LETTER

A Study on Consistency between MINAVE and MINMAX in SSIM Based Independent Perceptual Video Coding

Chao WANG[†a)], ***Student Member***, **Xuanqin MOU**[†], ***and*** **Lei ZHANG**[††], ***Nonmembers***

SUMMARY In this letter, we study the R-D properties of independent sources based on MSE and SSIM, and compare the bit allocation performance under the MINAVE and MINMAX criteria in video encoding. The results show that MINMAX has similar results in terms of average distortion with MINAVE by using SSIM, which illustrates the consistency between these two criteria in independent perceptual video coding. Further more, MINMAX results in lower quality fluctuation, which shows its advantage for perceptual video coding.

key words: *optimal bit allocation, MINAVE, MINMAX, perceptual video coding*

1. Introduction

The goal of optimal bit allocation in video encoding is to allocate the given quota of bits efficiently to different sources so that the best encoding performance can be achieved. Here, a source can be a group of pictures (GOP), a frame, or a sub image in a frame (e.g., basic unit (BU) in H.264/AVC [1]). The optimization of bit allocation among dependent sources is computationally complex and difficult to use in practical encoding [2]. Most optimization methods assume independent sources, e.g., GOPs in a sequence, or the BUs in an inter-frame [3]–[5].

Generally, the encoding results are evaluated by the average distortion or quality fluctuation of the sources. There are two commonly used optimization criteria for bit allocation problems: the minimum average criterion (MINAVE) and the minimum maximum criterion (MINMAX). MINAVE suggests minimizing the average distortion of the sources under the bits constraint, which can be solved with a constant multiplier by using the Lagrangian multiplier method [2]. Meanwhile, MINMAX suggests minimizing the maximum distortion of the sources under the bits constraint, which would result in uniform distortion for all sources [6].

Usually, the distortion metric used in optimal bit allocation is the mean square error (MSE) metric. However, it is well known that the MSE metric cannot faithfully reflect perceptual quality. Benefiting from the great progress in image quality assessment (IQA) studies in the past decade, perceptual based metrics such as SSIM [7] have been introduced into image/video encoding [4], [8], [9]. However, to the best of our knowledge, no work has compared the bit allocation performance of different optimization criteria when using perceptual distortion metrics, which motivated the research reported in this letter.

In this letter, we first discuss the properties of MINAVE and MINMAX based optimizations, and show that there is some consistency between their solutions when using perceptual based distortion metrics. We then validated this by encoding experiments. The results show that MINMAX has similar average distortion to MINAVE and lower quality fluctuation. This property suggests that MINMAX has the advantage over MINAVE for perceptual video coding, and should be used in practical encoding.

The rest of this letter is organized as follows. Section 2 briefly introduces the MINAVE and MINMAX criteria, and discusses their relationships when using MSE and SSIM metrics. Section 3 compares the encoding results on BU layer bit allocation. Finally, conclusions are drawn in Sect. 4.

2. MINAVE and MINMAX Criteria for Bit Allocation among Independent Sources

The purpose of optimal bit allocation is to find the best encoding parameters $X = \{x_1, x_2, \ldots, x_N\}$ for the constrained minimization problem:

$$\min_X f(X) \; s.t. \sum_{i=1}^{N} r_i(x_i) \le R_T, \tag{1}$$

where $f(X)$ is the objective function, N is the number of sources, $r_i(x_i)$ is the number of bits for the i^{th} source which is encoded by x_i, R_T is the bit constraint. In this letter, the encoding parameter x_i is the quantization parameter (QP).

In the MINAVE criterion, the objective function is:

$$f(X) = \sum_{i=1}^{N} d_i(x_i), \tag{2}$$

where $d_i(x_i)$ is the distortion of the i^{th} source. In the MINMAX criterion, the objective function is:

$$f(X) = \max_{i \in N} d_i(x_i). \tag{3}$$

2.1 The Optimal Bit Allocation in Independent Encoding

The MINAVE based constrained minimization problem can

Manuscript received December 24, 2014.
Manuscript revised February 12, 2015.
Manuscript publicized April 13, 2015.

[†]The authors are with the Institute of Image Processing and Pattern Recognition, Xi'an Jiaotong University, Xi'an, China.
[††]The author is with the Dept. of Computing, The Hong Kong Polytechnic University, Hong Kong, China.
a) E-mail: cwang.2007@stu.xjtu.edu.cn

DOI: 10.1587/transinf.2014EDL8253

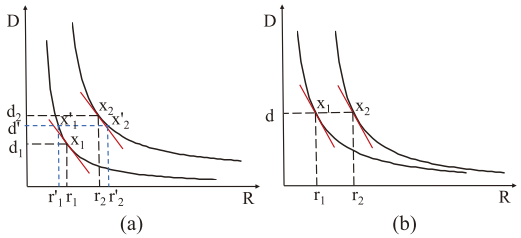

Fig. 1 An example of the bit allocation between two independent sources with rate constraint R_T. (a) The ideal solutions of MINAVE and MINMAX based optimization, respectively. (b) The situation when MINAVE and MINMAX have the same solution.

be converted into a non-constrained problem by using the Lagrange multiplier method with a constant multiplier λ [2]:

$$\min_X J(X) = \min_X \left(\sum_{i=1}^{N} d_i(x_i) + \lambda \cdot \sum_{i=1}^{N} r_i(x_i) \right) \tag{4}$$

$$= \sum_{i=1}^{N} \min_{x_i} \left(d_i(x_i) + \lambda \cdot r_i(x_i) \right), \tag{5}$$

where $J(X)$ is the Lagrangian cost function, and λ is determined by R_T. Equation (5) is derived from independent assumption. Geometrically, the solution of independent MINAVE problem is to find for each source a point (x_i) on its R-D curve, at where the tangent has a fixed slope of λ, and the total number of bits corresponding to these points equals to R_T.

In MINMAX based optimization, the maximum distortion of the sources is minimized, and thus the number of bits for the source which has the maximum distortion should be increased. Due to the bits constraint, the bits allocated to the other sources should then be decreased, and this would increase their distortions. As a result, the ideal balance point is that the one where all the sources have a constant distortion. Due to the discreteness of the QPs, constant distortion solution may not exist, the MINMAX method usually achieves the solution with the minimum quality fluctuation.

Figure 1 (a) shows the ideal solutions of MINAVE (x_1 and x_2 with constant slope λ) and MINMAX (x'_1 and x'_2 with constant distortion) by taking an example of bit allocation between two independent sources with bit constraint R_T, i.e., $r_1 + r_2 = r'_1 + r'_2 = R_T$. In some special case, at any given distortion d, if the corresponding slopes λ on all the R-D curves are identical to each other, the solutions of MINAVE and MINMAX based optimizations will be the same. This case correspond to that all the R-D curves are overlapped or parallel along the R-axis, as illustrated in Fig. 1 (b).

2.2 R-D Properties under Different Distortion Metrics

Since the sources have different contents from each other, their MSE based R-D properties vary a lot. Generally, a source with complex contents has higher MSE than a source with simple contents when encoded at similar bit rate. Thus,

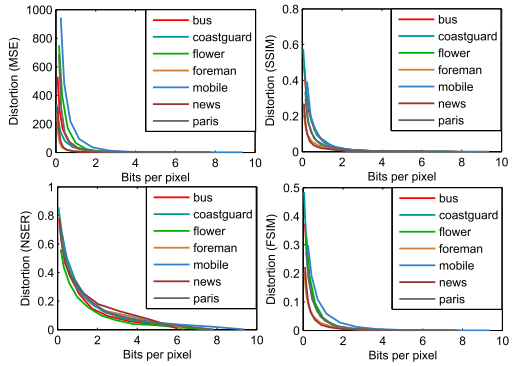

Fig. 2 The R-D curves of 7 intra frames with different contents. The distortions are measured by (from left to right and top to bottom): MSE, SSIM, NSER, and FSIM, respectively.

Table 1 Test sequences.

ID	Sequences	ID	Sequences
1	*bus@CIF*	7	*container@CIF*
2	*foreman@CIF*	8	*walk@CIF*
3	*coastguard@CIF*	9	*intotree@720P*
4	*paris@CIF*	10	*mobcal@720P*
5	*flower@CIF*	11	*parkjoy@720P*
6	*twoducks@CIF*	12	*stockholm@720P*

the R-D curve of complex source will lie above that of simple source. However, based on the error masking effect of HVS [10], the complex sources can tolerate more encoding error than the simple ones. This will reduce the perceptual distortion of the complex sources and do the opposite to the simple ones. As a result, when using perceptual based IQA metrics, the R-D curves of different sources will be closer to each other or even overlapped.

Figure 2 shows the R-D curves of 7 frames with different contents which are selected from the sequences in Table 1 and encoded by the H.264 intra frame encoder at $QP = \{10, 15, 20, 25, 30\}$, respectively. The distortions are measured by MSE and three perceptual based IQA metrics: SSIM [7], NSER [11], and FSIM [12]. Here, the perceptual distortion metrics are defined as:

$$D(x, y) = 1 - Q(x, y), \tag{6}$$

where x and y are the original and the reconstructed frames, respectively, and $Q(x, y)$ is the quality metric function of a specific IQA index. It is clear that, the R-D curves are more overlapped by using the perceptual distortion metrics, especially at moderate and high bit rates, while the MSE based R-D curves vary greatly between frames.

By encoding a frame with a QP, we can get a data pair (d, λ), where d is the distortion of the frame and λ is the slope of the tangent at the corresponding point on the R-D curve. We draw the data pairs (d, λ) of all the encoded frames in Fig. 3. MSE and SSIM metrics are used here. The SSIM based distortion metric is denoted as DSSIM hereafter. The slope λ of the i^{th} frame at QP x_i is calculated by:

$$\lambda_i(x_i) = \frac{1}{2} \left(\frac{d_i(x_i) - d_i(x_i - t)}{r_i(x_i - t) - r_i(x_i)} + \frac{d_i(x_i) - d_i(x_i + t)}{r_i(x_i + t) - r_i(x_i)} \right), \tag{7}$$

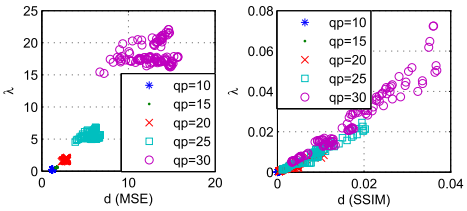

Fig. 3 The distribution of (d, λ) of independent frames under different distortion metrics MSE (left) and DSSIM (right).

where t is an integer that is set as 5 here.

We can see from Fig. 3 that, the (d, λ) points under DSSIM locate in a narrow band, which indicate that the corresponding R-D slopes are nearly the same under a given distortion. Recall the discussion in Sect. 2.1, these findings illustrate that by using perceptual distortion metrics, the output of MINMAX is consistent with MINAVE. However, this property does not hold for MSE. We will validate this conclusion by practical encoding in next section.

3. Experiments and Results

In this section, we compare the encoding performance of MINAVE and MINMAX based optimal methods on BU layer bit allocation in P frames. Both MSE and DSSIM metrics are used for comparing. The H.264/AVC baseline encoder [1] is used here, in which the encoding of BUs inside a P frame are independent of each other. The sequences used for comparison are listed in Table 1. In the encoding, each GOP has 15 frames, with an I frame followed by all P frames. Each sequence was encoded by three methods: the fixed QP method (which encodes all the BUs in a frame with a fixed QP), and the MINAVE and MINMAX based optimal bit allocation methods. In the two optimal methods, the I frames were encoded by the same QP used by the fixed QP method. The QPs for the BUs in each P frames for the two optimal methods were selected optimally by a full search method based on MINAVE and MINMAX criteria, respectively, constrained by using the same number of bits that used by the fixed QP method. In all the three methods, the reference frame number was set to 1 for simplicity. In the CIF sequences, a BU is constituted by a group of 2x2 MBs, and in the 720P sequences, it is a group of 5x5 MBs. The encoding performance was compared in terms of average distortion and quality fluctuation, respectively.

3.1 R-D Performance and QP Assignment

The encoding results of *bus* based on MSE and DSSIM metrics are shown in Figs. 4 and 5, respectively, where the QPs used by the fixed QP method are {10, 15, 20, 25, 30}. In the figures, $Mean(D)$ states for the average distortion of all the BUs, and $Std(D)$ states for the mean standard deviation of distortion among BUs in each frame. More details about the results are given in Fig. 6, which shows the QP assignments for the BUs by different methods when the QP used by the fixed QP method is 20.

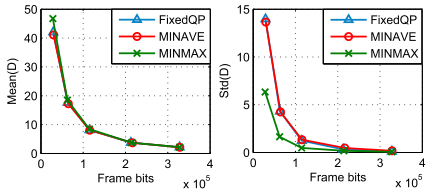

Fig. 4 The BU layer bit allocation results of *bus*. Left: the average MSE of BUs. Right: the standard deviation of MSE of BUs.

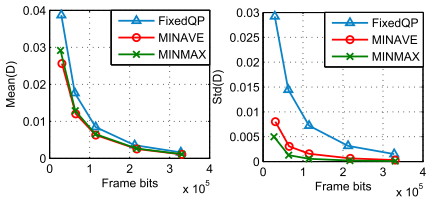

Fig. 5 The BU layer bit allocation results of *bus*. Left: the average DSSIM of BUs. Right: the standard deviation of DSSIM of BUs.

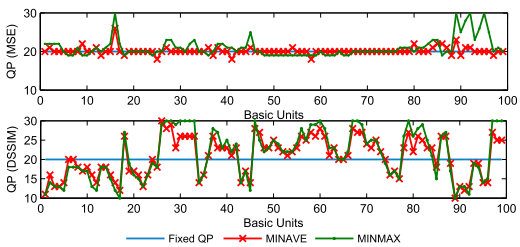

Fig. 6 The QP assignment for the BUs in the 1^{st} P frame of *bus* by using different distortion metrics. Top: MSE. Bottom: DSSIM.

From the results we have the following findings:

1) When using MSE (Fig. 4), the fixed QP method has similar results to the MINAVE method in terms of both $Mean(D)$ and $Std(D)$. The results of QP assignments (Fig. 6) show that the MINAVE method chooses nearly constant QP for BUs. This is because there is an approximately fixed relationship between QP and Lagrangian multiplier λ under MSE [13], and thus the fixed QP method corresponds to a constant λ method, which is actually MINAVE in independent encoding.

2) Under MSE (Fig. 4), the results of MINMAX has almost similar $Mean(D)$ and much lower $Std(D)$ than MINAVE. The QP assignments (Fig. 6) show that the quantization schemes are very different between these two methods. This is due to the difference on the R-D properties between BUs under MSE, as discussed in Sect. 2. The results imply that there are multi solutions for bit allocation under the same bit constraint with similar $Mean(D)$ but quite different values on $Std(D)$.

3) When using DSSIM (Fig. 5), the MINAVE method has much better results than the fixed QP method. This is because that the fixed relationship between QP and λ does not exist under IQA metrics. Besides, the MINAVE and MINMAX methods have similar results, which can be seen clearly in Fig. 6. Since the R-D properties of different con-

tents become closer under IQA metrics (refer to Sect. 2), the MINAVE and MINMAX methods choose more similar solutions.

4) Though the MINMAX method has similar $Mean(D)$ results to MINAVE with DSSIM, it has lower $Std(D)$, which demonstrates that the MINMAX method has advantage in perceptual (SSIM based) video encoding.

3.2 Numeric Comparison

In order to compare the encoding performance of different methods more directly, we convert the curves in Figs. 4 and 5 into numeric values. We use a method which is similar to that proposed in [14]. Taking the $bits - Mean(D)$ curve for example, we denote the bits number and $Mean(D)$ of the M data points on a specific curve by $R = \{r_1, r_2, \ldots, r_M\}$ and $D = \{d_1, d_2, \ldots, d_M\}$, respectively. The integration interval of R is $T = \{r_1, r_2 - r_1, \ldots, r_M - r_{M-1}\}$. Then, the value $Mean(D)$ is calculated by:

$$Mean(D) = \frac{\sum_{i=1}^{N} d_i \cdot t_i}{\sum_{i=1}^{N} t_i}, \tag{8}$$

where t_i is the i^{th} item of T. The value of $bits - Std(D)$ curves can be calculated similarly.

Results of the test sequences in Table 1 based on MSE and DSSIM are shown in Figs. 7 and 8, respectively. All the results have the same properties as we discussed before.

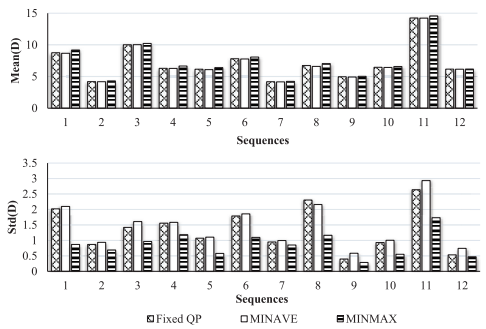

Fig. 7 Numeric comparison of the encoding performance on BU layer bit allocation based on MSE. Top: $Mean(D)$. Bottom: $Std(D)$.

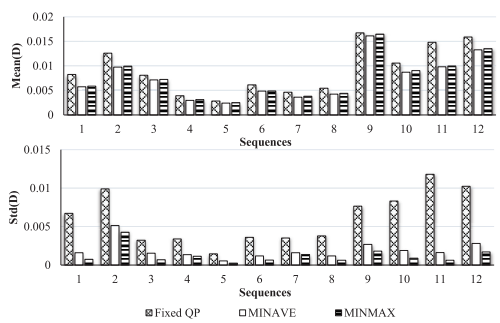

Fig. 8 Numeric comparison of the encoding performance on BU layer bit allocation based on DSSIM. Top: $Mean(D)$. Bottom: $Std(D)$.

4. Conclusions

In this letter, we studied the problem of optimal bit allocation among independent sources in perceptual video coding. Bit allocation on the BU layer was tested. Our experiments showed that by using SSIM, the MINMAX based optimization has similar average distortion to MINAVE, which demonstrates the consistency between the two criteria in perceptual video coding. The results also show that the MINMAX method achieves lower quality fluctuation, which shows the advantage of MINMAX. We suggest that the MINMAX criterion should be used to optimize the bit allocation among independent sources in perceptual video encoding.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China under Grant 61172163 and Grant 90920003, and in part by the Research Grants Council, Hong Kong, under Grant PolyU-5315/12E.

References

[1] T. Wiegand, G.J. Sullivan, G. Bjontegaard, and A. Luthra, "Overview of the H.264/AVC video coding standard," IEEE Trans. Circuits Syst. Video Technol., vol.13, no.7, pp.560–576, 2003.
[2] A. Ortega and K. Ramchandran, "Rate-distortion methods for image and video compression," IEEE Signal Process. Mag., vol.15, no.6, pp.23–50, 1998.
[3] Y. Shoham and A. Gersho, "Efficient bit allocation for an arbitrary set of quantizers," IEEE Trans. Acoust., Speech, Signal Process., vol.36, no.9, pp.1445–1453, 1988.
[4] T.-S. Ou, Y.-H. Huang, and H.H. Chen, "SSIM-based perceptual rate control for video coding," IEEE Trans. Circuits Syst. Video Technol., vol.21, no.5, pp.682–691, 2011.
[5] D.-K. Kwon, M.-Y. Shen, and C.-C.J. Kuo, "Rate control for H.264 video with enhanced rate and distortion models," IEEE Trans. Circuits Syst. Video Technol., vol.17, no.5, pp.517–529, 2007.
[6] G.M. Schuster, G. Melnikov, and A.K. Katsaggelos, "A review of the minimum maximum criterion for optimal bit allocation among dependent quantizers," IEEE Trans. Multimedia, vol.1, no.1, pp.3–17, 1999.
[7] Z. Wang, A.C. Bovik, H.R. Sheikh, and E.P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," IEEE Trans. Image Process., vol.13, no.4, pp.600–612, 2004.
[8] S.Q. Wang, A. Rehman, Z. Wang, S.W. Ma, and W. Gao, "Perceptual video coding based on SSIM-inspired divisive normalization," IEEE Trans. Image Process., vol.22, no.4, pp.1418–1429, 2013.
[9] C. Wang, X.Q. Mou, W. Hong, and L. Zhang, "Block-layer bit allocation for quality constrained video encoding based on constant perceptual quality," Proc. SPIE Electronic Imaging, 2013.
[10] A.B. Watson and J.A. Solomon, "Model of visual contrast gain control and pattern masking," Journal of the Optical Society of America A—Optics Image Science and Vision, vol.14, no.9, pp.2379–2391, 1997.
[11] M. Zhang, X.Q. Mou, and L. Zhang, "Non-shift edge based ratio (NSER): An image quality assessment metric based on early vision features," IEEE Signal Process. Lett., vol.18, no.5, pp.315–318, 2011.
[12] L. Zhang, X.Q. Mou, and D. Zhang, "FSIM: A feature similarity index for image quality assessment," IEEE Trans. Image Process.,

vol.20, no.8, pp.2378–2386, 2011.

[13] G.J. Sullivan and T. Wiegand, "Rate-distortion optimization for video compression," IEEE Signal Process. Mag., vol.15, no.6, pp.74–90, 1998.

[14] G. Bjontegaard, "Calculation of average PSNR differences between RD curves," Proc. 13th Meeting ITU-T Q.6/SG16VCEG, 2001.