OPTIMUM BIT ALLOCATION FOR FGS VIDEO CODING

Jun Xie, Liang-Tien Chia, Bu-Sung Lee

Center for Multimedia & Network Technology,
School of Computer Engineering, Nanyang Technological University, Singapore 639798
{xiejun@pmail.ntu.edu.sg, asltchia@ntu.edu.sg, ebslee@ntu.edu.sg}

ABSTRACT

This paper proposes a new bit allocation scheme for Fine-Granular-Scalability (FGS) video coding, through which we can achieve better video quality. Different from traditional rate-distortion (*R-D*) optimization schemes, we consider the characteristics of the bit-plane (BP) coding method. To be specific, we first find the approximate linear relationship between the bitrate of the FGS-layer and the percentage of non-zero binary-scaled coefficients (*NZBC*) in each BP; second, with mathematical justification, we derive an optimal strategy by analyzing the overall distortion with respect to *NZBC*. Finally, we perform our optimum bit allocation (OBA) on a FGS coder. Experimental results prove that our scheme can achieve smooth video quality with a higher average PSNR gain compared with uniform bit allocation (UBA). And for certain frames with lower PSNR, it has a gain of up to 3dB. It is highly source-independent and more robust compared with previous bit allocation schemes.

1. INTRODUCTION

The FGS video coding technique [1], proposed as a new flexible and simple framework for scalable coding, has been incorporated into MPEG-4 as an amendment for the streaming video profile. The major difference between FGS and conventional SNR scalability is the adoption of BP coding for DCT coefficients instead of the conventional quantization of DCT coefficients. If we encode and send all the BP information of FGS-layer, which is unnecessary, the bitrate will reach quite a high level that will exhaust network resources greatly and probably result in traffic congestion; realistically, video bitstreams are often transmitted in a bandwidth-constrained environment. Therefore, there is a need to regulate the bitstream of the FGS layer and accordingly perform an efficient bit allocation.

In a sentence, the aim of *R-D* analysis and OBA is to minimize the overall distortion, given a certain rate constraint. Recently several optimal rate allocation schemes [2, 3, 4, 5] for FGS coding have been proposed. In [3], the *R-D* relationship is extracted from experimental video data; such an empirical approach has not given us a deep insight into the FGS-coding behaviors. In [2, 3], the classical *R-D* model, used in general transform coders, is employed to express the coding behaviors of FGS layer approximately. However, without accurate *R-D* model we can't perform accurate optimal bit allocation [6]. According to our observation, traditional OBA solution is not reasonable enough for FGS coding. In the latest work, a mixed statistical model [7] is proposed and distortion analysis [5, 7] is performed considering some characteristics of FGS coding, but they do not develop a corresponding rate model to perform an analytic *R-D* optimization. Therefore, a more robust and reasonable bit allocation framework is required.

In this paper, we will propose a new bit allocation scheme for FGS coding based on analyzing the relationship between the *R-D* behaviors with respect to *NZBC*. We observe that the percentage of *NZBC*, $P1$, has an important effect on the bitrate of FGS layer; and at the same time the distortion is directly decided by the coded *NZBC*. Based on our *R-D* analysis, we come to an optimal strategy to encode those binary coefficients and then perform our bit allocation scheme.

The rest of the paper is organized as follows: in Section 2, we review previous *R-D* model and OBA for FGS coding briefly, and then highlight what we believe are some of the limitations; section 3 investigates some *R-D* characteristics of FGS coding and present our *R-D* analysis; in Section 4, the proposed bit allocation scheme is implemented for a practical FGS coder and compared with the UBA scheme; conclusions are given in Section 5.

2. BACKGROUND AND PROBLEMS

In this section, we give a short review of traditional OBA techniques; and point out some existing problems.

2.1. Optimum bit allocation

Let D_i and R_i be the distortion and rate for the input source $\{S_i | 1 \leq i \leq N\}$ respectively. Usually, for transform visual compression, the *R-D* model [8] is expressed as (1), where R_i denotes the average bits per pixel, σ_i^2 is the signal variance of S_i , and ϵ_i^2 is a constant dependant on the probability density function ($p.d.f$) of S_i, about 1.0 for uniform distribution, 1.4 for Gaussian, and 1.2 for Laplacian [9]. In practice, ϵ_i^2 is affected by the type of encoding used as well.

$$D_i = \epsilon_i^2 \sigma_i^2 2^{-2R_i} \tag{1}$$

The objective of OBA is to minimize the distortion, subject to a rate constraint. Thus, the problem of OBA can be expressed with the popular Lagrange multiplier method (2) and the optimal solution is given below (3), where R_T is the target rate.

$$J(\lambda) = \sum_{i=1}^{N} D_i(R_i) + \lambda \cdot (\sum_{i=1}^{N} R_i - R_T) \tag{2}$$

$$R_i = \frac{R_T}{N} + \frac{1}{2}\log_2 \frac{\epsilon_i^2 \sigma_i^2}{(\prod_{j=1}^{N} \epsilon_j^2 \sigma_j^2)^{\frac{1}{N}}} \tag{3}$$

2.2. Review of bit allocation strategies for FGS

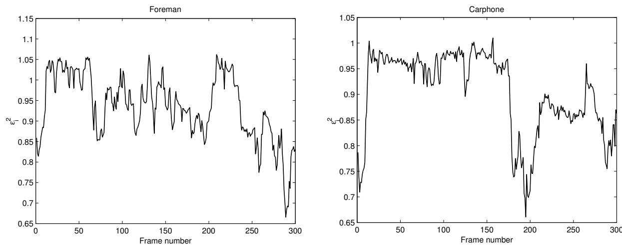
UBA is the simplest way among all bit allocation schemes for FGS coding. It allocates an average number of bits to each input source. The coding behavior will stop once the allocated bits are used up. Because there is no optimization technique, such an approach has low coding efficiency.

More efficient bit allocation schemes [2, 3, 4] have been developed based on the *R-D* optimization. The *R-D* analysis is the basis of OBA. It is therefore the key to achieve an accurate *R-D* function to model the characteristics of FGS-layer coding. Previously, there are two basic approaches to solving the problem. One approach follows the traditional analytic *R-D* model (1) and its solution (3), which leads to constant visual quality among frames at the same time [2, 3]. In another approach, several *R-D* points are extracted empirically during the encoding processing [3].

However, there are some practical drawbacks when we apply the traditional analytic *R-D* model to FGS coding directly. First, the $p.d.f$ of the source for FGS-layer coding, is uncertain because it is dependent not only on the original video frames but also on the quantization strategies of the base layer (BL). In the *R-D* model (1), ϵ_i^2 reflects the $p.d.f$ of the input source S_i. In existing schemes, they simply assume that all ϵ_i^2 is the same and then use the simplified solution (4), which means that all source data have the same $p.d.f$. However, we find that such a simplification isn't reasonable enough. For instance, the value of ϵ_i^2 for different frame in the FGS-layer is computed according to (1) and plotted in fig.1, which shows us that it is inaccurate and unreasonable to assume ϵ_i^2 is approximately constant.

$$R_i = \frac{R_T}{N} + \frac{1}{2}\log_2 \frac{\sigma_i^2}{(\prod_{j=1}^{N} \sigma_j^2)^{\frac{1}{N}}} \tag{4}$$

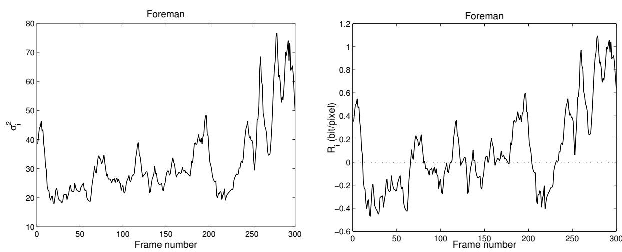
Second, there is obviously one practical constraint $R_i > 0$. However, when applying the theoretical solution (4) to a practical FGS coder, we probably get $R_i < 0$ for a certain percentage of the total frames. From (4), the first item is related to R_T and the second item is related to σ_i^2. Because



(a) Foreman (b) Carphone

Fig. 1. Value of ϵ_i^2 for different frames when BL is coded with VM18 rate control at 100kbit/s and FGS-Layer is coded with UBA at 100kbit/s.

R_T and σ_i^2 are independent, for a sequence with N frames, when there are many negative R_i in the optimal solution due to a small R_T or widely varying σ_i^2; a large rate control error will be generated if we simply set the R_i with negative values to 0 as mentioned in [3]. One example is shown in fig.2 to illustrate the bit allocation result with the traditional OBA schemes, which isn't robust. The result will be worse if the target bitrate is further reduced or σ_i^2 varies in a wider range.



(a) (b)

Fig. 2. Allocated number of bits, R_i, computed by previous bit allocation schemes, when both BL and FGS-Layer are encoded at 100kbit/s respectively. $N = 300$, Foreman. (a) σ_i^2; (b) R_i

3. *R-D* ANALYSIS OF FGS CODING

For convenience, we introduce some definition, which is also applicable to the Appendix.

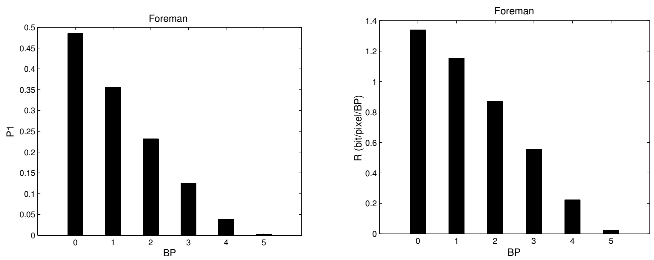
Subscript i: frame index;
Subscript j: BP level index, corresponding to the resolution level of 2^j;
Subscript k: the coefficient index, for both DCT and binary coefficients;
x: DCT coefficient;
b: binary coefficient;
N: the size of the data source;
J: the set of encoded BP levels;

In the standard FGS coder, source signals are transformed to DCT coefficients; then we decompose them into different BP and apply entropy coding to encode different BP starting first with the most significant bit (MSB) plane.

With a given coder, *R-D* function is totally dependent on the statistical characteristics of the DCT coefficients [10]. However, as illustrated in 2.2, the $p.d.f$ of the source DCT coefficients varies a lot and it is difficult to model accurately in a mathematically tractable form. As a result, we can't get an efficient bit allocation scheme by simply following the traditional method. In [10], the framework of *R-D* analysis is developed on the study of its relationship with the percentage of zero-valued quantized DCT coefficients, one advantage of which is little source-independent. In FGS coding, the source for the encoder will be those binary coefficients in different BP levels. Similarly, it must be necessary and beneficial to study the effect of those binary-scaled coefficients on our desired *R-D* model.

3.1. Rate Model

The actual distribution of *NZBC* and rate for different BPs is shown in fig.3. We can see that the percentage of *NZBC* in the BP j, $P1_j$ decreases gradually when the BP level j increases; and the ultimate bitrate of the BP j, R_j, has a strong correlation with $P1_j$. In FGS coding, $P1$ is quite different from BP to BP (e.g. as shown in fig.3), which results in different statistical characteristics for the rate model. In addition, four VLC tables are designed corresponding to MSB plane, MSB-1 plane, MSB-2 plane, and other BPs based on the statistics[1]. Therefore, we investigate the characteristics of our rate model on the basis of the BP.



(a)*NZBC* distribution (b)bitrate distribution

Fig. 3. Example of *NZBC* and rate distribution among BPs.

Next, let us see the relationship between the $P1_j$ and R_j from fig.4. Here note that the $P1_j$ stands for the percentage of the encoded *NZBC* in one BP if the whole BP is not encoded entirely. From the straight gradient shown in the fig.5, we can observe that there is a linear relationship between $P1_j$ and R_j. Let $b_{j,k}$ be the binary coefficient in the BP j of x_k. Then $P1_j$ can be computed by (5). The rate function for one BP (6) and the whole frame (7) are written below, where θ_j is a statistical slope and can be estimated according to experimental data and N_j is the size of BP j. Note it is possible that not all the *NZBC*s are encoded in the BP j among our encoded BPs, and in that case $P1_j$ stands for the percentage of the encoded *NZBC* among the encoded source.

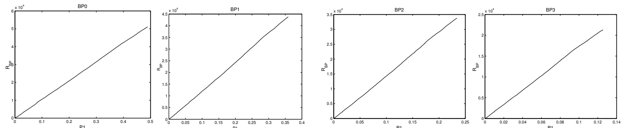


Fig. 4. Plot of $P1_j$ versus R_j in different BP from the same frame: (a)-(d) the BP level j from 0 to 3 respectively.

$$P1_j = \frac{\sum_{k=1}^{N_j} b_{j,k}}{N_j} \qquad (5)$$

$$R_j = \theta_j \cdot P1_j \cdot N_j \qquad (6)$$

$$R = \sum_{j \in J} \theta_j \cdot P1_j \cdot N_j \qquad (7)$$

3.2. Distortion Analysis with respect to *NZBC*

With the linear relationship between *R* and *NZBC*, *R-D* analysis can be transformed to minimize distortion at a certain budget number of *NZBC*. Next, we will study the relationship between the overall distortion and the strategy to encode *NZBC*.

The usual way to reflect the picture quality is the distortion measured by mean square error (MSE). We use the binary coefficients in different BPs to express the overall distortion. Assume that the source DCT coefficients for FGS coding, $X = \{x_k | 1 \leq k \leq N\}$, are encoded in the strategy $J = \{J_k | 1 \leq k \leq N\}$, the distortion expression can be written as the following (8). We can see that the distortion is affected only when $b_{j,k} = 1$ and this corresponds to a *NZBC*. It indicates that the decision in our strategy to encode *NZBC* is the key to minimize the overall distortion.

$$D = \sum_{k=1}^{N} (x_k - \sum_{j \in J_k} b_{j,k} \cdot 2^j)^2 \qquad (8)$$

From (8) we can draw the following two conclusions (see the Appendix) when we select *NZBC* to encode. First, the larger BP level *j* that the encoded *NZBC* is located in, the smaller *D*. In other words, we should select the *NZBC* in the higher BP to be encoded first. In the case of bit allocation among multiple frames, it indicates that we should ensure the higher BP among all frames to have a higher priority during encoding. Second, to minimize *D*, the *NZBC* decomposed from x_k with a larger residue of x_k should have a higher priority to be encoded. What this implies is that in the same BP j, we prefer to encode the one decomposed from the DCT coefficient with a larger residue, if the remaining number of bits isn't sufficient to encode the whole BP.

Our above *R-D* analysis is based on the distribution of *NZBC* among BPs. It is little dependence on the the $p.d.f$ of the source. In addition, our distortion model (8) is the actual distortion, which is more accurate than the traditional one estimated by the rate-quantization theory.

4. EXPERIMENTAL RESULTS

With the rate model taking *NZBC* into consideration and the distortion analysis with respect to *NZBC*, we have derived our optimal strategy.

We apply it to a basic FGS coder. Foremen, Carphone sequences at QCIF resolution, 300 frames, 4:2:0, are tested. The encoding frame rate is 25 frame/s for both BL and FGS-layer. A comparison between our proposed OBA scheme and UBA scheme is given in fig. 5. From the figure, we can see that the picture quality appears to be smooth and it also achieve a higher average PSNR gain, 0.16dB for Foreman and 0.25 dB for Carphone. For frames whose BL has a very low PSNR, a gain of up to 3dB can be achieved compared with UBA. In comparison, it shows that our algorithm is more efficient.

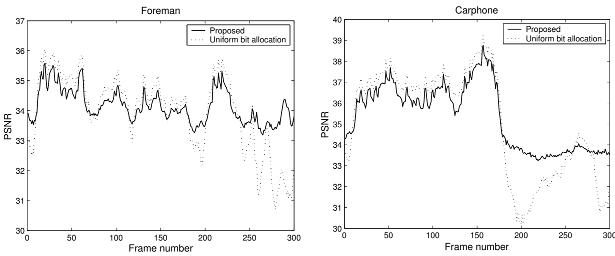


Fig. 5. PSNR of Foreman (a) and Carphone (b) Sequence, when BL and FGS-layer are encoded at 100kbit/s respectively. BL is controlled by the VM18 rate control algorithm. Solid line: our proposed work, dotted line: UBA

5. CONCLUSION

In this paper, the problem of bit allocation for FGS coding was highlighted. At first, we point out the shortcomings of applying traditional OBA schemes to FGS coding. Next, we propose our new bit allocation scheme, which is based on the analysis of *R-D* with respect to the *NZBC*. The main contribution of our work are: first, the rate model for FGS coding based on the linear relationship between the rate and the percentage of *NZBC*, which is more source independent, simple and accurate, was developed. Second, we analyze the relationship between the distortion and *NZBC*. From the theoretical derivation, it is reasonable to develop a strategy to optimize the overall distortion. Supporting experimental results have been provided. Overall, the proposed approach can provide higher video quality and is more robust.

Appendix:Proof of the strategy to encode *NZBC*

Assume the data source $X = \{x_k | 1 \leq k \leq N\}$ with an arbitrary encoding strategy $J = \{J_k | 1 \leq k \leq N\}$.

Next, if we only can encode one more *NZBC*,which one should be encoded to minimize the D? Select two arbitrary un-encoded *NZBC*: $b_{j1,k1}$ and $b_{j2,k2}$. According to (8), the distortion is shown below, where $D_{j1,k1}$ is the D when selecting $b_{j1,k1}$ and $D_{j2,k2}$ is in the case of selecting $b_{j2,k2}$. In the following, prove our strategy in three steps by comparing $D_{j1,k1}$ and $D_{j2,k2}$.

$$\begin{cases} D_{j1,k1} = \sum_{k=1,k\neq k1}^{N} (x_k - \sum_{j\in J_k} b_{j,k} \cdot 2^j)^2 \\ \qquad\qquad + (x_{k1} - \sum_{j\in\{J_{k1}+j1\}} b_{j,k1} \cdot 2^j)^2 \\ D_{j2,k2} = \sum_{k=1,k\neq k2}^{N} (x_k - \sum_{j\in J_k} b_{j,k} \cdot 2^j)^2 \\ \qquad\qquad + (x_{k2} - \sum_{j\in\{J_{k2}+j2\}} b_{j,k2} \cdot 2^j)^2 \end{cases}$$

1. When $k1 = k2$, and $j1 \neq j2$, $\Delta D = D_{j1,k1} - D_{j2,k2} =$
$$\{2 \cdot (x_{k1} - \sum_{j\in J_{k1}} b_{j,k1} \cdot 2^j) - 2^{j2} - 2^{j1}\}(2^{j2} - 2^{j1})$$
When $j1 > j2$, $\Delta D < 0$. It prove that we should choose the one located in the highest BP level on such a condition.

2. When $k1 \neq k2$, and $j1 \neq j2$.
Suppose $j1 > j2$, then we can get (9). Following the above step, we assume $j1$ and $j2$ satisfy the above conclusion, then we can get (10).
$$2^{j1} \geq 2 \cdot 2^{j2} \tag{9}$$
$$2^{j2} > x_{k2} - \sum_{j\in\{J_{k2}+j2\}} b_{j,k2} \cdot 2^j \tag{10}$$
$$\Delta D = \{2^{2\cdot j2} + 2^{j2+1} \cdot (x_{k2} - \sum_{j\in\{J_{k2}+j2\}} b_{j,k2} \cdot 2^j)\}$$
$$-\{2^{2\cdot j1} + 2^{j1+1} \cdot (x_{k1} - \sum_{j\in\{J_{k1}+j1\}} b_{j,k1} \cdot 2^j)\}$$
Combining (9) and (10), we can derive $\Delta D < 0$. It prove that we should select the one located in the higher BP level.

3. When $k_1 \neq k_2$, and $j_1 = j_2$. ΔD is
$$2^{j1+1}\{(x_{k2} - \sum_{j\in J_{k2}} b_{j,k2} \cdot 2^j) - (x_{k1} - \sum_{j\in J_{k1}} b_{j,k1} \cdot 2^j)\}$$
When $x_{k1} - \sum_{j\in J_{k1}} b_{j,k1} \cdot 2^j > x_{k2} - \sum_{j\in J_{k2}} b_{j,k2} \cdot 2^j$, $\Delta D < 0$, namely we should choose the one with a larger residue on such a condition.

6. REFERENCES

[1] W. Li, "Overview of fine granularity scalability in MPEG-4 video standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 11, pp. 301–317, Mar. 2001.

[2] Q. Wang, Z. Xiong, F. Wu, and S. Li, "Optimal rate allocation for progressive fine granularity scalable video coding," *IEEE Signal Processing Lett.*, vol. 9, pp. 33–39, Feb. 2002.

[3] X. M. Zhang, A. Vetro, Y. Q. Shi, and H. Sun, "Constant quality constrained rate allocation for FGS-coded video," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, pp. 121–130, Feb. 2003.

[4] L. Zhao, J. W. Kim, and C.-C. Kuo, "Constant quality rate control for streaming MPEG-4 FGS video," in *IEEE ISCAS*, vol. 4, Mar. 2002, pp. IV.544–547.

[5] S. Parthasarathy and H. Radha, "Optimal rate control methods for fine granularity scalable video," in *ICIP*, vol. 2, Sept. 2003, pp. III – 805–9.

[6] W. Ding and B. Liu, "Rate control of MPEG video coding and recording by rate-quantization modeling," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 6, pp. 12 –20, Feb. 1996.

[7] M. Dai, D. Loguinov, and H. Radha, "Statistical analysis and distortion modeling of MPEG-4 FGS," in *ICIP*, vol. 3, Sept. 2003, pp. III – 301–4.

[8] N. S. Jayant and P. Noll, *Digital Coding of Waveforms Priciples and Applications to Speech and Video*. Englewood Cliff, New Jersey: Prentic-Hall, 1984.

[9] H.-M. Hang and J.-J. Chen, "Source model for transform video coder and its application. I. Fundamental theory," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 7, pp. 287 –298, Apr. 1997.

[10] Z. He, Y. K. Kim, and S. K. Mitra, "Low-delay rate control for DCT video coding via -domain source modeling," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 11, pp. 928 –940, Aug. 2001.