Chinese Journal of Electronics
Vol.21, No.4, Oct. 2012

A Rate-Distortion Model Based Frame Layer Rate Control Algorithm for Stereoscopic Video Coding*

WANG Qun, ZHUO Li, ZHANG Jing and LI Xiaoguang

(*Signal and Information Processing Laboratory, Beijing University of Technology, Beijing 100124, China*)

Abstract — Rate control plays an important role in video coding and transmission. In this paper, a novel rate-distortion model has first been proposed to characterize the coding characteristics of stereoscopic video coding, where the weighted average of the left and right viewpoint measured with the Video quality metric (VQM) is adopted as the stereoscopic video coding distortion metric, instead of Mean square error (MSE). Then a frame layer rate control method for stereoscopic video coding has been presented based on the proposed R-D model. Experimental results demonstrate that, the proposed R-D model can accurately characterize the relationship among coding distortion, coding rate and quantization parameter and the proposed rate control method can efficiently control the output bit rate consistent with the target bit rate while the R-D coding performance is superior to that of JMVC 4.0.

Key words — Stereoscopic video, Video quality metric (VQM), Rate distortion model, Rate control.

I. Introduction

Rate control plays an important role in video coding and transmission. On the one hand, it is an essential component for robust video transmission, especially over time-varying and narrowband channel, where the transmission channels usually have fluctuated bandwidth. Hence, with the rate control technique, we can control the output bit rate according to the channel conditions and buffer size etc. On another hand, rate control is also beneficial to improve the video quality. The compression efficiency of the video encoder and the reconstructed video quality can be greatly improved through optimal bit allocation.

Source model was first presented by Hang H.M. *et al* in 1997[1], which describes the relationships among the output rate (R), coding distortion (D) and quantization parameter (QP). Source models are also called Rate-distortion (R-D) models to characterize the coding performance of the video encoder. Currently, many R-D model based rate control algorithm have been applied in many coding standards. Such as the first-order linear model has been used in MPEG-2 TM5 rate control algorithm[2], and the quadratic model has been used in MPEG-4 VM8 algorithm[3]. TMN8, proposed by Jordi Ribas Cobera *et al.*, has been used in H.263 rate control algorithm, and its R-D model is the combination of the logarithmic model and the quadratic model[4]. With recent advances in stereoscopic video coding, the research on rate-distortion model based rate control for stereoscopic videos has attracted high interest over the past years. Generally speaking, the rate-distortion models used for the stereoscopic video coding currently are mostly the improvements of traditional quadratic R-D model[5,6].

The rest of this paper is organized as follows. The proposed rate-distortion model for stereoscopic video coding is studied in the Section II. The frame layer rate control method based on the proposed R-D model is presented in the Section III. The experimental results are analyzed in Section IV. Section V concludes the paper.

II. Proposed Rate-Distortion Model for Stereoscopic Video Coding

1. Coding framework of stereoscopic video

In this paper, MVC is used as the coding framework of stereoscopic video. MVC was proposed by ITU-T and MPEG Joint Video Team (JVT) to achieve a high multi-view video coding efficiency. The coding architecture of stereoscopic video coding is shown in Fig.1[7]. It can be seen that the stereoscopic video encoder usually adopts the combination of Disparity compensated prediction (DCP) and Motion compensated prediction (MCP) method to completely remove the redundant information.

2. Coding distortion measurements of stereoscopic video

VQM, which is developed by the Institute of Telecommunication Sciences (ITS) and American National Standard

*Manuscript Received Dec. 2011; Accepted Jan. 2012. This work is supported by Program for New Century Excellent Talents in University, the National Natural Science Foundation of China (No.61003289, No.61100212), the Natural Science Foundation of Beijing (No.4102008), the Excellent Science Program for the Returned Overseas Chinese Scholars of Ministry of Human Resources and Social Security of China, Scientific Research Foundation for the Returned Overseas Chinese Scholars of MOE.

Institute (ANSI), is a standardized objective video quality metric[8]. VQM can represent both the perceived overall image quality and depth of 3-D video well and has shown a better performance in terms of stereoscopic video coding distortion compared with the traditional PSNR metric[9].

It's known that, according to the characteristics of human visual perception, if the right and left views are displayed with different quality and resolutions, the overall 3-D video quality is determined by the view with the better quality and resolution[10]. Based on this theory, in this paper, we use the weighted average VQM of the left and right views instead of the average VQM as the stereoscopic video coding distortion metric, where the weights are set to 0.7 (left) and 0.3 (right) respectively.

3. The actual R-D model for stereoscopic video coding

MVC coding framework supports two kinds of coding modes, *i.e.* Intra and Inter mode. The coding structure of I frame and P frame is quite different, thus the coding distortion property also varies. Therefore, we need to set up their R-D model respectively.

In order to analyze the rate distortion characteristics of stereoscopic video encoder, three kinds of 3-D video sequences with different motion characteristics are tested in this paper firstly. For each sequence, 250 frames are chosen for the test, and the frame rate is fixed to 25 frames/s. We take 2 views from 8 views video to test using the weighted average VQM of the left and right views as the coding distortion metric. To measure the R-D characteristics of I frame, the coding structure is set as all I frames. While for P frames, the IPPP···IPPP··· coding structures is utilized. The GOP size is set as 8 and the QP value as 5, 10, 15···50 respectively for each sequence.

Fig.2 shows the relationship curves of rate vs. distortion, quantization parameter vs. distortion, and rate vs. quantization parameter of I frame for the Ballroom sequence, which are simply called R-D, Q-D and QP-R curve respectively. Fig.3 shows the P frame results of Ballroom sequence. From Fig.2 and Fig.3, it can be seen that a quadratic polynomial model exhibits the best correlation with Q-D curve for both I frame and P frame. However, it is difficult to approximate the relationship of R-D and QP-R.

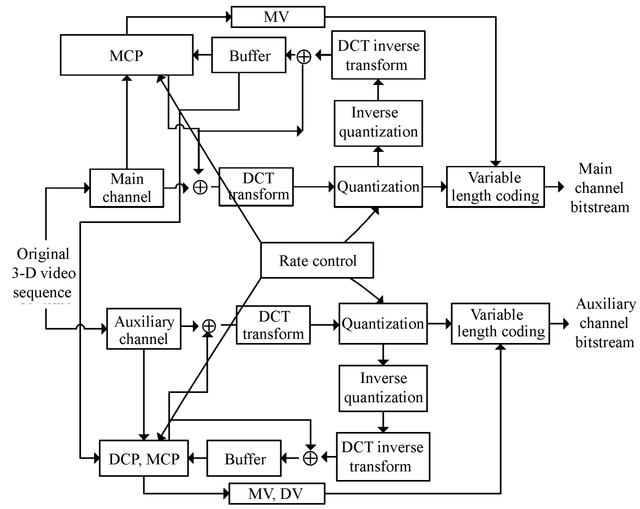


Fig. 1. The coding framework for stereoscopic video coding

4. The proposed R-D model for stereoscopic video coding

We have found through a large number of experiments that if we adopt exponential function directly to fit the R-D and R-Q curves in Fig.2 and Fig.3, its accuracy is very low and it

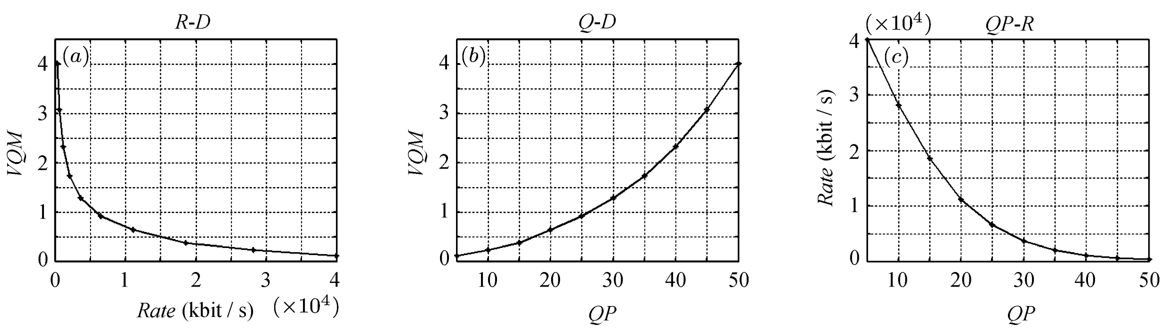


Fig. 2. *R-D*, *Q-D* and *QP-R* curves of Ballroom sequence I frame. (*a*) *R-D* curve; (*b*) *Q-D* curve; (*c*) *QP-R* curve

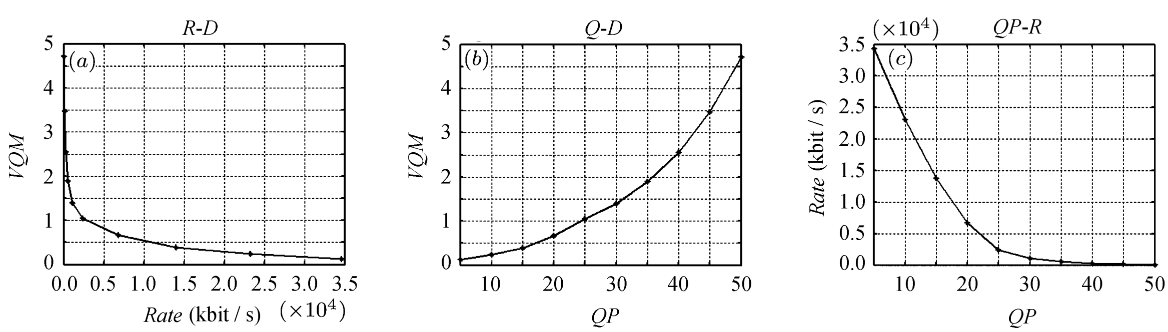


Fig. 3. *R-D*, *Q-D* and *QP-R* curves of Ballroom sequence *P* frame. (*a*) *R-D* curve; (*b*) *Q-D* curve; (*c*) *QP-R* curve

can lead to the rapid increase of computational complexity. However, if we compute logarithmic operation on R, and list the results again in Fig.4 and Fig.5, we can see clearly that a cubic polynomial model can describe the $\log R - D$ and $\log R - Q$ curves well whether for I frame or P frame, which will reduce the computational complexity of the model to a certain extent.

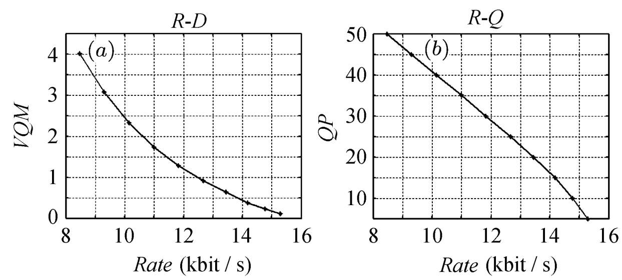


Fig. 4. log(R)-D and log(R)-Q curves of Ballroom I frame. (a) log(R)-D curve; (b) log(R)-Q curve

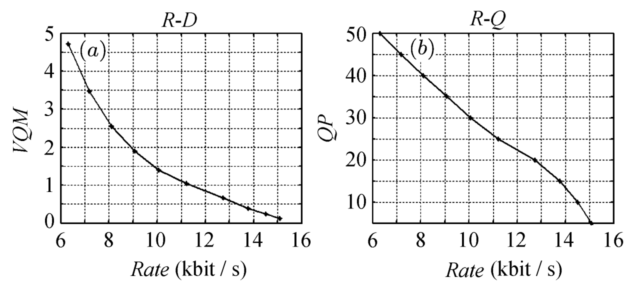


Fig. 5. log(R)-D and log(R)-Q curves of Ballroom P frame. (a) log(R)-D curve; (b) log(R)-Q curve

Therefore, considering the trade-offs between the model complexity and accuracy, we characterize the rate distortion characteristics of both I frame and P frame using the same simple cubic model. In this paper, rate-distortion model for stereoscopic video coding has been proposed as follows:

$$\begin{cases} R-D: D_p = p_1 \log^3(R) + p_2 \log^2(R) + p_3 \log(R) + p_4 \\ Q-D: D_p = p_5 QP^3 + p_6 QP^2 + p_7 QP + p_8 \\ R-Q: QP = p_9 \log^3(R) + p_{10} \log^2(R) + p_{11} \log(R) + p_{12} \end{cases} \tag{1}$$

where QP is the quantization step, R is the coding rate, D_P is the coding distortion measured using VQM, p_i, $i = 1, 2, \cdots, 12$ are the parameters of the proposed model.

5. Experimental results and analysis

The accuracy of R-D model is essential for the subsequent rate control results. Several 3D video sequences are tested and compared to validate the accuracy of R-D model presented in this paper. Fig.6 shows the comparison of the model data and actual test data of Vassar sequence. From Fig.6, it can be seen that the R-D model presented in this paper coincides well with the actual R-D curves, and the model can accurately characterize the relationship among coding distortion, coding rate and quantization parameter. Similar results have been achieved when other sequences were tested.

III. Proposed Frame Layer Rate Control Method

The goal of rate control method is to control the output bit rate of the encoder consistent with the target bit rate Under the given target bit rate constraint, the first step is to estimate the Quantization parameter (QP) of source encoder according to the proposed model. And then encode the current frame with the estimated QP parameters.

In this paper, the basic rate control unit is GOP. For the R-D model based rate control method, the key is to calculate the parameters of the model. Duo to the characteristics of the neighboring frames are very close to each other, the coding statistics of previous coded frames can be utilized to estimate the model parameters of the current frame. Therefore, the method can be implemented with the following steps:

Step 1 Initialization

• Encode the I frame with the preset QP parameters and obtain the output bit number B_I. Set $Encoded\_Frame\_Num = 1$. Subtract B_I from the target bits B_T and get the remaining bits B_P used for the P frames of GOP.

$$B_P = B_T - B_I \tag{2}$$

where $B_T = \dfrac{R_{target}}{Frame\_rate} \cdot GOP\_Size$, GOP_Size is the GOP size and R_{target} the target bit rate.

• Encode the 1st, 2nd, 3rd and 4th P frames with the preset QP parameters and achieve four sets of data, *i.e.* (R_1, QP_1), (R_2, QP_2), (R_3, QP_3) and (R_4, QP_4). Set $Encoded\_Frame\_Num = 5$.

Step 2 Determine the QP

• Calculate the target bits B_i of the remaining P frames in the same GOP using the Eq.(3):

$$B_i = \frac{B_p - \sum\limits_{i=1}^{Encoded\_Frame\_Num} R_i}{GOP - Encoded\_Frame\_Num} \tag{3}$$

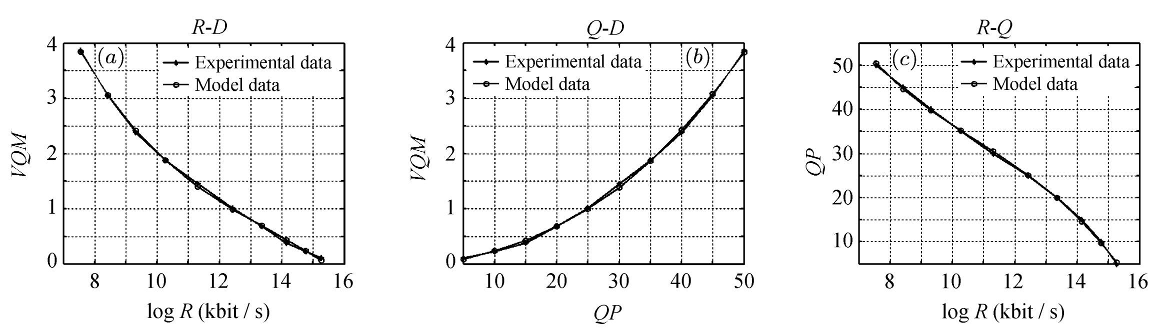


Fig. 6. Comparison between proposed model and actual data of vassar sequence. (a) R-D curve; (b) Q-D curve; (c) R-Q curve

• Estimate the model parameters of Eq.(1) using (R_1, QP_1), (R_2, QP_2), (R_3, QP_3) and (R_4, QP_4), then use these model parameters to determine the QP value of the current frame corresponding to the target bits.

• Encode the frame with the achieved QP and obtain the output bit number of the current frame.

Step 3 Update

• Update the coding data in Eq.(1) using the newly output bit number and QP. Use these coding data to estimate the model parameter of the next frame.

• Update $Encoded\_Frame\_Num = Encoded\_Frame\_Num + 1$.

Step 4 Loop over frames

• Repeat step 2 and 3 until all the P frames in the current GOP are encoded.

IV. Experimental Results and Analysis

In order to validate the effectiveness of our proposed rate control method in this paper, we performed experiments over several stereoscopic video test sequences with different motion characteristics. And we compared our proposed method with the fixed QP method of JMVC reference software. In the experiments, we test 250 frames for each sequence, and the frame rate is fixed to 25 frames per second. The coding structure of IPPP···IPPP···is utilized to encode each GOP. The GOP size is set as 15. The Rate control error (RCE) is used to measure the accuracy of rate control method:

$$RCE = \frac{|R_{target} - R_{actual}|}{R_{target}} \times 100\% \tag{4}$$

Table 1 illustrates the difference between the target bit rate and the actual output bit rate for various video sequences. From Table 1, it can be seen that the proposed rate control method can efficiently control the output bit rate consistent with the target bit rate. The average rate control error is about 3%. Fig.7 and Fig.8 show the R-D performance curves of Ballroom and Exit sequences using the proposed rate-controlled method and JMVC 4.0 method respectively. It can be seen from the Fig.7 and Fig.8 that, the R-D performance of the proposed RC algorithm is superior to that of JMVC 4.0 method, and a certain amount of coding gain have been achieved for some sequences.

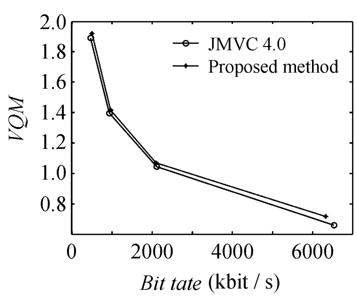


Fig. 7. Comparison of *R-D* performance with different methods for Ballroom sequence

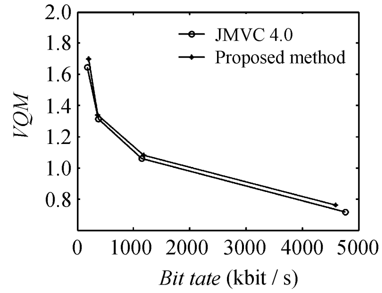


Fig. 8. Comparison of *R-D* performance with different methods for Exit sequence

Table 1. The difference between the output bit rate and the target bit rate for various video sequences

Test sequences	Bit rate (kbit/s)		RCE (%)
	Target bit rate	Actual bit rate	
Ballroom	6543.068	6324.1224	3.346
	2123.4784	2112.5580	0.514
	946.044	979.6448	3.552
Exit	4769.2176	4592.6684	3.702
	1145.7196	1175.4584	2.596
	372.3664	365.9608	1.720
Vassar	7177.8368	6873.018	4.247
	1944.4324	1805.6980	7.135
	373.874	381.7824	2.115
Race1	4271.6256	4111.7936	3.742
	2121.6644	2202.5396	3.812
	927.166	949.7148	2.432
Flamenco2	3359.4508	3517.4008	4.702
	1798.9368	1823.2428	1.351
	921.9116	955.1424	3.604

Generally speaking, the proposed rate control method can efficiently control the output bit rate consistent with the target bit rate while the R-D coding performance is superior to that of JMVC4.0 method. The proposed rate control method can be applied in the stereoscopic video coding and transmission applications.

V. Conclusion

In this paper, a novel R-D model for stereoscopic video coding is first proposed using the weighted average VQM of the left and right views as the stereoscopic video distortion metric, where the weights are set to 0.7 and 0.3 respectively. A cubic polynomial model is proposed to describe the correlation curves *i.e.* $\log R$ and distortion D, as well as $\log R$ and QP for both I frame and P frame. The experimental results indicate that, the R-D model presented in this paper coincides well with the actual stereoscopic video coding curves. Then a rate control method for stereoscopic video coding at frame layer has been presented based on the proposed R-D model. Experimental results demonstrate that, the proposed rate control method can efficiently control the output bit rate consistent with the target bit rate while the R-D coding performance is superior to that of JMVC4.0 method.

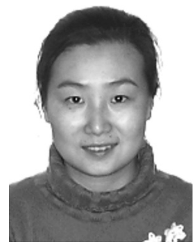
References

[1] H. Hang, J. Chen, "Source model for transform video coder and its application in fundamental theory", *IEEE Trans. Circuits Syst. Video Technol.*, Vol.7, No.2, pp.287–298, 1997.

[2] ISO/IEC JTC1/SC29/WG11, MPEG-4 test model5. 1993.
[3] ISO/IEC JTC1/SC29/WG11, MPEG-4 Video Verification Model Version 18.0, 2001.
[4] ITU-T/SG15.Video codec test model, TMN8. 1997.
[5] J.L. Chen, "Research on mulit-view video coding", *Ph.D. Thesis*, Zhejiang University, Hangzhou, China, 2006.
[6] Z.J. Zhu, F. Liang *et al.*, "Bit-allocation and rate-control algorithm for stereo video coding", *Journal on Communications*, Vol.28, No.7, pp.15–21, 2007.
[7] Z.J. Zhu, G.Y. Jiang, M. Yu, "Fast disparity estimation algorithm for stereo video coding", *Proceedings of 2002 IEEE Region 10 Conference on Computers, Communications, Control and Power Engineering*, Beijing, China, pp.285–288, 2002.
[8] M. Pinson, S. Wolf, "A new standardized method for objectively measuring video quality", *IEEE Transactions on Broadcasting*, Vol.50, No.3, pp.312–322, Sept. 2004.
[9] Chaminda T.E.R. Hewage *et al.*, "Quality evaluation of color plus depth map-based stereoscopic video", *IEEE Journal of Selected Topics in Signal Processing*, Vol.3, No.2, pp.304–318, 2009.
[10] S.L.B. Stelmach, W.J. Tam, D. Meegan and A. Vincent, "Stereo image quality: Effects of mixed spatio-temporal resolution", *IEEE Trans. Circuits Syst. Video Technol.*, Vol.10, No.2, pp.188–193, Mar. 2000.

WANG Qun was born in 1986. She received the B.S. degree in Communication Engineering from Shandong University of Science and Technology, Qingdao, China, in 2009. Since then, she has been a graduate student in Information and Communication Engineering at Beijing University of Technology. Her research interests mainly include stereoscopic video coding and transmission. (Email: wangqun@emails.bjut.edu.cn)

ZHUO Li was born in 1971. She received the B.S. degree in Radio Technology from the University of Electronic Science and Technology, Chengdu, China, in 1992, the M.S. degree in Signal and Information Processing from the Southeast University, Nanjing, in 1998, and the Ph.D. degree in Pattern Recognition and Intellectual System from Beijing University of Technology, in 2004. She has been a professor of Beijing University of Technology since 2007. Her research interests include secure multimedia signal processing, stereoscopic video coding and transmission. (Email: zhuoli@ bjut.edu.cn)



ZHANG Jing was born in 1975, Ph.D. Associate Professor and M.S. student supervisor of Beijing University of Technology. Her research interest is image/video signal and processing. (Email: zhj@bjut.edu.cn)

LI Xiaoguang was born in Beijing, China. He received the B.S. and Ph.D. degrees in Electronic Engineering from the Beijing University of Technology, in 2003 and 2008, respectively. He is currently a Lecturer and M.S. student supervisor of the Beijing University of Technology. His research interests include image super resolution and high dynamic range image processing. (Email: lxg@bjut.edu.cn)