Joint Rate Control for Multiple Sequences coding based on H.264 standard

Jing Yang, Xiangzhong Fang and Hongkai Xiong

Institute of Image Communication & Information Processing, Shanghai Jiaotong University, China.

Abstract

The objective of joint rate control is to dynamically distribute the channel capacity among video sequences according to their respective complexities, thus a more uniform picture quality and a more efficient utilization of channel capacity are achieved. Most existing approaches are based on MPEG2 coding platform. This paper presents a novel joint rate control scheme for multiple video sequences coding based on H.264 standard. A novel complexity measure that adapts to the characteristics of H.264 video coding is proposed. Experimental results show that the proposed scheme maintains a good balance in picture quality among the sequences as well as within a sequence.

1. Introduction

With recent advances both in digital video compression and in digital transmission technology, there are many occasions not only in high bit rate applications but also in low bit rate applications, that multiple sequences are transmitted over a single channel, such as in typical broadcasting systems (e.g. DBS or cable TV) and in real time video transmission applications (e.g. distant learning, video conferencing or TV over internet). How to reasonably allocate bit-rate among the sequences within a constrained fixed-rate network bandwidth, becomes a great deal to achieve consistent optimized video quality in the distributive application.

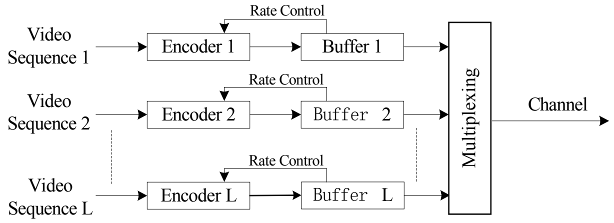

Fig.1. Independent coding of multiple sequences.

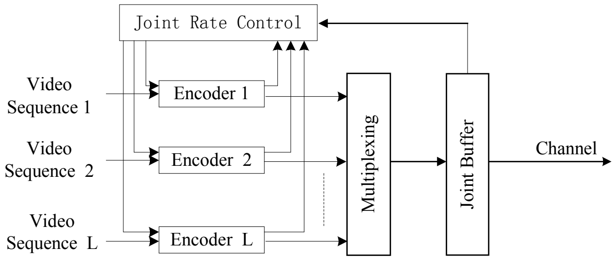

Fig. 2. Joint coding of multiple sequences.

Much research has shown that joint rate control is an effective way in maintaining a more uniform picture quality among sequences and within a sequence, as well as in better utilizing of channel capacity [2][4].

Basically, joint rate control of multi-sequence can be described as dynamically distributing the channel capacity among video sequences according to their respective complexities while keeping the total bit rate of multiple sequences a CBR conforming to the channel capacity. Existing approaches are mostly based on MPEG2 coding platform [1][2][3]. They basically followed the same complexity measure as that in TM5, which is the feedback statistical information from the previously coded frame. In [3], statistics characterizing the scene content of the video sources is added to make the complexity measure more accurate for current scene.

In this paper, we present a novel joint rate control scheme for H.264 coding of multiple video sequences. To achieve uniform picture quality between each sequence, a novel complexity measure that adapts to the characteristics of H.264 video coding is proposed, which can accurately indicate the coding complexity of the current frame.

In Section 2, the proposed scheme is described in detail. The complexity measure for each frame type is proposed. The target bits are assigned to the frame of each sequences in proportion to their complexities. Experimental results will be presented in Section 3.

2. Joint bit rate control

Assume that L sequences are to be delivered over one network, the independent coding and joint coding of multiple sequences are illustrated in Fig. 1 and Fig. 2, respectively. Notice that compared with independent coding in which each sequence has its own buffer control, there is only one channel buffer used in the joint coding. Moreover, in joint coding, data in relation to complexity measure from individual encoders as well as the channel buffer status are input parameters to the joint bit allocation controller. The number of target bit assigned to the frame of each sequence is output from the joint controller and is sent to the individual encoders. The channel buffer control is beyond our discussion in this paper. However, for the implementation of simulation, we apply an buffer control stratagy which is extended from that in the single sequence rate control scheme [6].

2.1. Complexity measure for each frame type

According to the latest JVT reference software JM76 and [6], a quadratic rate-distortion model same as in MPEG4 is used in the rate control [8][9], where the mean absolute difference (MAD) of the residual components is used as the complexity measure. Although MAD itself is a good indicator of the coding complexity of texture information of current frame, due to the specialties of H.264 video coding, two points should also be considered. First, for the known chicken and egg dilemma in H.264, the actual MAD of current frame is not available until after the rate-distortion optimization (RDO) process, whereas the quatization parameter must be decided before the RDO process begins. The MAD here used to determine the complexity of texture can be obtained only by prediction. Second, in the coding of H.264, flexible block-size and 1/4 fractional pixel are applied to obtain more precise prediction [7], whereas the bit count used to code motion information is also increased. Thus, the coding complexity of motion information should also be taken into account [5].

Here we assume that the number of successive B frames between I or P frames is the same among all the sequences. We need not discuss the target bit allocation for B frames as stated in [6], the quatization parameter of a B frame is calculated according to that of its adjacent P frames.

2.1.1. Complexity measure for I frame

In H.264, for I frames, the bits are mainly consumed to code residuals which are obtained by applying nine prediction modes for 4x4 block and four prediction modes for 16x16 block. Each prediction mode uses different weighted sum of horizontally and vertically adjacent reconstructed pixels to predict current block [7]. Thus, we define the intra prediction MAD as the complexity measure of I frame.

$$C_I = MAD_{intra} \tag{1}$$

In this paper, we propose a simple but very effective prediction model to predict the MAD of intra coded frames. The prediction value of MAD of intra-coded frame i is given by

$$\widetilde{MAD}_{intra}(i) = MAD_{intra}(i-1) \cdot \frac{\overline{M}_H(i) + \overline{M}_V(i) + \sigma \cdot \overline{act}(i)}{\overline{M}_H(i-1) + \overline{M}_V(i-1) + \sigma \cdot \overline{act}(i-1)} \tag{2}$$

If the coordinate of the top-left pixel of a 4x4 block is denoted by (M, N), then

$$M_H = \frac{1}{16}\sum_{m=0}^{3}\sum_{n=0}^{3} \left| lum(M+m, N+n) - lum(M-1, N+n) \right|$$

$$M_V = \frac{1}{16}\sum_{m=0}^{3}\sum_{n=0}^{3} \left| lum(M+m, N+n) - lum(M+m, N-1) \right|$$

$$act = \frac{1}{16}\sum_{m=0}^{3}\sum_{n=0}^{3} \left| lum(M+m, N+n) - \overline{lum} \right|$$

where $lum(M+m, N+n)$ is the luminance value of pixel $(M+m, N+n)$ and $\overline{lum}$ is the mean luminance value of the 4x4 block. $\overline{M}_H$, $\overline{M}_V$ and $\overline{act}$ is the average of M_H , M_V and act of the total frame, respectively. Although for different sequences, the best prediction results can be achieved by choosing different value of σ , in our experiment, we set σ = 0.7. Note $MAD_{intra}(i-1)$ denotes the actual intra MAD of the previous P frame, which will not bring excess computation since it is a by-product of the coding process of P frames, as the RDO of P frames always contains intra prediction.

The accuracy of the proposed prediction model is demonstrated in Fig.3 and Fig.4, where we can see the prediction value tracks the actual value very closely.

2.1.2. Complexity measure for P frame:

For P frames, the bits are mainly consumed not only on the residuals obtained by inter prediction but also on motion information.

The complexity of texture information of P frame is given by

$$C_{TEXTURE} = MAD_P \tag{3}$$

To solve the chicken and egg dilemma in the coding of P frames, we apply the linear model in [6] to predict the MAD of the current frame. Suppose that the predicted MAD of the current frame is denoted by $\widetilde{MAD}_P(i)$, the actual MAD of the previous frame is denoted by $MAD_P(i-1)$, the linear model is given by

$$\widetilde{MAD}_P(i) = a_1 \cdot MAD_P(i-1) + a_2 \tag{4}$$

The initial value of a_1 and a_2 are set to 1 and 0, respectively. They are updated after coding each frame.

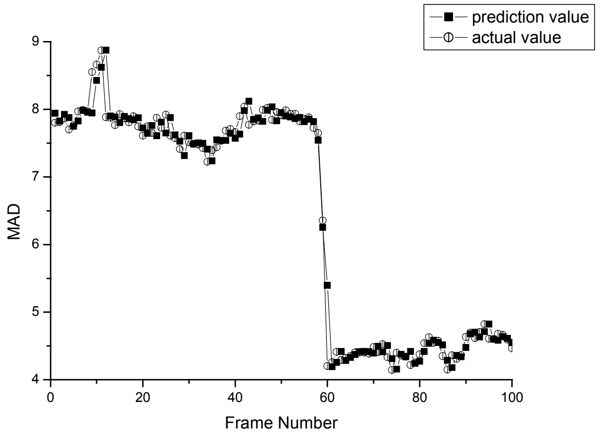

Fig. 3. Comparison between prediction value and actual value of intra MAD for sequence *trevor.qcif*

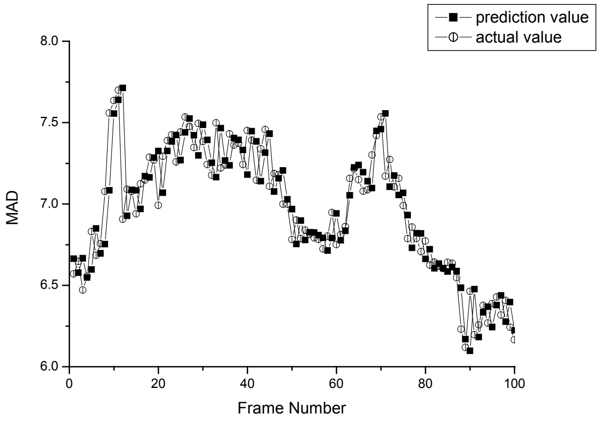

Fig. 4. Comparison between prediction value and actual value of intra MAD for sequence *foreman.qcif*

The complexity measure of motion information is defined as

$$C_{MOTION} = x_1 T_{MOTION} Q \tag{5}$$

where x_1 is an equivalent coefficient and its value is set to

$$x_1 = MAD_P / T_{TEXTURE} Q \tag{6}$$

T_{MOTION} and $T_{TEXTURE}$ denote the bit count used to code motion information and texture, respectively. Q denotes the quatization parameter. The coding complexity of motion information of current frame is estimated by that of the previous P frame, i.e.,

$$\widetilde{C}_{MOTION}(i) = C_{MOTION}(i-1) \tag{7}$$

The total complexity of P frame i is the sum of motion complexity and texture complexity,

$$C_p(i) = \widetilde{C}_{TEXTURE}(i) + \widetilde{C}_{MOTION}(i) \tag{8}$$

Instead of simply using the coding complexity of the previous frame to substitute for that of the current frame, the complexity measure described here more accurately indicates the complexity of the current frame. Meanwhile, since MAD , T_{MOTIOM} and $T_{TEXTURE}$ are also used in the single sequence rate control [6], the proposed complexity measure is suitable for H.264 coding platform, thus is easy to be implemented.

2.2. Target bit allocation

The number of bits assigned to each frame is in proportion to their complexity. The target bit rate for the frame of each sequence is sent to their respective encoder, where the rest of the single sequence rate control is performed and the final quantization parameter for each frame is decided.

3. Experimental results

To evaluate the proposed bit allocation strategy, we choose two groups of video sequences for testing. The first group contains four sequences with CIF resolution, and the second group contains six sequences with QCIF resolution. For each group, the sequences are coded both independently and jointly at a frame rate of 15 HZ and with two B frames between I- or P- frames. All the encoders are based on JM76. The total channel bandwidth is set to 1024 kbits/s for test group one and 288 kbits/s for test group two. In independent coding, the channel bandwidth is shared equally by all the sequences while in joint coding, the bandwidth is dynamically allocated to each sequences using the proposed bit allocation strategy.

Figs. 5-8 show the comparisons in PSNR between independent coding and joint coding for the two test groups. In independent coding, the picture quality varies a lot for different contents of sequences. It is not ideal for the customer to view a program with visible picture quality jitter. By applying joint rate control, it is clear that a more uniform picture quality among sequences is achieved by joint coding. Moreover, the picture quality variation within a sequence tends to be much smoother in joint coding than in independent coding. The PSNRs of the more complex sequences are significantly increased at the cost of moderate decreases of the easier sequences.

4. Conclusion

In this paper, we proposed a novel joint rate control scheme for H.264 coding of multiple video sequences. Bits are dynamically allocated to the frame of each sequence based on the proposed complexity measure which can accurately indicate the coding complexity of current frame. Simulation demonstrates that the proposed

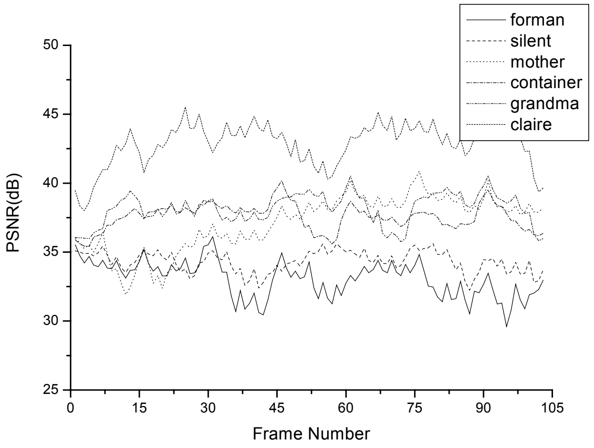

Fig. 7. PSNRs of independent coding for six QCIF sequences.

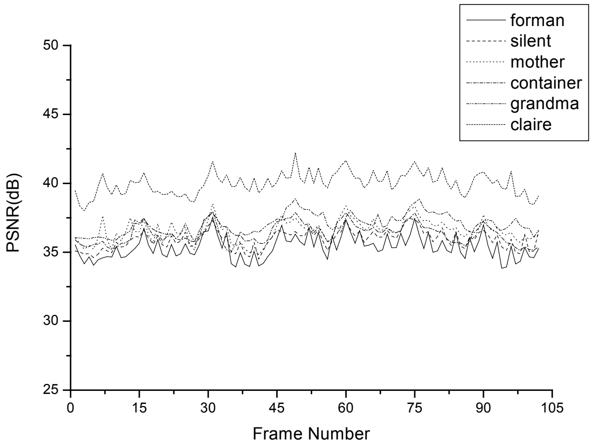

Fig. 8. PSNRs of joint coding for six QCIF sequences.

scheme maintains a good balance in picture quality among the sequences as well as within a sequence, compared to independent coding. Meanwhile, the scheme is easy to be implemented as all the input data to the joint rate controller can be easily extracted from individual H.264 encoder.

References

[1] Moving Pictures Expert Group, MPEG-2 Test Model 5, Doc ISO/IEC JTC1/SC29/WG11/ N0400, April 1993.

[2] L. Wang and A. Vincent, “Bit allocation and constraints for joint coding of multiple video programs,” IEEE Trans. Circuits Syst. Video Technol., vol. 9, no. 6, pp. 949-959, Sept. 1999.

[3] L. Boroczky, A. Y. Ngai and E. F. Westermann. “Joint rate control with look-ahead for multi–program video coding,” IEEE Trans. Circuits Syst. Video Technol., vol. 10, no.7, pp. 1159–1163, October 2000.

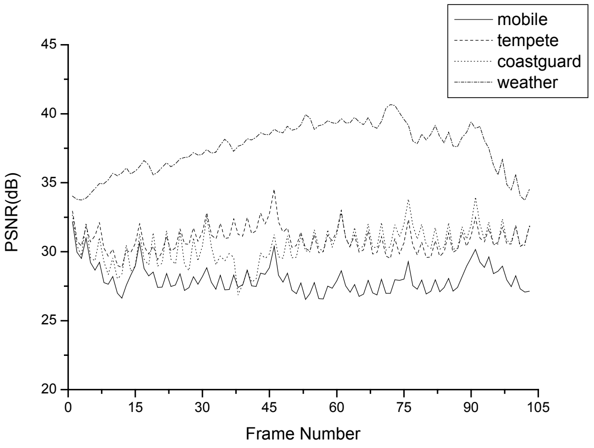

Fig. 5. PSNRs of independent coding for four CIF sequences.

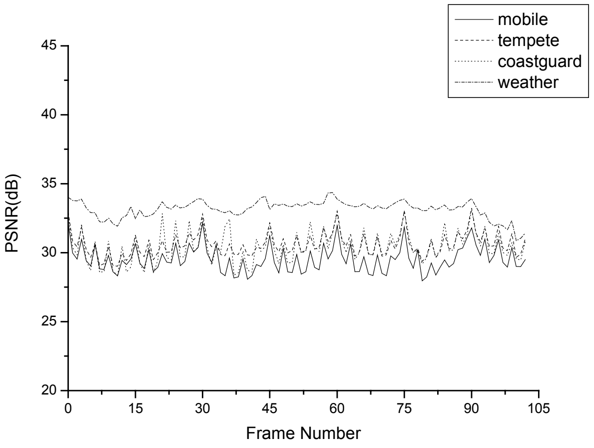

Fig. 6. PSNRs of joint coding for four CIF sequences.

[4] P. Corriveau, A. Vincent, “Subjective Evaluation of Joint Versus Independent Coding for Multi-Program Transmission”, ITU-R WP11E/WP11B, DOC llE/36-E: 34-E, 25 March 1997.

[5] A. Vetro, H. Sun, and Y. Wang. “MPEG-4 rate control for multiple video objects,” IEEE Trans. Circuits Syst. Video Technol., 9(1): 186–199, February 1999.

[6] “Proposed Draft of Adaptive Rate Control,” Joint Video Team (JVT) of ISO/IEC MPEG & ITU-T VCEG 8th Meeting: Geneva, May 20-26, 2003 Document: JVT-H017 Filename: JVT-H017.doc

[7] “Draft ITU-T Recommendation and Final Draft International Standard of Joint Video Specification,” Joint Video Team (JVT) of ISO/IEC MPEG & ITU-T VCEG 7th Meeting: Pattaya, Thailand, 7-14 March, 2003 Document: JVT-G050 Filename: JVT-G050d35.doc

[8] T. Chiang and Y.-Q. Zhang, “A new rate control scheme using quadratic rate-distortion modeling,” IEEE Trans. Circuits Syst. Video Technol., vol 7, no. 1, pp. 246-250, Feb. 1997.

[9] H.J. Lee, T. Chiang, and Y.-Q. Zhang, “Scalable rate control for MPEG-4 video,” IEEE Trans. Circuits Syst. video Technol., vol. 10, pp. 878–894, Sept. 2000.