Smooth Extraction of SVC Fine-Granular SNR Scalable Videos with a Virtual-GOP-Based Rate-Distortion Modeling

Jun Sun[*a], Wen Gao[a], Debin Zhao[b]
[a]Institute of Digital Media, EECS, Peking University, Beijing, 100871, China;
[b]Department of Computer Science, Harbin Institute of Technology, Harbin, 150001, China

ABSTRACT

Fine-Granular SNR scalable (FGS) technologies in H.264/AVC-based scalable video coding (SVC) provide a flexible and effective foundation for scaling FGS enhancement layer (EL) to accommodate different and variable network capacities. To support smooth quality extraction of SVC FGS videos, it's important to obtain the Rate-Distortion (R-D) function of each picture or group of pictures (GOP). In this paper, firstly, we introduce the R-D analysis of SVC FGS coding in our prior work. Then, with the analysis and models, we present virtual GOP concept and a virtual-GOP-based packet scheduling algorithm is proposed to acquire the optimal packet scheduling sequence in a virtual GOP. Based on the packet scheduling algorithm and the R-D analysis of FGS EL, an effective and flexible D-R model is proposed to describe the D-R function of the virtual GOP. Then, with the R-D model of virtual GOPs, a practical non-search algorithm for smooth quality reconstruction is introduced. Compared to the quality layer method, the reconstructed video quality is improved not only objectively but also subjectively.

Keywords: Smooth quality, rate-distortion analysis, scalable video coding (SVC), video coding

1. INTRODUCTION

Internet is experiencing explosive growth of video streaming. Since the Internet is a shared environment, it is desirable to encode video with fine-granular SNR scalable (FGS) technologies, so that it can be encoded once, but transmitted and reconstructed many times at different targeting rates. The developing scalable video coding (SVC) standardization project [1] chooses scalable extension of H.264/AVC as a start point, which realizes the fine-granular SNR scalability through sub-bitplane-based progressive refinement of the FGS enhancement layer (EL). Since the gap between SVC FGS scheme and single-layer coding can be quite small, the coding scheme gained significant interest. Note that FGS coding mode has been taken away from SVC final amendment, a phase 2 SVC project is started, which may include FGS coding mode [2].

To best utilize SVC FGS videos, a bit-stream extraction (rate allocation) algorithm should be employed to optimally transfer the targeting bit rate into the rate assigned to each FGS picture. Typically, there are two optimization goals. The first is optimal extraction with rate distortion sense, which minimizes the average distortion subject to the rate constraint [3]. The second is smooth quality extraction, which expects to achieve constant quality within the constraint of targeting bit rate. Psychological research indicates that the human visual system prefers a video sequence having consistent visual quality [4]. In fact, a video sequence with higher average PSNR, but larger variation in visual quality, is generally considered inferior to a video sequence with lower average PSNR, but smaller variation.

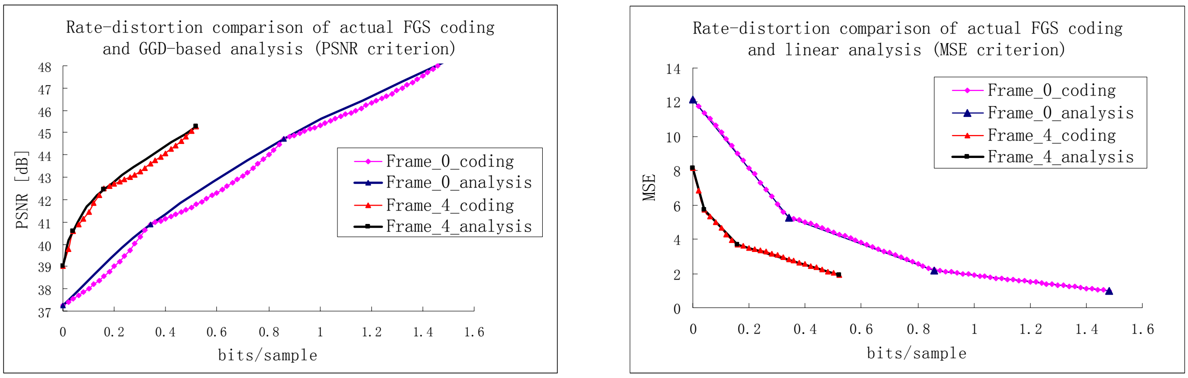
In this paper, since the cascading quantization is a nice method in a GOP [1], we propose smooth reconstruction between different GOPs and optimal extraction in a GOP. Firstly, we introduce the R-D analysis and models of scalable video coding in the authors' prior work. Then, with above R-D models, virtual GOP concept is introduced and a virtual-GOP-based packet scheduling algorithm is proposed to acquire the optimal packet scheduling sequence in a virtual GOP. Based on the packet scheduling algorithm and the R-D analysis of FGS EL, an effective and flexible D-R model is proposed to describe the D-R function of virtual GOP. Finally, using the virtual-GOP-based R-D model, a practical non-search algorithm is proposed for smooth quality reconstruction. Compared to the quality layer method, the reconstructed video quality is improved not only objectively but also subjectively.

[*] This work was supported by National Key Technology R&D Program under contract No. 2006BAH02A13 and China Postdoctoral Science Foundation 20060400030.

2. RATE-DISTORTION ANALYSIS OF SVC FGS CODING

2.1 Rate-Distortion Analysis and Modeling of SVC FGS EL

In our prior work of [5], we analyze the D-R function of SVC FGS EL with generalized Gaussian model and conclude the D-R function with PSNR criterion should be the left half of a concave function as a whole. After that, in [6], we further examine the sub-bitplane technology of SVC FGS coding and conclude that the D-R function with mean square error (MSE) criterion should be linear within a FGS layer. Considering the above two properties of SVC FGS coding, a piecewise linear model is proposed to describe the R-D function of FGS EL. Fig. 1 shows the distortion-rate functions and its models in the Foreman sequence.



(a) PSNR-based distortion-rate function

(b) MSE-based distortion-rate function

Fig. 1. The distortion-rate functions and its models with PSNR criterion (a) and MSE criterion (b) in the Foreman sequence.

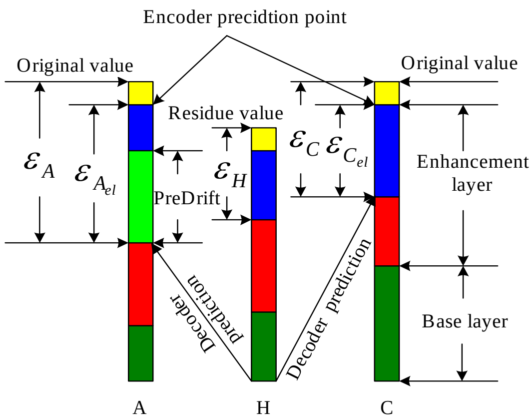


Fig. 2. The prediction structures of SVC hierarchical B coding.

2.2 Distortion Modeling of Reconstructed Frame with Drift

In the closed-loop coding with hierarchical B pictures, the current frame distortion is dependent on the encoder-decoder mismatch of reference frames and the reconstructed residue frame. Fig. 2 shows the typical prediction relation between the current frame and the reference frames using a pixel-based analysis. ε_A and ε_C show the differences between the original values and the reconstructed values in the reference frames *A* and *C*, respectively. $\varepsilon_{A_{el}}$ and $\varepsilon_{C_{el}}$ show the corresponding encoder-decoder mismatch values of the reference frames *A* and *C*, respectively. ε_H shows the difference between the original residue value and the reconstructed residue value in the current residue frame *H*. The pre-drift in Fig. 2 represents the drift of frame *A* caused by reference frames of *A*. All variables ε_A, ε_C, $\varepsilon_{A_{el}}$, $\varepsilon_{C_{el}}$, and ε_H may be positive or negative.

In [6], using a pixel-based analysis, a simple frame-based distortion model is proposed to estimate the reconstructed frame distortion with drift:

$$E(\varepsilon_B^2) = E(\varepsilon_H^2) + (p_{fwd} + \tfrac{1}{4} p_{Biwd})E(\varepsilon_A^2) + (p_{bwd} + \tfrac{1}{4} p_{Biwd})E(\varepsilon_C^2) + \mu E(\varepsilon_A^2)E(\varepsilon_C^2) + \nu \tag{1}$$

where $E(\varepsilon_B^2)$ is the reconstruction distortion of target frame *B*, $E(\varepsilon_H^2)$ is the distortion of FGS EL in frame *B*, $E(\varepsilon_A^2)$ and $E(\varepsilon_C^2)$ are the reconstruction distortion of reference frames *A* and *C* , and p_{fwd}, p_{bwd}, and p_{biwd} denote the fraction of forward-predicted, backward-predicted, and bi-predicted pixels in frame *B*. μ and ν are two parameter, which could be estimated using two decoding passes such as base layer extraction and full stream extraction. With the frame-based drift distortion model and the piecewise linear R-D model of FGS EL, we can estimate the reconstruction distortion of each frame just based on the allocated bit rate of each picture.

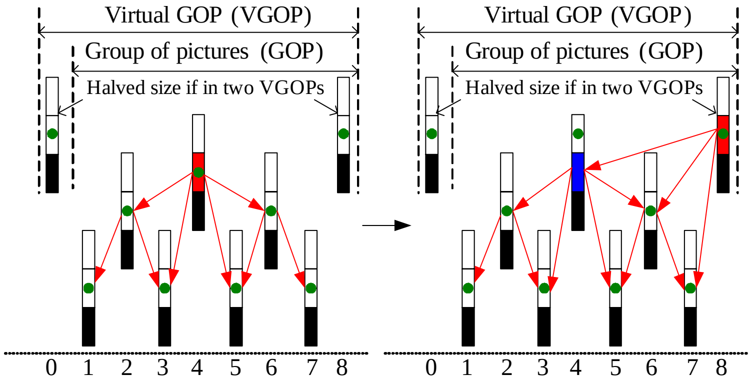


Fig. 3. Virtual GOP definition and the optimal scheduling algorithm.

3. VIRTUAL-GOP-BASED PACKET SCHEDULING AND RATE-DISTORTION MODELING

3.1 Virtual-GOP-based optimal packet scheduling algorithm

From a given FGS-coded bitstream and a specific average bit rate, there are many different ways to extract a sub-bitstream. The same target bit rate could be achieved by selecting and truncating different FGS levels (or packets). However, the coding efficiency is significantly influenced by the extraction scheme. Typically, the FGS levels should be assigned different priority according to the prioritization order, which is determined by the D-R slope of each FGS level. However it is not trivial to acquire the specific D-R slope of an FGS level since it is not only dependent on the properties of the FGS level, but also dependent on the definite scheduling sequence of the related FGS levels. For a GOP size 8 and 2 FGS layers, there need at least (8+7+6+5+4+3+2+1)*2=72 testing arrangements to achieve the optimal order even just in a GOP, not considering the interference of different GOPs. So it's not easy to acquire the optimal scheduling order in a video sequence.

According to the hierarchical *B* structure, the drift of FGS levels propagates within a GOP except the levels of the key pictures that could use the base layer as reference. So for simplicity of the priority setting in the entire bitstream, the optimal priority setting algorithm of this paper is confined in a GOP and the key picture levels is considered in a special way. The new concept of virtual GOP is proposed to show the target scheduled pictures and the special treatments of key pictures. Fig. 3 shows the virtual GOP definition in this paper, where the virtual GOP contains two consecutive key pictures and all the *B* pictures between the two key pictures. For the two-layers FGS coding with GOP size 8 in Fig. 3, the virtual GOP size is 9, and the number of FGS levels that should be considered for priority setting is 2*9=18. The FGS level size of a key picture that belongs to two virtual GOPs is halved in the virtual GOP. The optimal priority setting algorithm in a virtual GOP is described as follow:

1) Calculate the piecewise linear model of each frame *i* during the FGS encoding process as in Section 2.1. That is, H_i^l is derived, which represents the quantization error of residue frame *i* at FGS level *l*, where *l=0*

represents the quantization error of residue frame in the base layer. The parameters $p_{fwd,i}$, $p_{bwd,i}$ and $p_{Biwd,i}$ of frame i in Equation (1) are also collected during the encoding process.

2) Estimate the two parameters μ_i and ν_i of Equation (1) in each frame i. First, two extraction patterns of the base layer extraction and full layers extraction are selected and decoded to acquire the MSE of each frame. Then with the data H_i^l, $p_{fwd,i}$, $p_{bwd,i}$ and $p_{Biwd,i}$ in step 1), the two parameters μ_i and ν_i are calculated. For example, let E_i^b and E_i^f represent the decoded MSE of frame i in the base layer extraction and full layers extraction respectively. Then the drift model parameters μ_4 and ν_4 of frame 4 in Fig. 3 can be calculated by the following two equations:

$$\begin{cases} E_4^b = H_4^0 + (p_{fwd,4} + \frac{1}{4} p_{Biwd,4})E_0^b + (p_{bwd,4} + \frac{1}{4} p_{Biwd,4})E_8^b + \mu_4 E_0^b E_8^b + \nu_4 \\ E_4^f = H_4^2 + (p_{fwd,4} + \frac{1}{4} p_{Biwd,4})E_0^f + (p_{bwd,4} + \frac{1}{4} p_{Biwd,4})E_8^f + \mu_4 E_0^f E_8^f + \nu_4 \end{cases}.$$

3) Set the priority order and obtain the corresponding PSNR-rate slopes of FGS levels in each virtual GOP:

For each virtual GOP in the bitstream

A) Initialize the MSE array of the virtual GOP $MSE_{VGOP}[1..(gop\_size+1)]$ as the reconstruction MSE of the base layer, which is $[E_0^b, E_1^b, E_2^b, E_3^b, E_4^b, E_5^b, E_6^b, E_7^b, E_8^b]$ for the first VGOP (Results could be derived directly in step 2)). Initialize the level array as the 1st FGS level in the virtual GOP. In Fig. 3, the level array could be described as *array_level* =[(0,1), (1,1), (2,1), (3,1), (4,1), (5,1), (6,1), (7,1), (8,1)] (the green circle packets in Fig. 3), where (*,*) represents (*frame_no*, *FGS_level_no*).

B) Select continuously the best FGS level with the largest PSNR-rate slope to schedule until the *array_level* is NULL:

i) Calculate the distortion decrease of each FGS level in current *array_level*. In Fig. 3, for example, if the FGS level (4,1) is tested and added, the MSE decrease of frame 4 is first derived (that is $decMSE_{(4,1)}[4] = H_4^0 - H_4^1$, $decMSE_{(4,1)}[k \neq 4] = 0;$). After that, from temporal level 2 to temporal level 3, the MSE decrease of frame 2, 6, 1, 3, 5, 7 with FGS level (4,1) are calculated orderly using Equation (1). For example, the MSE decrease of frame 2 with FGS level (4,1) could be calculated as

$$\begin{aligned} decMSE_{(4,1)}[2] = MSE_{VGOP}[2] - (H_2^0 + (p_{fwd,2} + \tfrac{1}{4} p_{Biwd,2})(MSE_{VGOP}[0] - decMSE_{(4,1)}[0]) \\ + (p_{bwd,2} + \tfrac{1}{4} p_{Biwd,2})(MSE_{VGOP}[4] - decMSE_{(4,1)}[4]) \\ + \mu_2 (MSE_{VGOP}[0] - decMSE_{(4,1)}[0])(MSE_{VGOP}[4] - decMSE_{(4,1)}[4]) + \nu_2) \end{aligned}$$

After the process, the distortion decrease $decMSE_{(4,1)}[0..\ (gop\_size+1)]$ of the FGS level (4,1) in current virtual GOP could be derived, and so the PSNR increase of the FGS level (4,1) could be calculated as :

$$\begin{aligned} \mathrm{incPSNR}_{(4,1)} &= 10 mean(\log_{10}(MSE_{VGOP}[1..(gop\_size+1)])) \\ &- 10 mean(\log_{10}(MSE_{VGOP}[1..(gop\_size+1)] - decMSE_{(4,1)}[1..(gop\_size+1)])) \end{aligned}$$

ii) Calculate the PSNR-rate slope of each FGS level in the current *array_level*, and then select the FGS level with the largest PSNR-rate slope to schedule and output its corresponding PSNR-rate information.

iii) Update *array_level* and MSE_{VGOP}: The optimal FGS level in this cycle is removed from *array_level* and the higher level value of the corresponding frame is loaded in *array_level*. MSE_{VGOP} is updated with the optimal FGS level added. For example, in Fig. 3, if (4,1) is the best level of this cycle, now *array_level* is set as [(0,1), (1,1) , (2,1) , (3,1) , (4,2) , (5,1) , (6,1) , (7,1) , (8,1)] (the green circle packets in the right of Fig. 3), and MSE_{VGOP}=MSE_{VGOP}-$decMSE_{(4,1)}$.

4) Calculate the D-R slope of each FGS level: The level D-R slope of the key pictures in two virtual GOP is averaged, and the D-R slope of other packets is same as the one in the virtual GOPs.

In the above algorithm, two significant data are obtained. The first is the priority order of FGS levels in each virtual GOP, the other is the corresponding D-R information of the virtual GOP. The priority order and D-R information of the virtual GOPs is modeled and used for smooth quality reconstruction.

3.2 Rate-Distortion Modeling of Virtual GOP

In Section 2.1, we conclude that the derivative of D-R function (PSNR criterion) in FGS EL decreases as bit rate increases, that is, the D-R slope of lower-layer FGS packet should be steeper than the higher-layer FGS packet in the same picture. And in the algorithm of Section 3.1, the lower-layer packet is loaded first to the *array_packet*, and the packet with the highest D-R slope in the *array_packet* is always scheduled first until all the packets in the virtual GOP are scheduled. So, generally, we can assume that the derivative of the D-R function (PSNR criterion) in a virtual GOP decreases continuously as the rate increases. Then, as a similar method of our previous work in [7], the complete formula of the proposed R-D model in a virtual GOP can be described as follows:

$$PSNR(R) = a * R + A - (A - B)/(1 + b * R) . \quad (2)$$

where R is the bit rate (bit/sample), B is average PSNR of base layer in a virtual GOP, a, A and b are the control parameter. Usually, a and b could also be selected as constants for a coarse approximation.

To verify the virtual-GOP-based R-D model, three experiments are designed. In the first experiment, a is set as the constant 6.0 and b is set as the constant 8.0. The parameter A is acquired by the nonlinear least-squares data fitting. In the second experiment, b is set as the constant 8.0, and the other two parameters (a and A) of the asymptote are calculated using the same nonlinear least-squares data fitting. In the third experiment, all the three parameters (a, b and A) of the R-D model are acquired through the nonlinear least-squares data fitting. Table 1 shows the means of the maximum and average absolute estimate errors across all virtual GOPs in the eight test sequences. It is observed that the average absolute estimate error of the complete model on all the test videos is only about 0.037 dB in the third experiment. Generally, the proposed R-D model is accurate and flexible, and usually, two-control-parameter model is good enough to balance the accuracy and the complexity of the R-D model.

Table 1. The Average of Maximum and Average Absolute Estimate Error across the Virtual GOPs

Average across all GOPs	Foreman			Mobile			Bus			Football		
	Exp1	**Exp2**	**Exp3**	**Exp1**	**Exp2**	**Exp3**	**Exp1**	**Exp2**	**Exp3**	**Exp1**	**Exp2**	**Exp3**
Max error (dB)	0.182	0.151	0.147	0.419	0.160	0.160	0.468	0.129	0.092	0.722	0.144	0.063
Ave error (dB)	0.075	0.058	0.046	0.224	0.050	0.050	0.292	0.038	0.029	0.421	0.060	0.023
Average across all GOPs	**City**			**Crew**			**Harbour**			**Soccer**		
	Exp1	**Exp2**	**Exp3**	**Exp1**	**Exp2**	**Exp3**	**Exp1**	**Exp2**	**Exp3**	**Exp1**	**Exp2**	**Exp3**
Max error (dB)	0.210	0.201	0.179	0.397	0.117	0.107	0.616	0.207	0.113	0.433	0.132	0.092
Ave error (dB)	0.052	0.049	0.047	0.219	0.041	0.038	0.385	0.057	0.034	0.263	0.044	0.032

4. SIMPLE ALGORITHM FOR SMOOTH QUALITY RECONSTRUCTION (SASQR)

4.1 Simple Algorithm for Smooth Quality Reconstruction (SASQR)

In SVC coding, although the cascading of the quantization parameters over hierarchy levels results in relatively large PSNR fluctuations inside a group of pictures, subjectively, the reconstructed video appears to be smooth [1]. So it is nice to smooth the video quality between different GOPs. In the environments of virtual GOP model, the key problem is how to truncate virtual GOPs to both match the available average bandwidth $\bar{R}$ and achieve a constant quality D_{target} for each virtual GOP. To obtain the D_{target}, we need to calculate a combined function $C(D)$, which is constrained by $\bar{R}$:

$$C(D) = \frac{1}{N}\sum_{i=0}^{N-1} R_i(D) = \bar{R} , \quad (3)$$

where $R_i(D)$ is the R-D function of virtual GOP i, N is the number of smoothed virtual GOPs. Through (3), we can obtain the target constant quality $D_{target} = C^{-1}(\bar{R})$. Then the allocated bit rate of virtual GOP i can be calculated by $R_i(D_{target}) = R_i(C^{-1}(\bar{R}))$. However it's difficult to get a closed-form solution of C^{-1} for the known R-D models and a search algorithm for D_{target} is a burden for streaming server since the $\bar{R}$ changes continually in the actual streaming. Using the R-D model of virtual GOPs, where b is fixed as 8.0, the algorithm SASQR is described as follows:

1) Calculate the average distortion $\bar{D}$ with uniform bit rate allocation $\bar{R}$:

$$\bar{D} = \left[\bar{R} * \sum_{i=0}^{N-1} a_i + \sum_{i=0}^{N-1} A_i - (\sum_{i=0}^{N-1} A_i - \sum_{i=0}^{N-1} B_i)/(1 + 8.0 * \bar{R})\right] \Big/ N \,, \tag{4}$$

where a_i, A_i and B_i are the corresponding values of (2) in virtual GOP *i*.

2) Calculate the initial bit rate allocation $init\_rate_i = R_i(\bar{D})$, where $R_i(D)$ is the inverse function of (2) in virtual GOP *i*. Then we calculate the average tune bit rate:

$$tune\_rate = \frac{1}{N}(\sum_{i=0}^{N-1} init\_rate_i) - \bar{R} \,. \tag{5}$$

3) Calculate the tune weight of each virtual GOP *i*:

$$tune\_weight_i = N * [D_i^{'}(init\_rate_i)]^{-1} \Big/ \sum_{i=0}^{N-1} [D_i^{'}(init\_rate_i)]^{-1} \,, \tag{6}$$

where $D_i^{'}(init\_rate_i)$ is the derivative of the D-R function at the bitrate $init\_rate_i$. $[D_i^{'}(init\_rate_i)]^{-1} = R_i^{'}(\bar{D})$ approximates the bitrate requirement of one unit distortion change at the distortion point $\bar{D}$ of virtual GOP *i*.

4) Calculate the transmitting bit rate in a virtual GOP by

$$trans\_rate_i = init\_rate_i - tune\_rate * tune\_weight_i \,. \tag{7}$$

After acquiring $tran\_rate_i$ of each virtual GOP, the bit rate of each frame could be calculated as follows:

1) The allocated bits of the virtual GOP *i* is first calculated

$$bitsVGOP_i = tran\_rate_i * width * height * 1.5 * (gop\_size - 1 + KeyframesIn1VGOP + 0.5 * KeyframesIn2VGOP) \,, \tag{8}$$

where the *width* and *height* is the width and height of encoded pictures respectively, *KeyframesIn2VGOP* is the number of key frames included by this virtual GOP and another neighbor virtual GOP, and *KeyframesIn1VGOP* is the number of key frames that is included just by this virtual GOP.

2) According to the scheduling sequence, for each FGS packet *k* in virtual GOP *i*

if ($bitsPacket_k < bitsVGOP_i$), the allocated bits of the packet *k*'s frame is increased by the size of packet *k*, and $bitsVGOP_i -= bitsPacket_k$;

else the allocated bits of the packet *k*'s frame is increased by the remaining $bitsVGOP_i$.

At last, the allocated bit rate of key pictures in two virtual GOP is added together.

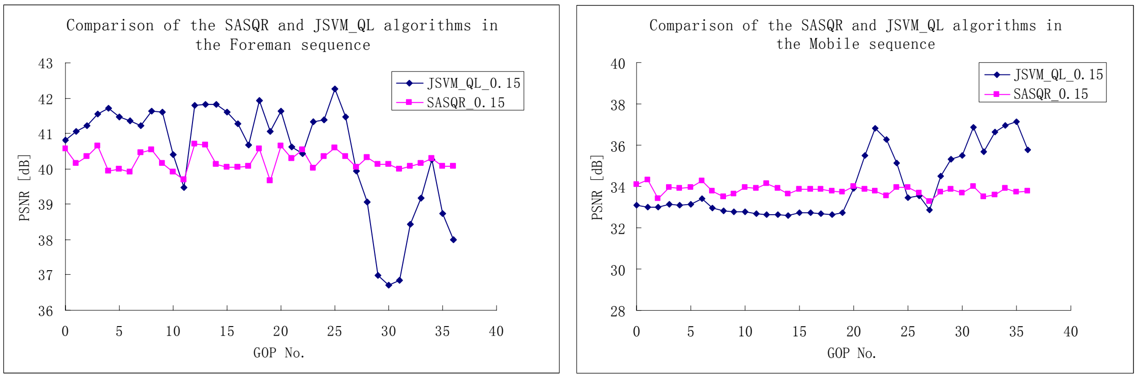


Fig. 4 Comparison of JSVM-9 default quality layer extraction and the proposed SASQR at the bit rate 0.15 bits/sample in the Foreman and Mobile CIF sequences.

4.2 Experimental Results

Smooth quality reconstruction is the favorable application. To validate the effectiveness of the SASQR, the SVC reference software (JSVM 9, version JSVM_9_0) [8] is used. Eight standard sequences (Foreman, Mobile, Bus, Football, City, Crew, Harbour, and Soccer) are encoded at CIF resolution. The SNR scalable configuration file of JVT-Q009 [9] is used with four parameters changed: the GOP size is set to 8, the FREXT mode is on for using adaptive 4*4 or 8*8 transform, the number of FGS Levels is set to 2, the prediction loop is operated at the highest FGS level. The target bit rate of FGS EL is 0.05 bits/sample (about 228 kbits/s), 0.10 bits/sample (about 456 kbits/s) and 0.15 bits/sample (about 684 kbits/s) respectively in the three experiments. The smoothed GOPs are from 0 to 36 in the Foreman, Mobile, City, Crew, Harbour, and Soccer sequences, from 0 to 31 in the Football sequence, and from 0 to 17 in the Bus sequence. Two different algorithms are applied in the experiments. The first is the proposed SASQR. The default quality layer method is also applied for comparison.

Table 2. Comparison of JSVM-9 Default Quality Layer Extraction and the Proposed SASQR in the Eight Standard CIF Sequences

Sequences		JSVM-9 QL(dB)			Proposed SASQR(dB)		
		Exp1	Exp2	Exp3	Exp1	Exp2	Exp3
Foreman	**Min**	34.88	36.58	36.71	36.92	38.55	39.66
	Max	39.44	41.66	42.27	38.33	39.96	40.69
	Var	1.01	2.32	2.32	0.14	0.10	0.08
Mobile	**Min**	29.61	31.67	32.60	30.15	32.39	33.30
	Max	33.54	33.82	37.15	32.10	33.39	34.33
	Var	2.10	0.24	2.47	0.13	0.04	0.05
Football	**Min**	33.32	33.97	35.37	35.43	36.41	37.48
	Max	43.85	44.77	44.77	39.65	39.98	40.31
	Var	7.66	6.01	4.19	0.71	0.40	0.22
Bus	**Min**	32.14	34.98	34.98	33.36	34.85	35.91
	Max	34.88	37.16	38.76	34.37	35.63	36.74
	Var	0.55	0.33	1.14	0.08	0.03	0.04
City	**Min**	35.76	39.08	39.13	37.55	39.47	40.44
	Max	39.31	41.77	41.80	39.00	40.41	41.36
	Var	1.21	0.71	0.63	0.11	0.05	0.05
Crew	**Min**	34.92	36.78	37.41	36.51	37.94	38.76
	Max	40.09	41.14	42.82	37.48	38.82	39.87
	Var	2.12	1.20	2.32	0.08	0.06	0.06
Harbour	**Min**	32.83	32.98	35.51	33.31	34.69	35.75
	Max	35.82	35.82	36.56	34.39	35.57	36.45
	Var	0.99	0.47	0.15	0.06	0.01	0.03
Soccer	**Min**	34.94	36.99	38.08	36.54	38.46	39.42
	Max	40.92	42.17	44.19	38.07	39.57	40.65
	Var	3.48	1.51	3.45	0.10	0.07	0.06

Fig. 4 compares the proposed SASQR with JSVM-9 default quality layer extraction at the bit rate 0.15 bits/sample in the Foreman and Mobile sequences, respectively. Table 2 shows the corresponding variation statistics of all the three experiments in the eight test sequences. It can be seen that, compared to the JSVM-9 default quality layer extraction, the smoother quality can be obtained with the proposed SASQR. Fig. 5 shows the comparison of subjective quality between JSVM-9 default quality layer extraction and the proposed SASQR at the bit rate 0.15 bits/sample in the frames 26 and 246 of the Foreman sequence. It can be seen that the subject difference of frame 26 between JSVM-9 default quality layer extraction and the proposed SASQR is little though the corresponding PSNR of JSVM-9 default method is 0.7 dB higher than the one of the proposed SASQR. However the subjective quality of frame 246 is improved noticeably in the proposed SASQR and the overall subjective quality is smoothed and improved.

5. CONCLUSION

In this paper, first, the virtual GOP concept is introduced and a virtual-GOP-based priority setting algorithm is proposed to obtain the optimal priority order of FGS levels in a virtual GOP. Second, a new efficient and flexible R-D model is presented for approximating the R-D function of virtual GOPs, with which a simple algorithm for smooth quality reconstruction is employed to improve the subjective quality of the reconstructed video. Extensive experiments show the effectiveness and efficiency of the models and algorithms.

(a) JSVM-9 default quality layer extraction

(b) The proposed SASQR

Fig. 5. Subject comparison of JSVM-9 default quality layer extraction (a) and the proposed SASQR (b) at the bit rate 0.15 bits/sample in the frame 26 and 246 of the Foreman CIF sequence.

REFERENCES

1. H. Schwarz, D. Marpe, and T. Wiegand, “Overview of the scalable video coding extension of the H.264/AVC standard,” *IEEE Trans. Circuits. Syst. Video Technol.*, vol. 17, no. 9, pp. 1103–1120, Sep. 2007.
2. J. Ridge and M. Karczewicz, AHG report: FGS applications and design simplification, Joint Video Team (JVT) of ISO-IEC MPEG & ITU-T VCEG, JVT-X006 Jul. 2007.
3. I. Amonou, N. Cammas, S. Kervadec, S. Pateux, “Optimized rate distortion extraction with quality layers in the scalable extension of H.264/AVC,” *IEEE Trans. Circuits. Syst. Video Technol.*, vol. 17, no. 9, pp. 1186–1193, Sep. 2007.
4. T. V. Lakshman, A. Ortega, and A. R. Reibman, “VBR video: Tradeoffs and potentials,” in *Proc. IEEE*, vol. 86, May 1998, pp. 952–972.
5. J. Sun, W. Gao, D. Zhao, “Statistical analysis and modeling of rate-distortion function in SVC fine-granular SNR scalable videos,” *IEEE International Conference on Multimedia. & Expo, ICME 2007*, Beijing, Jul. 2007.
6. J. Sun, W. Gao, D. Zhao, “Rate-distortion modeling and it’s application to quality layer assignment in SVC fine-granular SNR scalable videos,” *Picture Coding Symposium, PCS 2007,* Lisbon, Nov. 2007.
7. J. Sun, W. Gao, D. Zhao, and Q. Huang, “Statistical model, analysis and approximation of rate-distortion function in MPEG-4 FGS videos,” *IEEE Trans. Circuits Syst. Video Technol.*, Vol. 16, No. 4, pp. 535-539, Apr. 2006.
8. J. Reichel, H. Schwarz, M. Wien, and J. Vieron, Joint Scalable Video Model 9 of ISO/IEC 14496-10:2005/AMD3 Scalable Video Coding, Joint Video Team (JVT) of ISO-IEC MPEG & ITU-T VCEG, JVT-X202. Geneva, Jul. 2007.
9. M. Wien, and H. Schwarz, “AHG on Coding eff & JSVM coding efficiency testing conditions,” in Joint Video Team (JVT), Docs. JVT-Q009, Oct. 2005.