Research Article

Macroblock Layer Rate Control Based on Structural Similarity and Mean Absolute Difference for H.264

Xiao Chen^{1,2} and Dongjue Gu²

¹ Jiangsu Key Laboratory of Meteorological Observation and Information Processing, Nanjing University of Information Science and Technology, Nanjing 210044, China

² School of Electronic and Information Engineering, Nanjing University of Information Science and Technology, Nanjing 210044, China

Correspondence should be addressed to Xiao Chen; rainofsun@nietease.com

Received 20 April 2013; Revised 24 December 2013; Accepted 2 January 2014; Published 11 February 2014

Academic Editor: Costas Kotropoulos

Copyright © 2014 X. Chen and D. Gu. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In the process of the H.264 video coding, special attention should be paid to the subjective quality of the image. This paper applies the structural similarity (SSIM) based subjective evaluation to the rate control in the H.264 coding and proposes to combine the SSIM and the mean absolute difference (MAD) to perform the macroblock layer bit allocation instead of the MAD. Experimental results show that the proposed method is correlating better with the human visual system and thus achieves better subjective image quality.

1. Introduction

The JVT-G012 algorithm for the H.264 video coding uses the mean absolute difference (MAD) linear prediction model to solve the “chicken and egg dilemma” [1]. The basic unit layer number of bits (including the macroblock layer) is allocated averagely. As JM10.1 adopts the method, the distribution scheme mainly allocates the bit rate from the content of the complexity of the natural image. It has some disadvantages—it does not allocate the bit rate according to the human’s subjective characteristic. The obtained image is not consistent with the subjective characteristics of the human eyes.

Some scholars made a lot of improvement to overcome these shortcomings. For example, Lu put forward the motion extrapolated coding complexity to allocate bits between the region of interest (ROI) and other regions which reduced the distortion in the ROI using the rate distortion optimized macroblock level bit allocation [2]. Han et al. used a content complexity factor to allocate the bits more accurately and a quantization parameter adaptation factor to adjust the current quantization parameter according to the encoded frames history information [3]. Chen and Lu put forward the frame complexity factor to optimize the frame layer target bit allocation [4]. Liu et al. combine an inter-frame subtraction with the background subtraction to detect the target region

[5]. To achieve the purpose of high-definition of the target region under the low bit rate, an optimized rate control scheme is proposed. But these methods are not, according to the HVS (human visual system) model, to allocate the bits of the macroblock.

Wang et al. proposed the SSIM (structural similarity) as a new indicator to measure the similarity of two images. It is a method used for evaluating the quality of image [6]. Due to the fact that the human visual can extract image structure information easily, the calculation of two-image structure similarity can be used as a standard to evaluate the image quality. The bigger the value is, the better the quality is. The biggest of the SSIM is 1. The SSIM is better than the MSE (Mean Squared Error) and the PSNR (peak signal to noise ratio) in the image similarity evaluation.

Cui et al. put forward an empirical model of the SSIM linear distortion model and combined it with the improved secondly rate-quantitative model using Lagrange multiplier method to get the closed solution of the SSIM optimal MB layer quantization [7]. Cui et al. also used the SSIM to guide the RDO frame mode selection [8]. Wang et al. put forward a method of interest video quality evaluation method [9]. The method presented the extraction of the frame’s interest area to weigh the calculated SSIM of the block in the frame. Yang et al. introduced the SSIM as the distortion measurement for

the H.264 interframe prediction [10]. Yang et al. found the myopia relationship between the SSIM and rate, and then it was integrated into SSE as distortion measure of RDO function and combined with human visual characteristics [11]. It built the distortion measure of RDO model combining the SSE with the SSIM. Cui and Zhu applied the SSIM based subjective distortion to RDO-based intramode decision in H.264 I frame coding, and they further proposed a frame layer adaptive Lagrange multiplier adjustment scheme to get better tradeoff between the rate and the SSIM distortion [12]. Wu et al. used the relationship between the reconstructed macroblock and the best prediction macroblock from mode selection [13]. Cui and Zhu put forward a kind of the adaptive skipped frame scheme by recovering the skip frame in the decoding port through the reference frame [14].

The above methods did not use the SSIM to allocate the bits of the macroblock layer. This paper will focus on how to use the SSIM theory to allocate the bit of the macroblock layer in order to improve the subjective video quality.

2. Improved Rate Control in MB

The pixels of the image are related. This contains the structural information of the image. In the observation of the image, the human eyes are more sensitive to changes of the structure information than individual pixel values. In the SSIM theory, the SSIM index is defined from the view of image composition, which reflects the scene object structure properties. It is independent of the brightness and contrast of the image. The distortion modeling is built as the combination of three different factors—brightness, contrast, and structure. The SSIM of the original image x to the distorted image y is defined as [6]

$$\text{SSIM} = \frac{(2u_x u_y + c_1)(2\delta_{xy} + c_2)}{(u_x^2 + u_y^2 + c_1)(\delta_x^2 + \delta_y^2 + c_2)}, \quad (1)$$

where u_x and u_y are the brightness mean of the images x and y , δ_x^2 and δ_y^2 are the standard deviation the images x and y , δ_{xy} is the covariance between them, and c_1 and c_2 are the small constants in order to avoid that the denominator is zero. In the calculation of frame SSIM, the general is to divide the image into the fixed size MBs and then calculate the SSIM of each MB. The SSIM measurement has a boundary, which is not more than 1. The closer the value is to 1, the better the subjective quality of the distorted image is. The higher SSIM value implies that the similarity between the distorted image and the original image is high, so the image quality is better. Therefore, the distortion of the i th MB in the P frame can be defined as $1 - \text{SSIM}(i)$.

The MBs in the same position of the adjacent frames have strong temporal correlation and thus have the similar SSIM distortion characteristics. Therefore, the SSIM of the current MB can be estimated by the actual SSIM of the same position MB in the previous frame.

In the JVT-G012 algorithm for the H.264 video coding, the bit rate of the MB is allocated averagely. The obtained image is not consistent with the subjective characteristics

of the human eyes. This paper applies the SSIM based subjective evaluation to the rate control in the H.264 coding and proposes to combine the SSIM and the mean absolute difference to perform the MB layer bit allocation instead of the mean absolute difference.

The implementation of rate control for the MB layer mainly includes the bit allocation, calculations of the quantization parameter, and buffer control. First, we allocate the target bits according to the remaining bits, SSIM and MAD. Then, we calculate the quantization parameter and adjust it. Last, we perform rate-distortion-optimization (RDO) and update the model parameters.

Firstly, the bit $R(i)$ of the i th MB in the frame is allocated by

$$R(i) = T_r \left[a \frac{[1 - \text{SSIM}(i)]^2}{\sum_{k=1}^N [1 - \text{SSIM}(k)]^2} + b \frac{\text{MAD}_{\text{EMV}}(i)^2}{\sum_{k=1}^N \text{MAD}_{\text{EMV}}(k)^2} \right], \quad (2)$$

where i is the serial number of the MB in the frame, N is the total number of the MBs in the frame, and k is the loop variable. a and b are the weighted coefficients to allocate bits and $a + b = 1$. T_r is the remaining bits in the frame. $\text{SSIM}(i)$ represents the SSIM of the i th MB in the frame; $\text{MAD}_{\text{EMV}}(i)$ represents the MAD of the i th MB in the frame. a and b are 0.4 and 0.6 after taking a lot of experiments.

Then, the quantization parameter of the current MB is computerized by using the quadratic R-D model. However, the inaccuracy of the model can produce the unexpected coded bits. Therefore, we adjust the current quantization parameter by the previous MB coding information, which can achieve more accurate rate control. In order to avoid the adjacent MBs' image quality jump too much, we adjust the quantization parameter. If the current macroblock is the first basic unit in the current frame, $\text{QP} = Q_{\text{apf}}$, Q_{apf} is the average quantization parameter of all the basic units in the previous frame. If the bit number of current frame is less than zero, the quantization parameter should be greater than that of the previous basic unit in order that the sum of generated bits is closed to the current frame bits. Therefore, $\text{QP} = \text{QP}_{\text{prev}} + 1$ where QP_{prev} is the quantization parameter of the previous macro block. Otherwise, using the quadratic model to compute a quantization parameter, $\text{QP} = \max\{\text{QP}_{\text{prev}} - 1, \min\{\text{QP}, \text{QP}_{\text{prev}} + 1\}\}$.

Finally, the algorithm has taken into account the coded frame information fully after the adjustment. We perform RDO and update the model parameters. In this paper, the SSIM is incorporated into the RDO framework as a quality metric [15] for all MBs in the current basic unit. For convenience, the method in reference 15 is called PRDO method.

3. Experimental Results and Discussions

The proposed algorithm in this paper is realized by enhancing the JM10.1 test model. The main experimenters are shown in Table 1. The purpose of the proposed algorithm is to improve the video subjective quality, so the original JM10.1 rate control algorithm is selected as a reference for the comparison,

FIGURE 1: The image subjective quality comparison of the 3rd (a) and 76th (b) frame for the carphone sequence.

because the bit allocation method in the JM10.1 algorithm is better than the original JVT-G012 algorithm. The proposed algorithm is also compared with the PRDO algorithm.

3.1. Subjective Visual Quality. Figures 1 and 2 give some subjective renderings for the carphone and foreman sequences of the proposed, JM10.1, and PRDO algorithms at the bit rate of 64 kbps. From the view of human being's eyes, the image

structure information of the proposed algorithm is better than the JM10.1 algorithm as shown in the figures (the top is original image, the middle is obtained from the proposed algorithm, and the bottom is obtained from JM10.1).

3.2. SSIM. Figures 3 and 4 show the SSIM comparison from the first to the one hundredth frame of the carphone and foreman sequences by using the proposed, JM10.1, and PRDO

FIGURE 2: The image subjective quality comparison of the 69th (a) and 82nd (b) frame for the foreman sequence.

TABLE 1: The main experimental parameters.

Coding mode	Format	Frame rate	Frame number	Target bitrate	Iframe circle	Search range	ME precision	Reference frame number	Entropy coding	RDO
IPPP	QCIF	30 fps	100	64 kbps	20	16	1/4	1	CAVLC	on

algorithms. As can be seen from the figures, the proposed algorithm improves the SSIM of the video sequence. The SSIM obtained from the proposed algorithm is better than other two algorithms. So it has the better subjective video quality.

Tables 2, 3, and 4 show the average SSIM of the proposed, JM10.1, and PRDO algorithms for the carphone, crew, highway, claire, and foreman sequences at the bit rates of 64, 128, 256, and 512 kbps when the initial quantization parameter is set to 23, 28, and 33, respectively. According to tables,

the proposed algorithm gets better average SSIM than the other algorithms for all video sequences at all bit rates. So the proposed algorithm can achieve better subjective video quality than other algorithms during the video coding.

3.3. PSNR. Figures 5 and 6 show the PSNR comparison from the first to the one hundredth frame of the carphone and foreman sequences by using the proposed, JM10.1, and PRDO algorithms. As can be seen from figures, the PSNR differences

TABLE 2: Average SSIM comparison of the proposed, PRDO, and JM10.1 algorithms (QP = 23).

Sequence	Target bit rate (kbps)	JM10.1	PRDO	Proposed
Crew	64	0.6032	0.6113	0.6140
Carphone	64	0.7851	0.7926	0.7954
Highway	64	0.7810	0.7879	0.7908
Foreman	64	0.7365	0.7468	0.7500
Claire	64	0.7498	0.7569	0.7597
Crew	128	0.7495	0.7560	0.7581
Carphone	128	0.8252	0.8321	0.8348
Highway	128	0.8235	0.8341	0.8362
Foreman	128	0.8452	0.8515	0.8540
Claire	128	0.8015	0.8078	0.8107
Crew	256	0.8634	0.8694	0.8722
Carphone	256	0.8853	0.8911	0.8943
Highway	256	0.8481	0.8544	0.8565
Foreman	256	0.8984	0.9025	0.9058
Claire	256	0.8433	0.8494	0.8519
Crew	512	0.9249	0.9280	0.9305
Carphone	512	0.9105	0.9146	0.9173
Highway	512	0.8742	0.8787	0.8815
Foreman	512	0.9368	0.9402	0.8424
Claire	512	0.8566	0.8619	0.8648

TABLE 3: Average SSIM comparison of the proposed, PRDO, and JM10.1 algorithms (QP = 28).

Sequence	Target bit rate (kbps)	JM10.1	PRDO	Proposed
Crew	64	0.6330	0.6409	0.6434
Carphone	64	0.7703	0.7788	0.7819
Highway	64	0.7748	0.7815	0.7846
Foreman	64	0.7611	0.7686	0.7717
Claire	64	0.7135	0.7215	0.7250
Crew	128	0.7710	0.7791	0.7815
Carphone	128	0.8430	0.8497	0.8529
Highway	128	0.8142	0.8198	0.8219
Foreman	128	0.8263	0.8332	0.8359
Claire	128	0.7753	0.7829	0.7855
Crew	256	0.8614	0.8678	0.8706
Carphone	256	0.8770	0.8829	0.8864
Highway	256	0.8428	0.8477	0.8503
Foreman	256	0.8862	0.8908	0.8942
Claire	256	0.8319	0.8378	0.8409
Crew	512	0.9236	0.9275	0.9296
Carphone	512	0.9071	0.9122	0.9146
Highway	512	0.8694	0.8742	0.8775
Foreman	512	0.9328	0.9359	0.9380
Claire	512	0.8437	0.8491	0.8524

of three algorithms are very small. The proposed algorithm achieves a little higher PSNR than the other two algorithms.

3.4. Bit Rate. Tables 5, 6, and 7 show the bit rates of the proposed, JM10.1, and PRDO algorithms for the carphone, crew, highway, claire, and foreman sequences at the target

bit rates of 64, 128, 256, and 512 kbps when the initial quantization parameter is set to 23, 28, and 33, respectively. As summarized in the tables, the proposed algorithm can control the bit rates more accurately than the other two algorithms.

The experimental results show that the proposed algorithm in this paper makes the macroblock layer bit rate

TABLE 4: Average SSIM comparison of the proposed, PRDO, and JM10.1 algorithms (QP = 33).

Sequence	Target bit rate (kbps)	JM10.1	PRDO	Proposed
Crew	64	0.6385	0.7461	0.7492
Carphone	64	0.7839	0.7908	0.7939
Highway	64	0.7540	0.7611	0.7641
Foreman	64	0.7572	0.7643	0.7670
Claire	64	0.6750	0.6835	0.6869
Crew	128	0.7683	0.7763	0.7787
Carphone	128	0.8333	0.8405	0.8434
Highway	128	0.7982	0.8050	0.8078
Foreman	128	0.8195	0.8266	0.8295
Claire	128	0.7860	0.7939	0.7972
Crew	256	0.8592	0.8657	0.8684
Carphone	256	0.8745	0.8801	0.8834
Highway	256	0.8274	0.8339	0.8366
Foreman	256	0.8820	0.8869	0.8893
Claire	256	0.8204	0.8256	0.8284
Crew	512	0.9142	0.9193	0.9214
Carphone	512	0.8974	0.9015	0.9037
Highway	512	0.8561	0.8608	0.8633
Foreman	512	0.9254	0.9334	0.9353
Claire	512	0.8419	0.8470	0.8499

TABLE 5: Bit rate comparison of the JM10.1, the PRDO, and the proposed algorithms (QP = 23).

Sequence	Target bit rate (kbps)	JM10.1 (kbps)	PRDO (kbps)	Proposed (kbps)
Crew	64	78.59	65.89	65.32
Carphone	64	64.38	64.42	64.26
Highway	64	64.83	64.71	64.54
Foreman	64	64.28	64.25	64.23
Claire	64	64.31	64.27	64.25
Crew	128	128.37	128.32	128.29
Carphone	128	128.60	128.53	128.47
Highway	128	129.71	129.43	129.27
Foreman	128	128.51	128.36	128.33
Claire	128	128.62	128.45	128.38
Crew	256	256.13	256.11	256.17
Carphone	256	256.34	256.27	256.21
Highway	256	257.07	256.89	256.77
Foreman	256	256.54	256.43	256.31
Claire	256	257.04	256.78	256.73
Crew	512	512.14	512.16	512.22
Carphone	512	512.25	512.21	512.16
Highway	512	512.34	512.24	512.21
Foreman	512	512.56	512.37	512.29
Claire	512	513.42	513.12	512.95

allocation more reasonable and improves the subjective image quality, which allocates the macroblock layer rate from the two angles of the image structure similarity and the close degree of the content. What is more, it improves the continuity of the video sequences. Compared with JM10.1 and PRDO, the video effect of the proposed algorithm is also better.

4. Conclusions

This paper presents a rate control algorithm for H.264 video coding by using the SSIM and the MAD to allocate the bit number of the macroblock, and then it performs some measurements to adjust the quantization parameter. The

TABLE 6: Bit rate comparison of the JM10.1, the PRDO, and the proposed algorithms (QP = 28).

Sequence	Target bit rate (kbps)	JM10.1 (kbps)	PRDO (kbps)	Proposed (kbps)
Crew	64	64.35	64.29	64.25
Carphone	64	64.38	64.31	64.29
Highway	64	64.96	64.72	64.57
Foreman	64	64.31	64.28	64.25
Claire	64	64.31	64.29	64.23
Crew	128	128.11	128.25	128.27
Carphone	128	128.57	128.36	128.31
Highway	128	129.30	128.98	128.79
Foreman	128	128.47	128.35	128.29
Claire	128	128.60	128.49	128.45
Crew	256	256.08	256.13	256.18
Carphone	256	256.44	256.31	256.19
Highway	256	257.53	257.12	256.88
Foreman	256	256.50	256.39	256.34
Claire	256	257.15	256.87	256.56
Crew	512	512.04	512.24	512.29
Carphone	512	512.07	512.05	512.13
Highway	512	512.21	512.19	512.17
Foreman	512	512.48	512.43	512.34
Claire	512	514.41	513.57	513.21

TABLE 7: Bit rate comparison of the JM10.1, the PRDO, and the proposed algorithms (QP = 33).

Sequence	Target bit rate (kbps)	JM10.1 (kbps)	PRDO (kbps)	Proposed (kbps)
Crew	64	64.35	64.29	64.31
Carphone	64	64.36	64.34	64.28
Highway	64	64.80	64.67	64.53
Foreman	64	64.26	64.31	64.25
Claire	64	64.34	64.43	64.32
Crew	128	128.10	128.17	128.25
Carphone	128	128.55	128.48	128.36
Highway	128	129.45	128.87	128.57
Foreman	128	128.50	128.46	128.41
Claire	128	128.58	128.44	128.39
Crew	256	256.10	256.31	256.24
Carphone	256	256.06	256.17	256.19
Highway	256	256.26	256.31	256.24
Foreman	256	256.46	256.35	256.21
Claire	256	257.16	256.88	256.63
Crew	512	512.04	512.21	512.26
Carphone	512	511.93	512.11	512.06
Highway	512	512.09	512.07	512.05
Foreman	512	512.38	512.28	512.26
Claire	512	504.10	512.25	512.23

FIGURE 3: The SSIM of carphone.

FIGURE 5: The PSNR of carphone.

FIGURE 4: The SSIM of foreman.

FIGURE 6: The PSNR of foreman.

experiment results show that the algorithm can improve the SSIM of the video sequence in the condition of ensuring the PSNR. So it can improve the subjective quality of the video sequence.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

This work was supported by the Qing Lan Project and the Priority Academic Program Development of Jiangsu Higher Education Institutions.

References

- [1] Zh. Li, F. Pan, K. P. Lim et al., "Adaptive basic unit layer rate control for JVT," in *Proceedings of the 7th Joint Video Team*

Meeting of ISO/IEC MPEG and ITU-T VCEG, Pattaya, Thailand, 2003.

- [2] Y. Lu, H. Li, Y. Zhang, and H. Cui, "Consistent perceptual quality rate control in H.264 region of interest video coding," *Journal of Tsinghua University*, vol. 49, no. 1, pp. 90–93, 2009.
- [3] Z. Han, K. Tang, and H. Cui, "Rate control method based on the standard H.264," *Journal of Tsinghua University*, vol. 48, no. 1, pp. 59–61, 2008.
- [4] X. Chen and F. Lu, "A reformative frame layer rate control algorithm for H.264," *IEEE Transactions on Consumer Electronics*, vol. 56, no. 4, pp. 2806–2811, 2010.
- [5] X. Liu, Y. Li, Y. Gao, and H. Zhang, "A coding strategy of H. 264 based on high- definition display of target region," *Computer Technology and Development*, vol. 20, no. 6, pp. 29–31, 2010.
- [6] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.
- [7] Z. Cui, Z. Gan, and X. Zhu, "Structural similarity optimal MB layer rate control for H.264," in *Proceedings of the International Conference on Wireless Communications and Signal Processing (WCSP '11)*, November 2011.
- [8] Z.-G. Cui and X.-C. Zhu, "Subjective rate-distortion performance improvement scheme for H.264 based on SSIM," *Journal of Electronics and Information Technology*, vol. 34, no. 2, pp. 433–439, 2012.
- [9] Z. Wang, G. Hu, J. Ming, and H. Wu, "Video quality assessment based on SSIM and ROI," *Chinese Journal of Scientific Instrument*, vol. 30, no. 9, pp. 1906–1911, 2009.
- [10] C.-L. Yang, H.-X. Wang, and R.-K. Liang, "Improved inter prediction based on structural similarity for H.264," *Chinese Journal of Computers*, vol. 32, no. 8, pp. 1603–1610, 2009.
- [11] C.-L. Yang and D.-Q. Xiao, "Improvements for H.264 intra mode selection based on SSE and SSIM," *Journal of Electronics and Information Technology*, vol. 33, no. 2, pp. 289–294, 2011.
- [12] Z. Cui and X. Zhu, "Subjective quality optimized intra mode selection for H.264 I frame coding based on SSIM," in *Proceedings of the 6th International Conference on Image and Graphics (ICIG '11)*, pp. 157–162, August 2011.
- [13] G. Wu, Y. Fu, and S. Chien, "System design of perceptual quality-regulable H. 264 video encoder," in *Proceedings of The IEEE International Conference on Multimedia and Expo*, pp. 509–514, 2012.
- [14] Z. Cui and X. Zhu, "SSIM-based content adaptive frame skipping for low bit rate H.264 video coding," in *Proceedings of the 12th International Conference on Communication Technology (ICCT '10)*, pp. 484–487, November 2010.
- [15] T.-S. Ou, Y.-H. Huang, and H. H. Chen, "SSIM-based perceptual rate control for video coding," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 21, no. 5, pp. 682–691, 2011.