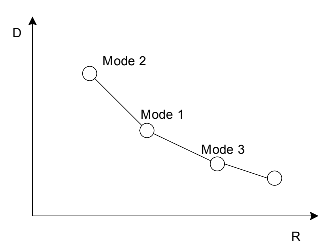

Figure 3 Mode Selection

From the above description of mode decision and RDO, it is recognized that the selection of Lagrangian multiplier will influence the selection of coding modes. Figure 3 illustrates the effect of updating the Lagrangian multiplier. Mode 1 is the case where the Lagrangian multiplier is calculated according to (5). If we increase the Lagrangian multiplier, the coding mode will transfer to mode 2 which means fewer bits are allocated. If we decrease the Lagrangian multiplier, the coding mode will transfer to mode 3 which means more bits are allocated.

One of the important characteristics of human vision system (HVS) is called edge masking that HVS is less sensitive to errors along a prominent edge in an image. In other words, a dominant edge pattern will obscure the perception of other lower contrast variations in this block [8]. Due to this observation, more bits may be allocated to the dominant edge pattern, while fewer bits may be allocated to the texture pattern. This observation can also be illustrated in Figure 3. If the MB is texture pattern, fewer bits should be allocated to it and its coding mode is more likely mode 2. On the other hand, if the MB is edge pattern, more bits should be allocated to it and its coding mode is more likely mode 3. Therefore, we can connect the bits, the Lagrangian multiplier and the MB pattern together to allocate bits perceptually. That is, we update the Lagrangian multiplier according to the MB pattern, and this will change the mode selection perceptually, and correspondingly the allocated bits.

The steps of updating Lagrangian multiplier are listed as follows.

Step 1: Calculate the Sobel operators.

$$G_x = I_{x-1,y+1} + 2I_{x,y+1} + I_{x+1,y+1} - I_{x-1,y-1} - 2I_{x,y-1} - I_{x+1,y-1} \quad (6)$$

$$G_y = I_{x+1,y-1} + 2I_{x+1,y} + I_{x+1,y+1} - I_{x-1,y-1} - 2I_{x-1,y} - I_{x-1,y+1} \quad (7)$$

Step 2: Calculate the average squared gradients.

$$G_{xx} = \sum_W G_x^2, \quad G_{yy} = \sum_W G_y^2, \quad G_{xy} = \sum_W G_x G_y \quad (8)$$

Step 3: Calculate the coherence of the squared gradient [9].

$$Coh = \frac{\sqrt{\left(G_{xx} - G_{yy}\right)^2 + 4G_{xy}^2}}{G_{xx} + G_{yy}} \quad (9)$$

Step 4: Determine the block patterns according to edge threshold T_{Edge} and texture threshold $T_{Texture}$.

$$Pattern = \begin{cases} Edge, & if\ Coh > T_{Edge} \\ Texture, & if\ T_{Texture} < Coh \le T_{Edge} \\ Background, & others \end{cases} \quad (10)$$

Step 5: Update the Lagrangian multiplier.

$$a = \sum_W a_i\ ,\ a_i = \begin{cases} 0.5, & if\ Pattern = Edge \\ 1.0, & if\ Pattern = Texture \\ 0.5, & if\ Pattern = Background \end{cases} \quad (11)$$

$$\lambda' = a\lambda \quad (12)$$

where λ' is the updated Lagrangian multiplier.

(a) Original Image (b) Importance Map

Figure 4 The 37th Frame of Sequence "Bike"

Figure 4 shows an example of segmentation result of the 37th frame of sequence "Bike". It can be seen that most of the important edges and textures can be recognized in the importance map.

4. IMPROVED RATE CONTROL SCHEME

Based on the motion complexity measure and perceptual mode decision, we propose an improved rate control scheme which includes frame-level and MB-level rate control. At frame-level, we estimate the target bit of current frame according to motion complexity. At MB-level, we update Lagrangian multiplier according to the perceptual characteristics.

At the frame-level rate control, target bit estimation includes the total number of bits of a group of pictures (GOP), the remaining bits of a GOP, and the target bits of the current frame. The total number of bits of a GOP is determined by the following equation:

$$B = N \times \frac{R_B}{R_F} \quad (13)$$

where N is the number of frames in a GOP, R_B is the bit rate, R_F is the frame rate. The remaining bits of a GOP when encoding frame i is:

$$B_{R,i} = \begin{cases} B & i = 1 \\ B_{R,i-1} - B_{i-1} & i = 2,3,\cdots,N \end{cases} \quad (14)$$

where B_i is the bits of frame i. The target bits of the ith frame is:

$$B_{T,i} = \frac{C_i \times B_{R,i} + \beta_i \times (F_s - F_i + \frac{R_B}{R_F})}{N - i + 1} \qquad i = 2,3,\cdots N \quad (15)$$

where C_i is the motion complexity, β_i ($0 \le \beta_i \le 1$) is a coefficient to control buffer fullness, F_s is the buffer size, F_i is the buffer fullness at time i.

At the MB level rate control, we estimate QP and calculate the original Lagrangian multiplier. Then we use (6) — (12) to update the Lagrangian multiplier.

5. EXPERIMENTAL RESULTS

We implement our rate control scheme in a H.264 reference model version JM6.1e and compare it with JVT-H014 [1]. In the experiments, we firstly encode the test sequences by using a fixed QP to find out the actual bits it consumes, and use it as the target bit rate in the rate control scheme. We use three CIF video sequences in total and only consider the features enabled in the Main Profile of H.264. We evaluate the experimental results with both subjective and objective assessment.

5.1. Subjective Quality Evaluation

For subjective quality evaluation, we use the double stimulus continuous quality scale (DSCQS) test method which is described in ITU-R BT.500-10 [10]. The mean opinion score (MOS) scales for observers to vote for the quality after viewing are excellent, good, fair, poor and bad. Correspondingly, the scores are 5, 4, 3, 2 and 1. Four observers were involved in the experiments. All of them have the knowledge of image processing. Table 1 lists the evaluation results. The higher the score gain is, the better the perceptual quality is achieved. From the table, we can see that the subjective score of our proposed scheme is better than that of JVT-H014. For sequences "Bike" and "Bus", our scheme earns higher scores. For sequence "Stefan", both schemes have the same scores. This shows that our scheme can achieve better perceptual quality.

Table 1 Subjective Score

Sequence	QP	Target Bit Rate (Kbps)	Score		Score Gain
			H014	Proposed	
Bike	28	961.63	4.50	4.50	0.00
	32	604.12	4.00	4.25	0.25
	36	365.95	4.00	4.25	0.25
	40	217.00	4.00	4.00	0.00
Bus	28	1121.56	4.00	4.25	0.25
	32	610.79	4.00	4.25	0.25
	36	326.78	4.00	4.25	0.25
	40	180.65	4.00	4.00	0.00
Stefan	28	1070.66	4.00	4.00	0.00
	32	529.76	4.00	4.00	0.00
	36	274.77	4.00	4.00	0.00
	40	155.16	4.00	4.00	0.00

5.2. Objective Quality Evaluation

We choose PSNR as objective performance measure. To see the overall performance of our proposed scheme for all the test sequences, we employ average PSNR difference and average bit rate difference as performance measure [11]. These measures are often used to compare RD performance between two different methods. The results are shown in Table 2. From the table, we can see that our scheme performs better than JVT-H014. Especially for sequence "Bike", our scheme obtains an average PSNR gain of up to 0.138dB and at the same time an average bit rate saving of 2.069%.

Table 2 Average PSNR Difference and Bit Rate Difference

Sequence	Average PSNR Difference over Full Range of Bit Rate (dB)	Average Bit Rate Difference over Full Range of PSNR (%)
Bike	0.138	-2.069
Bus	0.009	-0.215
Stefan	0.004	-0.077

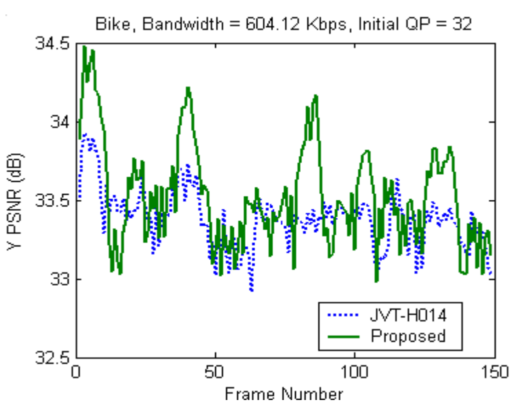

Figure 5 PSNR vs. Frame Number

Figure 5 shows the PSNR result with frame by frame for sequence "Bike" at initial QP = 32. From the figure, we can see that in most cases, our scheme performs better than JVT-H014.

6. CONCLUSION

In this paper, we aim at improving perceptual video quality and accurately estimating target bit for H.264 rate control. We have proposed a new motion complexity measure to represent the complexity of a frame's motion contents. We have also proposed a perceptual mode decision algorithm by updating Lagrangian multiplier according to the perceptual characteristics of video contents, so that more bits are allocated to edge patterns. Based on the motion complexity and perceptual mode decision, we have presented a new rate control scheme for H.264. In this scheme, we employ motion complexity to estimate a frame's target bit so that the bit allocation is in accordance with the complexity of frame's motion contents, and we employ perceptual mode decision to allocate the MB's bits perceptually. Our experimental results have shown that our rate control scheme outperforms the existing H.264 proposal.

7. REFERENCES

[1] Z. G. Li, W. Gao, F. Pan, S. W. Ma, K. P. Lim, G. N. Feng, X. Lin, S. Rahardja, H. Q. Lu, and Y. Lu, "Adaptive Rate Control with HRD Consideration", *8th JVT Meeting*, JVT-H014, Geneva, Switzerland, May 2003.

[2] ISO/IEC JTC1, *Information Technology — Coding of Audio-Visual Objects — Part 10: Advanced Video Coding*, ISO/IEC FDIS 14496-10, 2003.

[3] M. Q. Jiang, X. Q. Yi, and N. Ling, "Improved Frame-Layer Rate Control for H.264 Using MAD Ratio", *Proceedings of 2004 IEEE International Symposium on Circuits and Systems*, Vol. III, pp. 813-816, Vancouver, Canada, May 2004.

[4] S. Winkler, "A Perceptual Distortion Metric for Digital Color Video", *Proceedings of SPIE*, Vol. 3644, pp. 175-184, San Jose, USA, Jan. 1999.

[5] C. J. Tsai, C. W. Tang, C. H. Chen, and Y. H. Yu, "Adaptive Rate-Distortion Optimization using Perceptual Hints", *Proceedings of 2004 IEEE International Conference on Multimedia and Expo*, Taipei, China, June 2004.

[6] M. M. Ghandi and M. Ghanbari, "A Lagrangian Optimization Rate Control Algorithm for the H.264/AVC Encoder", *Proceedings of 2004 IEEE International Conference on Image Processing*, pp. 123-126, Singapore, Oct. 2004.

[7] T. Wiegand, H. Schwarz, A. Joch, F. Lossentini, and G. J. Sullivan, "Rate-Constrained Coder Control and Comparison of Video Coding Standards", *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 13, No. 7, pp. 688-703, July 2003.

[8] F. Pan, "Adaptive Image Compression Using Local Pattern Information", *Pattern Recognition Letters*, Vol. 23, Issue 14, pp. 1837-1845, Dec. 2002.

[9] A. M. Bazen and S. H. Gerez, "Systematic Methods for the Computation of the Directional Fields and Singular Points of Fingerprints", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 24, No. 7, pp. 905-919, July 2002.

[10] ITU-R, *Methodology for the Subjective Assessment of the Quality of Television Pictures*, ITU-R Recommendation BT.500-10, Mar. 2000.

[11] G. Bjontegaard, "Calculation of Average PSNR Differences between RD-Curves", *ITU-T Q.6/SG16 VCEG 13th Meeting*, Document VCEG-M33, Austin, USA, Apr. 2001.