A ROI QUALITY ADJUSTABLE RATE CONTROL SCHEME FOR LOW BITRATE VIDEO CODING

Ling Yang[1], *Li Zhang*[2], *Siwei Ma*[3], *Debin Zhao*[4]
{yanglin, zhanglili, swma, dbzhao}@jdl.ac.cn
[1] Graduate University, Chinese Academy of Sciences, Beijing
[2]Key Lab of Intelligent Information Processing, Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China
[3] Institute of Digital Media, School of Electronic Engineering and Computer Science, Peking University, Beijing, China
[4] Department of Computer Science and Engineering, Harbin Institute of Technology, Harbin, China

ABSTRACT

This paper proposed a novel Region of Interest (ROI) based rate control algorithm for video communication systems, in which the subjective quality of ROI can be adjusted according to users' requirements. In the proposed scheme, in order to analyze the relationship between subjective quality and encoding parameters, a Structural Similarity Index Map – quantization parameter (SSIM-QP) model is established. Through this relation, the possible visual quality range of ROI is defined according to the range of ROI QP, which is predicted by rate control algorithm. Then, with interest levels being identified within the visual quality range, resources allocated to ROI is determined. Finally, considering both the quality of ROI and the entire frame, resource allocation is slightly adjusted. Experimental results demonstrate that the subjective visual quality of ROI in our proposed scheme can be adjustable improved compared to the existing rate control algorithms.

***Index Terms*—** region of interest, rate control, SSIM

1. INTRODUCTION

Recently, with the rapid demands for Region of Interest (ROI) in the extensive application of video conferencing, videophone, video monitoring, surveillance and telemedicine, ROI-based rate control has become a compelling topic in the past decade. On these occasions, users pay much more attention on the interested area, while less attention for the other area, named non-ROI. Therefore, the scalable enhancement of visual quality (VQ) of ROI according to its importance defined by the user as well as the maintenance of the visual quality of the entire frame has become a major concern, especially in low bit rate condition. Several researches have been dedicated in improving video coding with ROI. In [1-2], weights are put upon distortion mean square error (MSE) and bit budgets to improve visual quality of important areas. In [3] of Ping-Hao Wu *et al*, a "base encoder" is used for coding full view version at low resolution and another "region of interest encoder" is used for ROI at high resolution. The bits for ROI encoder is determined according to the distortion obtained from the corresponding region in the base encoder. In [4] of Yang Liu *et al*, ROI is first detected using both direct frame differences and skin-tone information. Next, coding resources are adaptively adjusted according to the importance of each MB.

In fact, for scalable adjustment of visual quality of ROI, it is very necessary to employ visual quality criteria as the compass to direct bit allocation scheme between ROI and non-ROI area. The prevalent works employ peak to signal noise ratio (PSNR) to evaluate the performance of rate control algorithms. However, such measurement may not in good agreement with human visual quality judgment. The widely commended structural similarity (SSIM) metric of Wang *et al*. [5] is chosen for such purpose. The metric operates based on the notion that the human vision system has evolved to extract structural information from natural images, and therefore, a high-quality images is the one whose structural most closely matches that of the original. Such measurement has received wide agreement on its effectiveness and is now implemented in X264 codec for VQ assessment.

In this paper, the relation between QP and SSIM is established as a facility for visual quality assessment and prediction. Then, the maximum and minimum visual quality of the ROI region is estimated according to current bandwidth, so that subjective quality ranks with equal visual quality difference can be identified. And an adaptive bit allocation scheme will be given to meet the interest level designated by the user.

The rest of the paper is organized as follows. In section 2, the relationship between QP and SSIM is analyzed. In section 3, the ROI visual quality adjustable rate control scheme is presented. Experimental results are given in section 4. Section 5 concludes the whole paper.

2. ANALYSIS OF SSIM AND QUANTIZATION PARAMETERS RELATION

Meanwhile, subjective image and video quality assessment methods have been extensively studied, and many models were established [5-6]. The chosen SSIM metric [5] follows the philosophy that 'human visual system is highly adaptive in extracting structural information'. Therefore, it employs a modified measure of spatial correlation between the pixels of the original and distorted images to quantify the extent to which the image's structure has been distorted.

Therefore, the frames of reconstructed sequences can be compared with the original ones for visual quality evaluation, computed as follows:

$$SSIM(x,y) = l(x,y) \bullet c(x,y) \bullet s(x,y) \tag{1}$$

where $l(x,y)=\dfrac{2\mu_x\mu_y+C_1}{\mu_x^2+\mu_y^2+C_1}$, $c(x,y)=\dfrac{2\sigma_x\sigma_y+C_2}{\sigma_x^2+\sigma_y^2+C_2}$, $s(x,y)=\dfrac{\sigma_{xy}+C_3}{\sigma_x\sigma_y+C_3}$, x and y are the image signals of the original and reconstructed frame, μ_x and μ_y are the mean luminance of x and y respectively, σ_x and σ_y are the standard deviation of x and y respectively, σ_{xy} is the covariance of x and y, C_1, C_2 and C_3 are three small constants.

To predict subjective quality at various encoding configuration, a uniform relation between visual quality and quantization of DCT coefficient is exploited. The visual quality is measured using SSIM metric.

10 CIF sequences are used in building this SSIM-QP model, including '*Akiyo*', '*deadline*', '*Paris*', '*mthr_dotr*', '*foreman*', '*missA*', '*salesman*', '*carphone*', '*students*', '*news*'. The experimental quantization parameter ranges from 4 to 51, with interval 4. The resulting points are shown in Fig. 1.

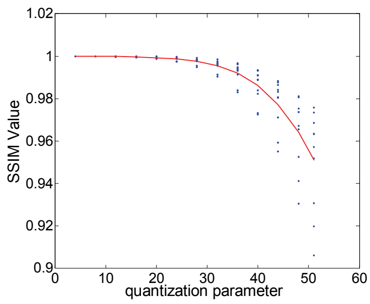


Fig. 1 SSIM-QP estimation curve

Considering fitting accuracy and computational complexity, the following function is selected to simulate the SSIM-QP relation,

$$S(QP)= aQP^4+ bQP^3+ cQP^2+ dQP+ e \tag{2}$$

where $a=-2.3351\times10^{-8}$, $b=1.3105\times10^{-6}$, $c=-2.9737\times10^{-5}$, $d=0.0002608$, $e=0.99924$.

From (2), the visual quality of the encoding result by a certain QP can be predicted. Conversely, the quantization parameter corresponds to a target visual quality SSIM can also be determined.

Such SSIM-QP relation is used to properly allocate bits to different areas of the frame. From Fig. 3 we can observe that when QP is smaller than 30, more bits expenditure will not render with equivalent much visual quality enhancement, which means, when the quantization parameter of ROI is beneath a certain threshold, it is more beneficial to maintain visual quality of ROI at the threshold level and allocate the rest bits on the non-ROI region, so that the visual quality of the entire frame will be improved.

3. ROI BASED SCALABLE VISUAL QUALITY VIDEO CODING SCHEME

To balance the bits allocation between ROI region and non-ROI region according to users' interest, the framework of encoding one frame is shown in Fig.2.

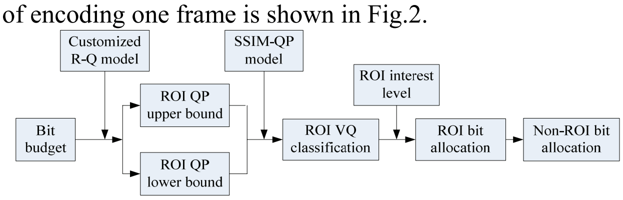


Fig. 2 The framework of the ROI based scalable visual quality bit allocation scheme

In the proposed scheme, the visual quality of ROI can be adjusted according to users' interest level. The R-Q model is customized with the fact that the texture content and visual quality of ROI and non-ROI are different to achieve more accurate rate control. According to the customized R-Q model, the upper and lower bound of ROI QP, QP_{ROIMAX} and QP_{ROIMIN}, are obtained. Then, utilizing the SSIM-QP model established in the former section, the possible visual quality range is estimated and equal difference visual quality ranks of ROI are given. Therefore, the target subjective quality for ROI, also measured in SSIM metric, can be set according to the user's interest. With the target SSIM, quantization parameter for encoding ROI is first predicted and then adjusted to further save bits for non-ROI region to enhance the entire frame quality. After ROI encoding is done, the rest bits are allocated for non-ROI region by FMO technique [7], and its encoding parameters can also be derived using customized R-Q model.

3.1. ROI Customized R-Q Model

The current H.264/AVC standard utilize a quadratic rate-distortion model [8][9] to calculate the corresponding quantization parameter of a given bit budget. For video coding with ROI, an improved R-VQ performance can be expected if mean absolute difference (MAD), is customized applied with different scene content and regions of different VQ in a sequence. Therefore, the MAD of ROI and non-ROI is predicted separately. And the R-D model is also customized.

Assuming in the ith frame, the bit budget of the entire frame is R_i, and the bit budget for ROI is $R_{R,i}$, the actual bits that encodes ROI is $R_{RA,i}$, and the bits for the non-ROI region is R_i-$R_{RA,i}$, the following equations are derived.

$$\frac{R_{R,i}}{MAD_{R,i}}=\frac{c_{11}}{Q_{R,i}}+\frac{c_{12}}{Q_{R,i}^2} \tag{3}$$

$$\frac{R_i-R_{RA,i}}{MAD_{N,i}}=\frac{c_{21}}{Q_{N,i}}+\times\frac{c_{22}}{Q_{N,i}^2} \tag{4}$$

where $Q_{R,i}$ is the quantization step of ROI area, and $Q_{N,i}$ is the quantization step of the non-ROI area, c_{11} and c_{12} are coefficients that could be updated after encoding of ROI and c_{21} and c_{22} are coefficients that could be updated after encoding of non-ROI. And $MAD_{R,i}$ and $MAD_{N,i}$ are the MAD of ROI and non-ROI respectively, which predicted by the linear model [10].

$$MAD_{R,i} = a_{11} \times MAD_{R,i-1} + a_{12} \quad (5)$$

$$MAD_{N,i} = a_{21} \times MAD_{N,i-1} + a_{22} \quad (6)$$

3.2 Lower and Upper Bound of ROI Quantization Parameter

In the proposed scheme, the interest level of ROI can be flexibly defined in several ranks according to the actual need.

The minimum interest level means ROI will be regarded as the same as non-ROI region

$$QP_{ROIMAX} = QP_F, \text{ when } I=1 \quad (7)$$

where QP_F denotes the frame QP obtained from frame layer rate control regardless of ROI and I denotes the interest level designated by the user.

On the other hand, the visual quality of the ROI area cannot be maximized unlimitedly. Since the purpose of rate control is to maintain consistency buffer status, when proper bits are allocated to a frame by bit allocation scheme, we must adjust the encoder parameter to satisfy the bit budget. In the proposed scheme, an extreme condition is that the entire bit budget for one frame is used solely on ROI. At this stage, it reaches to the minimum possible encode parameter for ROI region QP_{ROIMIN}.

$$\frac{R_i}{MAD_{R,i}} = \frac{c_{11}}{Q_{ROIMIN,i}} + \frac{c_{12}}{Q^2_{ROIMIN,i}} \quad (8)$$

where Q_{ROIMIN} is the minimum quantization step of the QP for ROI on the highest user interest level.

$$QP_{ROIMIN} = 6 \times \log_2 Q_{ROIMIN,i} + 4 \quad (9)$$

3.3. Bit Allocation among ROI and non-ROI

In applications such as surveillance system, video conference and so on, ROI region has priority and its quality must be ensured. Therefore the proposed scheme performs flexible macroblock order to encode ROI region prior to non-ROI region. Bit budget on ROI is set according the interest level, and also adaptively save bits for non-ROI region when surplus bits are not necessary for ROI region. Detail is described in as follows.

Assume I_{MAX} is the ceiling interest level choice. Rank 1 and Rank I_{MAX} determine the two ends of the possible subjective quality interval. Assume they are S_1 and $S_{I_{MAX}}$, respectively.

$$S_1 = S(QP_{ROIMAX}) \quad (10)$$

$$S_{I_{MAX}} = S(QP_{ROIMIN}) \quad (11)$$

To achieve equal increment in visual quality as the ranking increase, the target subjective quality of each level is defined as:

$$S_I = (S_{I_{MAX}} - S_1)\frac{I}{I_{MAX} - 1} + S_1 \quad (12)$$

where I is the interest level of the ROI region designated by the user.

Then, the quantization parameter to achieve target visual quality S_I can be estimated by (2).

On a special condition when the interest level is chosen as I_{MAX}, there is a possibility to save bits for non-ROI region. As depicted in Fig. 3, if QP_{ROI} is below 30, the subjective visual quality will not increase much. Therefore, the scheme adaptively adjusts ROI QP to maintain the ROI quality within the QP=30 level and allocate the rest bits on non-ROI region to enhance the entire frame quality.

A demonstration of ROI QP determination is shown in Fig. 3.

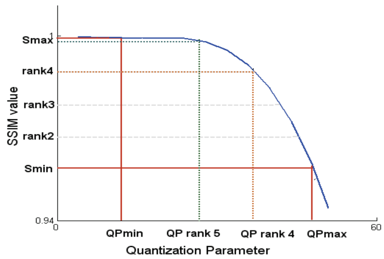


Fig. 3 Demonstration of ROI QP determination when the ceiling interest level is set to 5. The orange dot line denotes the condition when interest level = 4. The green dot line denotes the condition when interest level =5.

To avoid the extreme condition when all bits are used in ROI region and the quality of non-ROI will be unacceptably poor, if the interest level is set as the ceiling level and it is predicted that QP of non-ROI has exceed 51. The QP of ROI will be subtracted by 2.

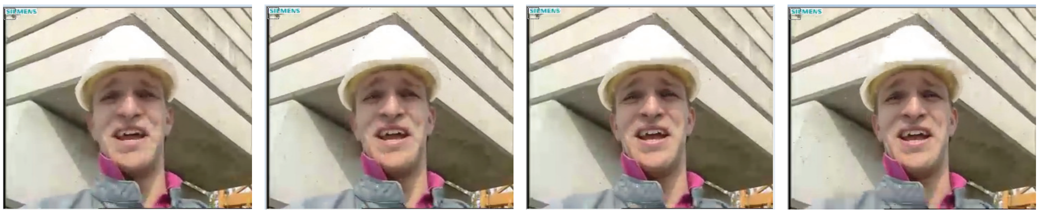
After the important ROI is encoded, the non-ROI region will be encoded with parameters in (4) and (6).

4. EXPERIMENTAL RESULT

The performance of the scalable ROI-quality bit allocation scheme is evaluated in this section. Several standard test sequences with CIF (352×288) spatial resolution are carried out in the experiments. The simulation is implemented with the JVT reference H.264/AVC software JM10.0. The sequences are encoded in IPPP structure, with GOP size equal to 30. The search range for ME is set to 32 pixels and increased to 64 in coding the left and up margin MBs of ROI. CABAC and RDO are enabled. The number of referencing frame is set to 3. Quarter-sample-accuracy ME is enabled. The maximum interest level is set to 4.

An example of subjective visual quality comparison is shown in Fig. 4. It could be observed that the subjective quality of ROI is obviously improved. The visual quality of the whole frame and ROI region are also compared, in terms of SSIM value.

Simulation results show that our proposed encoder can adjust subjective visual quality of ROI in relatively linear steps. The visual quality of ROI is improved and the visual quality of the entire frame remains acceptable.



(a) (b) (c) (d)

Fig. 4 various ROI quality when the maximum interest level is set to 4 when bit-rate is 128kb/s. (a)Interest level =1 (b)Interest level =2, (c)Interest level =3 (d)Interest level =4 (ROI area is set as foreman's face)

Table 1 The SSIM results of coded sequences with different bitrate

Sequence	Target bitrate (kb/s)		Original	Proposed Level 1	Proposed Level 2	Proposed Level 3	Proposed Level 4
Foreman	128	Actual bitrate (kb/s)	129.97	129.68	130.35	130.83	131.52
		Overall SSIM	0.9940	0.9939	0.9911	0.9843	0.9760
		ROI SSIM	0.9940	0.9940	0.9948	0.9959	0.9968
	256	Actual bitrate (kb/s)	256.63	256.91	257.42	257.36	257.65
		Overall SSIM	0.9971	0.9971	0.9969	0.9968	0.9969
		ROI SSIM	0.9971	0.9971	0.9969	0.9970	0.9970
news	128	Actual bitrate (kb/s)	129.28	128.58	129.41	129.85	131.53
		Overall SSIM	0.9969	0.9969	0.9958	0.9932	0.9882
		ROI SSIM	0.9969	0.9970	0.9974	0.9979	0.9984
students	128	Actual bitrate (kb/s)	128.47	128.84	130.12	129.05	132.19
		Overall SSIM	0.9952	0.9952	0.9926	0.9912	0.9867
		ROI SSIM	0.9952	0.9952	0.9964	0.9973	0.9979

5. CONCLUSIONS

This paper proposed a ROI quality adjustable rate control scheme for low bit-rate video coding. The scheme achieves visual quality ranks with equal intervals by establishing a SSIM-QP model to direct bit allocation.

To improve the proposed scheme, optimization can be used in balancing bit allocation between ROI and non-ROI, which will be our further study direction.

6. ACKNOWLEDGEMENT

This work is supported by grants from National Basic Research Program of China under contract No. 2009CB320907.

7. REFERENCES

[1] Chi-Wah Wong, Oscar C. Au, Bojun Meng, Hong-Kwai Lam "Perceptual Rate Control for Low-Delay Video Communications", *International Conference on Multimedia & Expo,* pp.361-364, July, 2003

[2] Hang Li, Zhibing Wang, Huijuan Cui, Kun Tang, "An Improved ROI-Based Rate Control Algorithm for H.264/AVC", *International Conference on Software Process,* 2006

[3] Ping-Hao Wu, Homer H. Chen, "Frame-Layer Constant-Quality Rate control of Region of Interest for Multiple Encoders With Single Video Source", *IEEE Trans. Circuit Syst. Video Technology*, pp. 857-867, July, 2007.

[4] Yang Liu, Zheng Guo Li, Yeng Chai Soh, "Region-of-Interest based Resource Allocation for Conversational Video Communication of H.264/AVC", *IEEE Trans. Circuit Syst. Video Technology*, pp. 134-139, vol. 18, No.1, January, 2008.

[5] Zhou Wang, Liguang Lu, Alan C. Bovik, "Video quality assessment based on structural distortion measurement," *Signal Processing: Image Communication*, 19(2004) 121-132

[6] X.K. Yang, W.S. Ling, Z.K. Lu, E.P. Ong, S.S. Yao, "Just noticeable distortion model and its applications in video coding" *Signal Processing: Image Communication* 20 (2005) 662–680

[7] Stephan Wenger, Michael Horowitz, FMO: Flexible Macroblock Ordering, Joint Video Team (JVT) of ISO/IEC MPEG & ITU-T VCEG (ISO/IEC JTC1/SC29/WG11 and ITU-T SG16 Q.6) Doc. JVT-C089, Fairfax, Virginia, USA, 6-10 May, 2002

[8] Siwei Ma, Zhengguo Li, Feng Wu, Proposed draft of Adaptive rate control, Joint Video Team (JVT) of ISO/IEC MPEG & ITU-T VCEG(ISO/IEC JTC1/SC29/WG11 and ITU-T SG16 Q.6),8th Meeting: Geneva, May 20-26,2003

[9] Yang Liu, Zhengguo Li, Yeng Chai Soh, "A novel rate control scheme for low delay video communication of H.264/AVC standard", *IEEE Trans. On Circuits and Systems for Video Technology*, Vol.17, no. 1, January 2007

[10] Z. G. Li, F. Pan, K. P. Lim, G. N. Feng, X. Lin, and S. Rahardaj, Adaptive basic unit layer rate control for JVT, Joint Video Team of ISO/IEC JTC1/SC29/WG11 and ITU-T SG16/Q.6 Doc. JVT-G012, Pattaya, Thailand, Mar. 2003.