BIT ALLOCATION FOR FINE-GRANULAR SNR SCALABILITY CODING WITH HIERARCHICAL B PICTURES

Jun Xu, Li Song, Shibao Zheng, Xiaokang Yang, Rong Xie

Institute of Image Communication and Information processing
Shanghai Jiao Tong University
Shanghai 200240, China

ABSTRACT

Hierarchical B pictures are devised to achieve temporal scalability in the scalable extension of H.264/AVC (SVC) which is under standardization. The fine-granular SNR scalability (FGS) can be provided by progressive refinement (PR) slices in SVC. In this paper, we firstly investigate error propagation in the case of discarding PR slices and obtain a rate difference distortion optimization criterion to improve coding efficiency of base layer. Then we consider the full rate case and propose a rate distortion slope criterion to enhance FGS coding efficiency at high rate. Finally the criterion to boost coding efficiency in the whole range of FGS rate is derived by combing the criterions derived previously. The proposed method is compared to the approach in SVC test model and up to 0.3dB coding gains are achieved.

1. INTRODUCTION

Scalable video coding (SVC) based on H.264/AVC, which is currently jointly developed by ISO/IEC MPEG and ITU-T VCEG [1] [2], has achieved significant improvement in coding efficiency when providing spatial, temporal and SNR scalability [3]. Temporal scalability is achieved by hierarchical B pictures. SNR scalability is divided into coarse-granular SNR scalability, medium-granular SNR scalability and fine-granular SNR scalability (FGS). FGS can be provided by progressive refinement (PR) slices. FGS coding increases the flexibility for bit-stream adaptations and the error robustness. It is thus especially suitable for streaming applications in which the video bit-rate has to be flexibly adapted to the channel conditions, or in combination with unequal error protection for error-prone environments. Yet the efficiency of FGS coding is quite lower than that of the single layer coding. Truncating or discarding PR slices will reduce reconstruction quality at the decoder and may cause error accumulation within hierarchical B pictures which is called drift effect. FGS coding should be carefully designed to improve coding efficiency in the whole range of FGS rate.

This work was supported by NSFC (60332030, 60625103), NCET-06-0409 and the 111 Project.

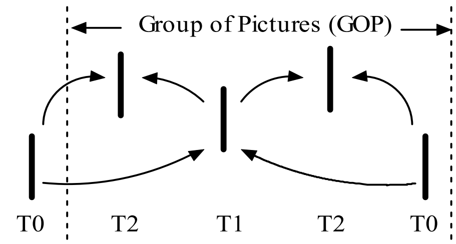

Fig. 1. Hierarchical B pictures with $T=3$ stages

Recently a linear distortion model for SNR scalable video coding with hierarchal B pictures is derived in [4] to adjust quantization parameters (QP) in base layer. It can enhance coding efficiency compared to SVC test model (JSVM). The parameters in the model are computed only if encoding the sequence using the unmodified encoder. After getting the model parameters, another pass of coding has to be performed. It is inconvenient in practice. In the JSVM, adaptive QP selection is not enabled for intra macroblocks. Yet adaptive QP selection for intra macroblocks is enabled in [4].

In this paper, we first analyze error propagation when discarding PR slices and acquire a rate difference distortion optimization criterion to enhance coding efficiency of base layer with low complexity. Then we consider the full rate case and propose a rate distortion (RD) slope criterion to enhance FGS coding efficiency at high rate. Finally the proposed scheme is to combine the criterions derived previously to enhance coding efficiency in the whole range of FGS rate. In the proposed scheme, QP in both base and enhancement layers are adjusted. Only one pass coding is needed and the bit-stream is compatible with the current standard.

2. INVESTIGATED SYSTEM

2.1. Hierarchal B pictures

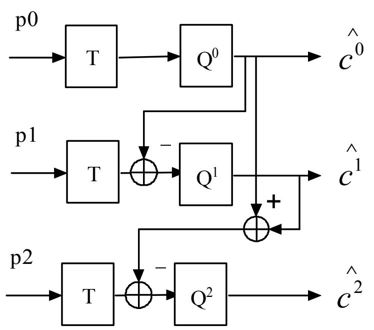

Fig. 2. Progressive refinement quantization.

Typical hierarchical B pictures with three dyadic stages are depicted in Fig. 1. $T0$ corresponds to the coarsest temporal resolution frames which are called key pictures. The pictures of a GOP are hierarchically predicted as illustrated in Fig. 1. It can be seen that this structure inherently allows for dyadic temporal scalability and that frames in the finest temporal resolution are not used as reference for motion compensated prediction.

2.2. Progressive refinement slices

Progressive refinement slices have been designed for efficiently representing SNR refinements and allowing for truncation of the corresponding NAL units. Basic quantization principle is illustrated in Fig. 2. Here, p^n is the unquantized prediction residual signal of quality level n. In this paper, we focus on the case of prediction loop closed at the highest rate, which means $p^0 = p^1 = p^2$. $\hat{c}^n$ denotes the corresponding quantized transform coefficients. T represents forward spatial transform. Quantization parameter associated with quality level n is denoted by Q^0. For each quality level, quantization is performed on the refinement information relative to the preceding level.

In the JSVM, for a macroblock which does not belong to key pictures and is inter-coded, quantization parameter Q^0 in base layer can be adaptively selected in range $[QP_{base} - \Delta_{QP}, QP_{base} + \Delta_{QP}]$ in a RD optimization way. Δ_{QP} can be specified by configuration files. After Q^0 has been decided. If non-zero transform coefficient is transmitted, QP for the enhancement layer Q^1 is equal to $Q^0 - 6$, which means the quantization step size in enhancement layer is half that in base layer. It is similar to bit-plane coding. Otherwise, Q^1 can be adaptively selected in range $[QP_{enh1} - \Delta_{QP}, QP_{enh1} + \Delta_{QP}]$ according to RD optimization criterion. Here, QP_{enh1} is equal to $QP_{base} - 6$. In following enhancement layers, the above process is applied similarly.

With PR slices, the transform coefficient levels are processed in several scan cycles. The coefficient scanning can be influenced by configuration files. It is possible to adjust the trade-off between decoder complexity, which increases with the number of scan cycles, and the quality of the coarse to fine representation, which determines the coding efficiency for truncated FGS layers [3]. Yet adjusting coefficient scanning to increase FGS coding efficiency is not the focus of this paper.

3. OPTIMIZATION CRITERION

3.1. Rate difference distortion optimization

In the following, for expression simplicity, we only investigate P frame which can be seemed as one directional predictive coding of B frame. Thus the results can apply to hierarchical B pictures.

In the encoder side, frame i is predicated by frame $i-1$. Let p_i denote the unquantized prediction residual signal, then

$$X_{org,i} = MC(\hat{X}_{enc,i-1}) + p_i + I_i. \qquad (1)$$

$X_{org,i}$ are the original pixels in frame i. $\hat{X}_{enc,i}$ and $\hat{X}_{dec,i}$ are the reconstructed frame i at encode side and decoder side, respectively. MC expresses the motion compensation and is assumed to be a linear operation. I_i is the intra prediction in frame i, such as intra DC prediction. We do not consider directional intra prediction here. These assumptions are also adopted in [4]. When prediction loop is closed in the highest rate and in the case of discarding PR slices, $\hat{p}_{enh,i}$ and $\hat{p}_{base,i}$ denote the reconstructed residual in frame i in encode side and decoder side, respectively. $\hat{X}_{enc,i}$ and $\hat{X}_{dec,i}$ are computed by

$$\hat{X}_{enc,i} = MC(\hat{X}_{enc,i-1}) + \hat{p}_{enh,i} + I_i, \qquad (2)$$

$$\hat{X}_{dec,i} = MC(\hat{X}_{dec,i-1}) + \hat{p}_{base,i} + I_i. \qquad (3)$$

Jointly considering (2) and (3), we get

$$\hat{X}_{enc,i} - \hat{X}_{dec,i} = MC(\hat{X}_{enc,i-1} - \hat{X}_{dec,i-1}) + (\hat{p}_{enh,i} - \hat{p}_{base,i}). \quad (4)$$

Jointly considering (1) and (3), we get

$$X_{org,i} - \hat{X}_{dec,i} = MC(\hat{X}_{enc,i-1} - \hat{X}_{dec,i-1}) + (p_i - \hat{p}_{base,i}). \quad (5)$$

If (4) is substituted into (5), then

$$X_{org,i} - \hat{X}_{dec,i} = \sum_{n=1}^{i} MC^n(\hat{p}_{enh,i-n} - \hat{p}_{base,i-n}) + (p_i - \hat{p}_{base,i}). \quad (6)$$

We find the difference of $\hat{p}_{enh,m}$ and $\hat{p}_{base,m}$ $(m = 0,1,...,i-1)$ is propagated along the motion compensation path to frame i. From the point of view of the whole sequence, weighted combination of $(p_m - \hat{p}_{base,m})^2$, which is the distortion of its own, and $(\hat{p}_{enh,m} - \hat{p}_{base,m})^2$, which is the difference distortion to propagate, to decide QP selection in frame m should be adopted as [4]. Yet the weight factor is hard to derive before encoding. [4] involves two pass coding. Adjusting QP modifies the reference frame, and thus the weight factor derived by one pass coding is no longer accurate. In order to avoid two pass coding which brings inaccuracy and complexity, we simplify the objective to only minimizing the difference distortion for frames used as reference for motion compensation. Here, we assume $\sum_{n=1}^{i} MC^n(\hat{p}_{enh,i-n} - \hat{p}_{base,i-n})$ is the dominant distortion in (6). This assumption is not strictly valid, but it can avoid two pass coding and produce satisfactory results. Thus, minimization of the quadratic distortion of a decoded sequence subject to a rate constraint is almost equivalent to minimization of difference between base layer and full enhancement layer under the base layer rate constraint within the frames which will be used as references. We put it in a rate difference distortion optimization framework. For a macroblock within a frame which does not correspond to the finest temporal resolution, the objective is to minimize

$$(\hat{p}_{enh} - \hat{p}_{base})^2 + \lambda R_{base}. \tag{7}$$

R_{base} denotes the number of coding bits used in a macroblock.

Let D_{base} represent the reconstructed distortion of a macroblock in base layer. For a macroblock within a frame which corresponds to the finest temporal resolution in hierarchical B pictures and will not propagate difference distortion to other frames in case of truncation, according to (5) the objective is to minimize

$$D_{base} + \lambda R_{base}. \tag{8}$$

This is the same as the implementation in the JSVM.

3.2. RD slope optimization

Due to cycle based scanning and linear truncation, RD performance of truncated FGS layers can be approximately modeled as the linear model. In the JSVM, if all-zero transform coefficients are transmitted in base layer, QP selection in enhancement layer is decided by RD optimization criterion. It is the slope of RD points between base layer and full enhancement layer that decides the coding efficiency of PR slices in the case of truncation. For example, in Fig. 3, point 1 is the RD optimization point and point 2 has larger absolute value of RD slope. It can be seen

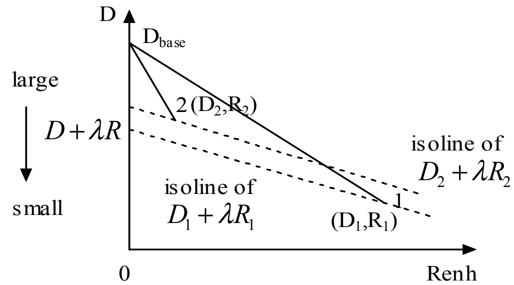

Fig. 3. FGS coding efficiency.

that point 2 is better regarding FGS coding efficiency since in the range of 0 to R_2 its RD performance is superior to that of point 1, although $D+\lambda R$ of point 2 is larger than that of point 1.

Our objective is to maximize the absolute value of RD slope of the whole sequence. Namely, we aim to maximize

$$\left(\sum_{i\in MB} D_{base,i} - \sum_{i\in MB} D_{enh,i}\right) \Big/ \sum_{i\in MB} R_{enh,i}. \tag{9}$$

$R_{enh,i}$ is the number of coding bits and $D_{enh,i}$ is the reconstructed distortion of macroblock i in enhancement layer. MB is the number of macroblocks in the sequence. For macroblock i, it can only adjust $R_{enh,i}$ and $D_{enh,i}$ of its own. This objective is hard to be decomposed into every macroblock, so we solve it heuristically. We assume $\sum_{i\in MB} R_{enh,i}$ changes faster than $(\sum_{i\in MB} D_{base,i} - \sum_{i\in MB} D_{enh,i})$ since the numerator of (9) is the difference term. So $\sum_{i\in MB} R_{enh,i}$ dominates the value of (9). The above assumption is not very reasonable and just used for finding a heuristic solution. For a macroblock that has all-zero transform coefficient in base layer, the objective is to minimize $R_{enh,i}$.

4. COMBINED OPTIMIZATION CRITERION

In order to boost FGS coding efficiency in the whole range of FGS rate, we combine the criterions acquired above to adjust QP selection in both base and enhancement layers.

We minimize (7) for macroblocks in frames not corresponding to the finest temporal resolution in hierarchical B pictures by adjusting QP selection in base layer. In the frames corresponding to the finest temporal resolution, QP selection in base layer is decided by minimizing (8). The above is in order to enhance coding efficiency in the case of discarding PR slices.

Minimizing $R_{enh,i}$ is applied in QP selection in the frames corresponding to the finest temporal resolution in enhancement layer. Thus the FGS coding efficiency at high rate is also improved.

5. EXPERIMENTAL RESULTS

We replace the cost function used for macroblock based QP selection according to Sec. 4. In our implementation, when adjusting QP selection in base layer within frames not corresponding to the finest temporal resolution, QP in enhancement layer is fixed to the maximal value in its selection range in order to minimize the difference between base layer and full enhancement layer. Given a QP, re-encoding the macroblock will produce the corresponding reconstruction information. In the JSVM, adaptive QP selection is not enabled for intra macroblocks, which is left unmodified.

We use hierarchical B pictures with T=5 temporal levels and two quality levels. The key pictures at temporal resolution T=0 are intra encoded. Δ_{QP} is set to 2. Entropy coding and coefficient scanning adopt default settings. Our proposed strategy is compared to RD optimization based QP selection of JSVM_7_10. The base quality is adjusted such that similar minimal bits rates are met in all tests.

Results are depicted in Fig. 4 for two test sequences, Foreman and Bus. "DeltaD+Enh RDslope(proposed)" represents the strategy in Sec. 4 is adopted. "DeltaD+Enh RDO" means only the strategy in Sec. 3.1 is utilized, others are the same as JSVM. It can be seen that at high rate, the FGS coding efficiency of "DeltaD+Enh RDO" approaches that of the JSVM and is dropped compared to the proposed scheme. "Base RDO+Enh RDslope" means only the criterion in Sec. 3.2 is applied, others are the same as the JSVM. When it is compared to the JSVM, its performance is only enhanced at high rate. Strategy in Sec. 3.1 or in Sec. 3.2 alone can not ensure improving the coding performance in the whole range of FGS rate. The proposed scheme in Sec. 4 combining the criterions in Sec. 3.1 and in Sec. 3.2 achieves better coding performance in the whole range of FGS rate. Up to 0.3dB coding gains are observed by the proposed method compared to the JSVM. Optimized RD extraction with quality layers [5] is performed on the proposed scheme and up to 0.2 dB coding gains are achieved.

If the criterion in Sec. 3.2 is applied on the ARFGS [6] in which QP is not allowed to adjust in base layer. Proposed scheme outperforms the JSVM up to 0.1dB. The figure is omitted due to space limitation.

6. CONCLUSION

We jointly consider QP selection of both base layer and enhancement layer instead of sequentially optimizing them according to a strategy combining rate difference distortion criterion aiming to enhance coding efficiency of base layer and RD slope criterion aiming to improve FGS coding efficiency at high rate in order to boost coding efficiency in the whole range of FGS rate. Only one pass coding is needed and the bit-stream is compatible with the current standard. The coding gains are up to 0.3 dB on the tested sequences. Our future work will comprise further refinement

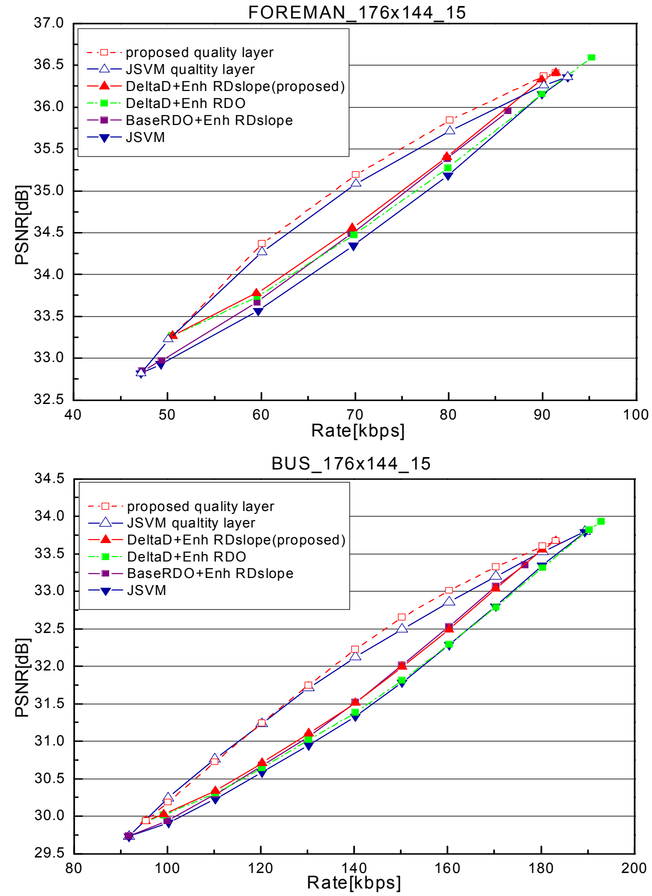

Fig. 4. Simulation results

of the criterions for macroblock based bit allocation for FGS coding with hierarchical B pictures.

7. REFERENCES

[1] T. Wiegand, G. Sullivan, J. Reichel, H. Schwarz, and M. Wien, "Joint Draft 8 of SVC Amendment," Doc. JVT-U201, Joint VideoTeam (JVT), 21th Meeting, Hangzhou, China, Oct. 2006.

[2] J. Reichel, H. Schwarz, and M. Wien, "Joint scalable video model JSVM-8," Doc. JVT-U202, Joint Video Team (JVT), 21th Meeting, Hangzhou, China, Oct. 2006.

[3] H. Schwarz, D. Marpe, and T. Wiegand, "Overview of the Scalable Extension of the H.264/MPEG-4 AVC Video Coding Standard," Doc. JVT-U145-L, Joint Video Team (JVT), 21th Meeting, Hangzhou, China, Oct. 2006.

[4] T. Rusert and J.-R. Ohm, "Macroblock Based Bit Allocation for SNR Scalable Video Coding with Hierarchical B Pictures," in *Proc. IEEE Int.Conference on Image Processing ICIP '06*, Atlanta, GA, USA, Oct. 2006.

[5] I. Amonou, N. Cammas, S. Kervadec and S. Pateux, "Optimized Rate-Distortion Extraction with Quality Layers," in *Proc. IEEE Int.Conference on Image Processing ICIP '06*, Atlanta, GA, USA, Oct. 2006.

[6] Y. Bao, M Karczewicz, X. Wang and J. Ridge, "FGS Coding with Adaptive Reference for Low-Delay Applications," in *Proc. IEEE Int.Conference on Image Processing ICIP '06*, Atlanta, GA, USA, Oct. 2006.