Perceptual Motivated Coding Strategy for Quality Consistency

Like Yu[1,2], Feng Dai[1], Yongdong Zhang[1], and Shouxun Lin[1]

[1] Institute of Computing Technology, Chinese Academy of Sciences, 100190 Beijing, China
[2] Graduate University of Chinese Academy of Sciences, 100190 Beijing, China
{yulike,fdai,zhyd,sxlin}@ict.ac.cn

Abstract. In this paper, we propose a novel quality control scheme which aims to keep quality consistency within a frame. Quality consistency is an important requirement in video coding. However, many existing schemes usually consider the quality consistency as the quantization parameter (QP) consistency. Moreover, the most frequently used metric to evaluate the quality consistency is PSNR, which has been well known that it is not good for subjective quality evaluation. These flaws of the existing methods are pointed out and proved to be unreasonable. For optimization, we take the effect of texture complexity on subjective evaluation into consideration to build a new D-Q model. We use the new model to adjust the quantization parameters of different regions to keep quality consistency. The simulation result shows that the new scheme gets better subjective quality and higher coding efficiency compared to traditional way.

Keywords: quality consistency, quality fluctuation, video coding, H.264/AVC, perceptual quality.

1 Introduction

Quality control is one of the most important requirements in video coding. In most cases, we demand low quality fluctuation within a frame and between two successive frames [1] because it plays a major negative role on visual perception.

Many schemes have been proposed to achieve constant quality [2-4]. They usually pay attention to the quality fluctuation between frames. The quality difference within a frame is less concerned. The purpose of fluctuation limitation is to provide better perceptual visual feelings. Therefore not only the quality variance between frames should be limited, a uniform visual quality within a frame is also important.

In most cases, researchers restrict the QP (quantization parameter) variation in order to achieve consistent quality. In [5], the authors assign fixed QP to all macroblocks (MBs) in a frame to maintain consistent picture quality. In JVT-G012 [6], which has been adopted in H.264/AVC reference software, the QP variation between each basic unit is limited in order to constrain quality fluctuation. However, in [7] it has been pointed out that coding with a constant value of *Qs* (quantization step) generally does not result in either constant bit-rate or constant perceived quality.

K.-T. Lee et al. (Eds.): MMM 2011, Part I, LNCS 6523, pp. 21–28, 2011.

In this paper, we propose a new quality control scheme which partitions a frame into several regions and assigns each of them different quantization steps to decrease quality fluctuation within a frame. Different from the existing methods, we abandon the PSNR quality metric because we demand uniform perceptual quality and PSNR is not a good metric for subjective evaluation among different image content. We adopt a "weighted MSE" metric [8], to help control the quality consistency. Based on the new metric, a new D-Q (Distortion-Quantization) model is established. Then different quantization parameters are assigned to different regions according to the new D-Q model and the quality consistency will be optimized. Moreover, the overall quality can be improved via the reasonable coding resource allocation.

The remainder of this paper is organized as follows. Section 2 presents our algorithm and the implementation steps. The experimental results and discussions are shown in Section 3. Finally, we give conclusion in Section 4.

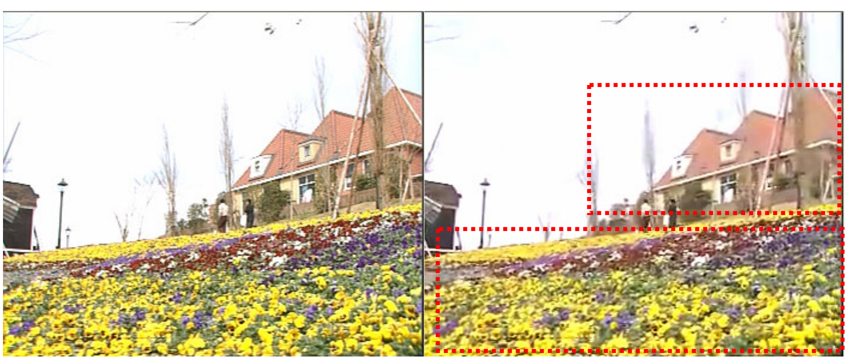
2 Proposed Method

2.1 Problem Analysis

The most common way which uses a fixed quantizer is based on a simple hypothesis: the D-Q characteristics are nearly the same within a frame. Obviously, it does not fit in most cases because D-Q characteristics are source dependent, or content dependent.

Fig. 1 is an example where a fixed QP is assigned to each macroblock. Picture (a) is the 121th frame of "flower garden" (CIF, 352*288). Picture (b) is the reconstructed image encoded by H.264/AVC JM software with the QP set to 40. Obviously the visual quality of the flower part and house part (inside the dashed line) are different. More specifically, the house part looks worse than the flower part although they use the same quantization parameter.

In this case, it is reasonable to increase the quality of the house, or decrease the quality of the flower. However, this adjusting strategy is based on a subjective observation which is unavailable in an encoding system. Therefore a quality evaluation metric should be employed to detect the quality fluctuation.



(a) original image (b) encoded with QP=40

Fig. 1. Quality fluctuation with a fixed QP

The most commonly used method in video coding is PSNR. Although easily calculated, it has been found to correlate poorly with subjective quality ratings [9]. In this case, the PSNR of the house part is 24.23db and the flower part has only 22.19db, which means it gives an opposite result compared to the observation conclusion. If we use PSNR as the quality evaluation metric, we might improve the quality of the flower part or decrease the house part. Obviously, this will lead to more serious quality fluctuation, although it may have better result in PSNR.

2.2 Distortion Model

Since PSNR/MSE is not good for subjective quality evaluation, we need a new one to replace it. In [8], an MOSp metric was proposed to give a weighted parameter to MSE in order to get more precise evaluation result. It is based on an obvious theory that visibility of artefacts in highly detailed regions is lower than in low detailed regions. The weighted parameter is based on the texture complexity of image content. The simulation result shows that it correlates better to subjective evaluation results than the original MSE. Considering the simplicity and effectiveness of this algorithm, we adopt it to help control the quality consistency. The metric is defined as:

$$Quality = 1 - k \cdot MSE \tag{1}$$

Quality has a value range of [0, 1]. 1 represents a perfect quality and 0 is a worst quality. *MSE* is the mean square error of the region. *k* is related to the edge strength of the region which can be referred as:

$$k = 0.03585 * \exp(-0.02439 * EdgeStength) \tag{2}$$

where *EdgeStrength* is the average edge strength of the region. The edge strength of a single pixel is computed by Sobel edge detecting filters [10] as:

$$EdgeStength(x, y) = \left|G_{Horizontal}(x, y)\right| + \left|G_{Vertical}(x, y)\right| \tag{3}$$

where *G* is the edge magnitude image and (x, y) is the pixel location.

2.3 New D-Q Model

After the new distortion metric selected, we are able to establish a new D-Q model. Firstly we introduce a classic and famous D-Q model, which is used in a zero-mean independent and identically distributed source [11]. It can be written as:

$$D = \frac{Q^2}{\varepsilon} \tag{4}$$

where *D* means "distortion", which is usually denoted by *MSE*. *Q* is quantization step and ε is a source dependent parameter. Then (1) and (4) can be combined into:

$$Q^2 = \varepsilon \cdot MSE = \frac{\varepsilon}{k}(1 - D_{new}) \tag{5}$$

where D_{new} is the new distortion level parameter with a value range of [0, 1]. This new D-Q model can be used to help control the quality fluctuation by allocating different quantization parameters to different regions.

2.4 Region Partition

In this section we will discuss how to make use of the new D-Q model to keep consistent quality within a frame.

The new distortion model is based on the detail complexity of the image. Therefore we divide a picture into different detailed regions, such as highly detailed part and low detailed part. The k-means clustering algorithm [12] is employed, which can partition n observations ($\mathbf{x}_1$, $\mathbf{x}_2$, …, $\mathbf{x}_n$) into k sets ($k < n$) $\mathbf{S}=\{S_1, S_2, \ldots, S_k\}$ so as to minimize the within-cluster sum of squares (WCSS):

$$\underset{S}{\operatorname{argmin}} \sum_{i=1}^{k} \sum_{x_j \in S_i} (x_j - \mu_i)^2 \tag{6}$$

where μ_i is the mean of S_i.

Fig. 2. Region partition according to the detail complexity

2.5 Our Proposed Scheme

Our new scheme divides a frame into different regions according to their detail complexity (shown in Fig. 2). According to the new D-Q model, each region will be assigned a proper quantization parameter (usually highly detailed region gets larger QP). With the QP adjustment, the quality difference between each region can be minimized and the quality consistency within a frame will be optimized.

The implementation of our coding scheme can be roughly divided into two steps:

Step 1: Calculate the average *EdgeStrength* of each macroblock by Sobel edge detecting filters (3). Then partition them into k sets, and calculate the average *EdgeStrength* of each set, which can be denoted as:

$$Set\_AVG\_ES_i = \frac{1}{N_i} \cdot \sum_{j=1}^{N_i} MB\_ES_j \tag{7}$$

where S_i is the i^{th} set, N_i is the MB number of S_i. Then the k values of each set can be achieved by using (2), which is:

$$k_i = 0.03585 * \exp(-0.02439 * Set\_AVG\_ES_i) \tag{8}$$

Step 2: A target distortion value has to be set before a frame is encoded, and each region of the current frame will be encoded to approach the target quality. This target value can be a constant one, which will keep the quality of the whole sequence consistent. In another way, the target value can also be adapted based on the quality of the previous frame. The range of the adaption is related to the bit budget remains. Once the target distortion is assigned, quantization steps of each region can be calculated by (5), which can be referred as:

$$Q_i = \sqrt{\frac{\varepsilon_i}{k_i}(1 - D_{Target})} \tag{9}$$

where k_i can be achieved by (8) and ε_i is source dependent which can be predicted by former encoded frames.

3 Experimental Results and Analysis

Three sequences are chosen for experiments which have significant variation in image complexity. They are “flower_garden”, “coastguard” and “stefan” with the resolution of 352*288. The experiments are based on H.264/AVC reference codec model JM12.2. For each sequence, each frame is divided into two parts, the high detailed and low detailed, using the k-means clustering algorithm.

As mentioned before, our new scheme is designed to decrease visual quality fluctuation within a frame and this will help improve the overall quality. And also we discussed the flaws of PSNR when doing evaluations of different image content in section 2.1. Therefore we will not simply using PSNR comparison to judge the effectiveness of our proposed scheme. The discussion about the coding efficiency is divided into two parts, the objective evaluation and the subjective one.

The objective test is done by encoding the selected sequences with different strategies. Different from traditional tests, the SSIM [13] metric, which is a famous perceptual metric, is employed together with the PSNR for comparison. The results

Table 1. Comparison of coding efficiency

	Ours		*JM12.2*		*Bit-rate (kbps)*
	PSNR	SSIM	PSNR	SSIM	
Coastguard	29.35	**0.76**	29.65	**0.724**	230
	29.94	**0.789**	30.3	**0.751**	290
	30.47	**0.811**	30.82	**0.771**	340
	31.31	**0.839**	31.61	**0.802**	420
	32.35	**0.871**	32.62	**0.835**	560
Flower Garden	25.29	**0.648**	25.33	**0.624**	230
	25.95	**0.666**	26.06	**0.642**	280
	26.52	**0.687**	26.61	**0.649**	310
	27.01	**0.695**	27.12	**0.659**	350
Stefan	27.32	**0.824**	27.56	**0.807**	210
	28.01	**0.835**	28.27	**0.818**	240
	28.66	**0.846**	28.89	**0.831**	270
	29.73	**0.867**	29.99	**0.844**	330

are shown in Table 1. We can see that our scheme achieves lower PSNR (0.2-0.3db) than JM12.2, which is predictable considering the strategy we carry out. However in SSIM metric, our scheme gets better results. That means our scheme gets better perceptual visual quality.

In order to further prove the efficiency of the new scheme, more subjective tests were done by following the guidelines in ITU-BT.500 [14], involving 12 naive evaluators. The result is shown in Figure 3. Our scheme achieves better subjective quality than the original JM12.2. Two obvious features should be mentioned. Firstly, the quality improvement is larger in lower bandwidth than in higher bandwidth. Secondly, in scenes with large range of texture variation (like flower_garden), the quality enhancement tends to be larger.

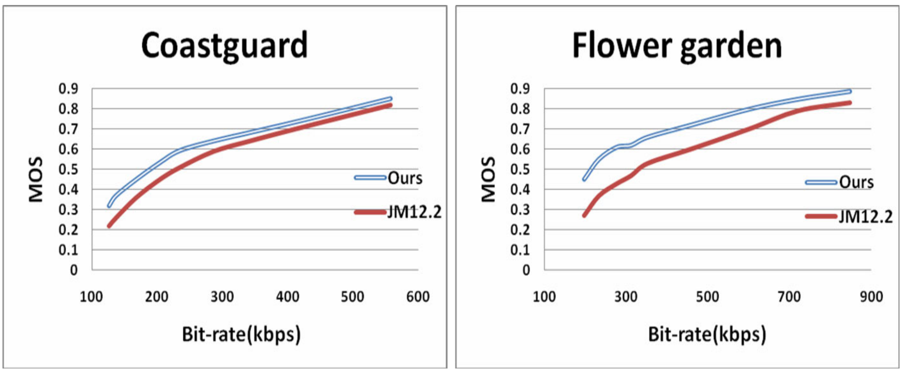
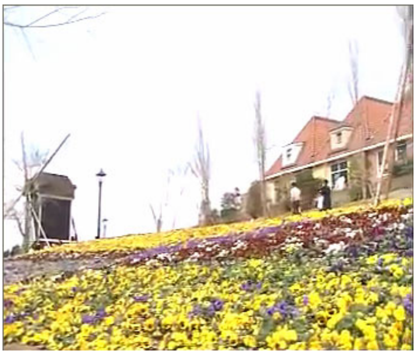


Fig. 3. Results of subjective tests

Figure 4 is a more detailed subjective quality comparison. It is about the 90^{th} frame of “flower garden”. (a) is the original frame. (b) and (c) are encoded by our scheme and JM12.2 respectively, with 311kbps. When evaluated by PSNR metric, (b) and (c) have similar values (26.08db and 26.14db). However, when using SSIM, it seems to have different result. The SSIM metric judges (b) as 0.691 and (c) as 0.646. That means (b) has better quality than (c), which is opposite to the conclusion made by PSNR metric. The subjective evaluation seems to be consistent with conclusions made by SSIM. We can see that the image coded by JM12.2 has poor quality with the tree branch and house roof (too many details are lost). In contrast, our scheme works much better, the whole image looks consistent and has better visual quality.

The reason of the quality improvement is easy to understand. We still take Figure 4 for example. Compared to using fixed QP (JM12.2), our scheme slightly raise the QP of the flower part to save bits which can be used to improve the quality of low detailed regions (such as house and trees). The degradation of flower part is less sensitive because of its high detailed texture, but the improvement of house and trees is much more sensitive. It may cause overall PSNR degradation, but it makes the overall perceptual quality much better. Based on the reasons mentioned above, it is easy to understand that our scheme works better in low bandwidth and in scenes with large range of texture variation.

(a) Original



(b) Ours
PSNR=26.08db
SSIM=0.691

(c) JM12.2
PSNR=26.14db
SSIM=0.646

Fig. 4. Subjective evaluation between our scheme and JM12.2 (the 90th frame of "flower garden" with bit-rate at 311kbps)

The extra calculation burden is added into our proposed scheme such as edge strength calculation and clustering in each frame. Therefore the complexity of the additional calculation is tested. Our platform is Intel Q9550 CPU (with 3G RAM), and the operating system is Windows XP. In the case of flower_garden (CIF) sequence, the processing speed of our scheme and original one are 1.67fps and 1.68fps, respectively. Therefore the edge strength and clustering calculations will not affect coding speed much (0.01fps). Our scheme can be used in real-time systems.

4 Conclusions

We have proposed a new quality control scheme which aims to keep quality consistency within a frame. Our new scheme modifies the traditional MSE into texture weighted distortion model. The simulation results show that the new scheme achieves better subjective quality and better coding efficiency, especially in scenes with large range of texture variation. In future work, we plan to implement the coding scheme to smaller control unit, such as MB level, in order to improve the preciseness of quality control.

Acknowledgements

This paper is supported by National Basic Research Program of China (973 Program, 2007CB311100), National Nature Science Foundation of China (60802028), Beijing New Star Project on Science & Technology (2007B071), Co-building Program of Beijing Municipal Education Commission.

References

1. Viterbi, A., Omura, J.: Principles of Digital Communication and Coding. McGraw-Hill, New York (1979)
2. Schuster, G.M., Katsaggelos, A.K.: Rate-Distortion Based Video Compression. Kluwer Academic Publishers, Norwell (1997)
3. Lin, L.J., Ortega, A.: Bit-rate control using piecewise approximated rate–distortion characteristics. IEEE Trans. Circuits and Systems for Video Technology 8, 446–459 (1998)
4. Chen, Z., Ngan, K.N.: Distortion variation minimization in real-time video coding. Signal Processing-Image Communication 21, 273–279 (2006)
5. Hong, S.H., Yoo, S.J., Lee, S.W., Kang, H.S., Hong, S.Y.: Rate control of MPEG video for consistent picture quality. IEEE Transactions on Broadcasting 49, 1–13 (2003)
6. Li, Z., Pan, F., Lim, K.P., Feng, G., Lin, X., Rahardja, S.: Adaptive basic unit layer rate control for JVT. JVT-G012 (2003)
7. Hoang, D.T., Linzer, E., Vitter, J.S.: Lexicographic bit allocation for MPEG video. Journal of Visual Communication and Image Representation 8, 384–404 (1997)
8. Bhat, A., Richardson, I., Kannangara, S.: A new perceptual quality metric for compressed video. In: IEEE International Conference on Acoustics, Speech and Signal Processing, pp. 933–936 (2009)
9. Girod, B.: What's wrong with mean-squared error. In: Digital Images and Human Vision. MIT Press, Cambridge (1993)
10. Ran, X., Farvardin, N.: A perceptually motivated three-component image model – Part 1: Description of the model. IEEE Trans. on Image Processing 4, 401–415 (1995)
11. Jayant, N., Noll, P.: Digital Coding of Waveforms, Englewood Cliffs, NJ (1994)
12. MacQueen, J.: Some methods for classification and analysis of multivariate observations. In: Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability, pp. 281–297 (1967)
13. Wang, Z., Bovik, A.C.: A universal image quality index. IEEE Signal Processing Letters 9, 81–84 (2002)
14. ITU-R BT.500 Methodology for the Subjective Assessment of the Quality for TV Pictures, ITU-R Std. (2002)