H.264/SVC Rate Allocation based on Graceful Degradation of Subjective Quality in Frame Rate Switching

Yu-Chieh Li, Zong-Yi Chen, Kai-Wen Liang, and Pao-Chi Chang
Dept. of Communication Engineering, National Central University, Jhongli, Taiwan

***Abstract* —This paper proposes a rate allocation method for Scalable Video Coding (SVC) temporal scalability based on subjective quality metric. The proposed method can gracefully lower the perceptual video quality by switching the frame rate under the situation of bandwidth fluctuation. Each temporal layer is measured by the subjective quality metric and allocated with a corresponding rate. The proposed method tends to increase the quality of low layer by sacrificing the high layer quality but human cannot perceive the degradation of high layer. Simulations show that the proposed method can efficiently allocate the rate of each temporal layer with closer subjective video quality when the bandwidth is insufficient. Compared with the JVT recommended method, the difference of subjective quality can be reduced from 4.03dB to 2.8dB.**

I. Introduction

With the improvement of video coding technology, storage capacity, display resolution, and CPU processing capability, the applications of multimedia systems become richer and more popular. Therefore, how to efficiently provide suitable video to users under different constraints is very important. Scalable video coding is one of the best solutions to this problem.

H.264/SVC [1], which is constructed based on H.264/AVC, is the current state-of-the-art scalable video coding standard. H.264/SVC contains multiple scalabilities so that not only it has high compression efficiency but also the encoded bitstream can be adapted to heterogeneous user/network environments without transcoding. Temporal scalability [2] can support multiple display frame rates (FR) with a wide range of bitrates. When the bandwidth is limited, the FR can be switched to satisfy the rate constraint. However, when adopting the JVT recommended QP setting [2] for H.264/SVC temporal layers, the SVC system exhibits a wide subjective quality gap between different layers in frame rate switching. The wide quality gap might be annoying when the FR switching occurs frequently. Thus how to efficiently allocate the bitrate among multiple temporal layers under a rate constraint to reduce the difference of subjective quality between different temporal layers is an important issue.

In the recent work [3], Cho *et al.* proposed a distortion model that takes dependency of temporal layers into consideration for temporal layer bit allocation. This distortion model results in a highly efficient bit allocation scheme, which outperforms the rate control algorithm in the JSVM 9.12 reference software codec [4]. However, PSNR, the distortion assessment used in the model, is not a subjective video quality metric which might not be able to truly reflect the distortion caused by frame rate variation [5]. The quality optimization of all temporal layers might lead to a considerable gap of subjective quality between layers. Therefore, in this work, we utilize the subjective quality metric (QM) [6], instead of the conventional objective measurement PSNR, to measure the video quality. Because the human eyes are more sensitive to the quality degradation of low FR than high FR, the proposed rate allocation scheme achieves graceful subjective quality degradation between different FRs by increasing the low temporal layer quality and decreasing the high layer quality.

The rest of this paper is organized as follows. An analysis of H.264/SVC hierarchical temporal layers is reported in Sec. II. Sec. III introduces the subjective quality metric used in this paper. Sec. IV describes the proposed rate allocation algorithm. The simulation results are shown in Sec. V. Finally, Sec. VI concludes this paper.

II. Hierarchical Temporal Layers of H.264/SVC

Frames in the lowest temporal layer are referred to key frames (which are typically I or P frames) in hierarchical B-pictures structure. A key frame and all the frames that are temporally located between two key frames are considered as a Group of Pictures (GOP). Within a GOP, frames are predicted in a dyadic structure as illustrated in Fig. 1. This structure can support temporal scalability better than the traditional ("IBBBBBBP…") coding structure.

For optimal overall coding efficiency, the quantization step, controlled by the Quantization Parameter (QP), differs in each temporal layer. According to the encoding order, the distortion of lower temporal layers will propagate to higher layers, so lower temporal layers are more important than higher layers by considering the ME references. Thus key frames typically have the lowest QP values and the highest PSNRs. And QP values often increases with temporal layers.

Coding experiments with H.264/SVC [2] suggest the following QP settings.

$$QP_T = QP_0 + 3 + T \qquad (1)$$

where T means the temporal layer, QP_0 is the QP of the lowest temporal layer.

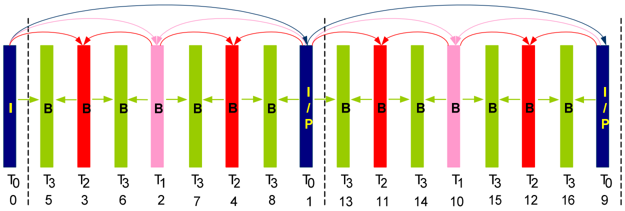

Fig. 1 Hierarchical B-pictures structure (GOP=8).

However the QP settings mentioned above result in subjective quality, i.e. QM, (which will be introduced in next section) differences between different FRs up to 4 dB depending on the video content. The large difference might be annoying if the FR switches frequently. The simulation result of Bus sequence is shown in Fig. 2, as an example.

By using JVT recommended setting, we observe that the perceptual quality varies widely in frame rate switching. In particular, lower QP value causes larger QM difference.

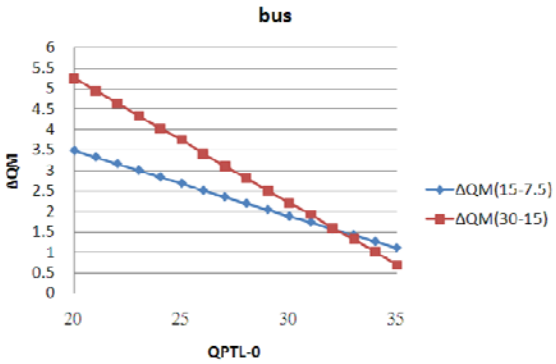

Fig. 2 QM variation with FR switching.

III. Subjective Quality Metric

Feghali *et al.* proposed a subjective quality metric (QM; Quality Metric) as follows [6].

$$QM = PSNR + c_1 m^{c_2}(30 - FR) \tag{2}$$

where c_1=0.986 and c_2=0.378. m is motion speed and is sequence dependent, which is the normalized largest 25% motion vectors in average. TABLE I shows the values of m for six video sequences. These values correspond well to the perceived motion speeds in these sequences. In (2), the full frame rate is 30fps. When the FR is 30, this metric degrades to the conventional objective quality metric, i.e. PSNR. If the FR becomes lower than 30, the second term is used to compensate for the PSNR term to provide a subjective quality metric close to human eyes.

IV. Proposed Rate Allocation Algorithm

The purpose of this paper is to propose a rate allocation mechanism among temporal layers to provide graceful degradation of subjective quality in frame rate switching. The proposed scheme is shown in Fig. 3. The procedure is described as following steps.

TABLE I
Normalized Average Magnitudes of Large Motion Vectors *m*

sequence	Motion; *m*
soccer	0.0527
harbour	0.0061
bus	0.0479
foreman	0.0158
ice	0.0261
crew	0.0177

Step 1: Given a target rate, calculate the rate of base layer by the estimation $R_{3.75}$ = target rate $\times R_0$ ratio, (detailed in Sec IV-A) where the suffix denotes the frame rate and the lowest frame rate is 3.75.

Step 2: Determine the corresponding $QP_{3.75}$ and $QM_{3.75}$ using R-Q model (detailed in Sec IV-B) and R-D model (detailed in Sec IV-C).

Step 3: Determine the $QM_{7.5}$, QM_{15}, QM_{30} using the following criterion to achieve graceful subjective quality degradation and meet target rate constraint ($R_{30}\leq$ target rate).

- Criterion for choosing the closest QM sets in all combinations: sort$\sqrt[3]{QM_{7.5}\cdot QM_{15}\cdot QM_{30}}$ (3)
  → select the top 3 largest values
- Criterion for choosing the best quality from the 3 candidates: $\max(QM_{7.5}+QM_{15}+QM_{30})$ (4)
  → select the largest value

Step 4: Repeat *Step1-3* by adjusting R_0 *ratio* to find the best QP setting.

Step 5: Find corresponding $QP_{7.5}$, QP_{15}, QP_{30} to encode video by H.264/SVC encoder.

The block of loop of Enhancement layer quality assignment follows the following rule for all possible combinations.

$$\Delta 1 = QM_{7.5} - QM_{3.75}; \Delta 2 = QM_{15} - QM_{7.5}; \Delta 3 = QM_{30} - QM_{15}$$

search $0 \leq \Delta 1 \leq \Delta 2 \leq \Delta 3 \leq 5$ with increment 0.2.

A. *Rate allocation of initial rate*

In our algorithm, the quality of 3.75 frame rate is not taken into consideration because it is hardly used in practical applications. But it is still critical to decide how many bits are allocated to the base layer. In our proposed scheme, the initial rate of $R_{3.75}$ is determined by encoding the first 16 frames (2 GOPs) of each sequence with JVT recommended method. Then R_0 ratio can be calculated by

$$R_0\ ratio = R_{3.75} / Total\ rate. \tag{5}$$

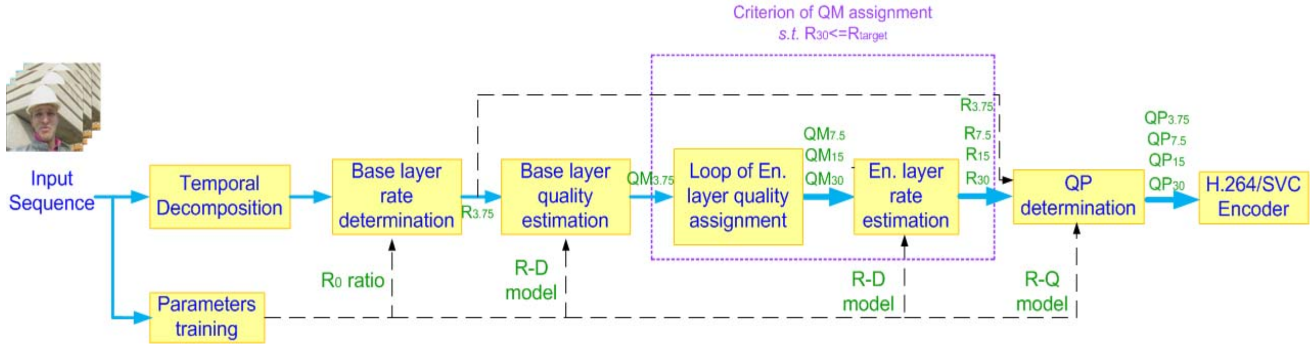

Fig. 3 The proposed rate allocation scheme.

Figure 4 shows an example of the R_0 ratios for four videos under different total rates. Once we get the R_0 ratio through the pre-encoding, we can decide the initial $R_{3.75}$ for given target rate.

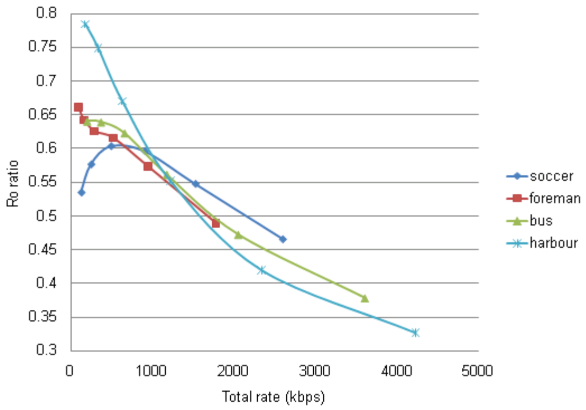

Fig. 4 R_0 ratio under different total rates.

B. *Rate-QP (R-Q) model*

A simple and fast method was proposed in the literature [7] to decide the starting value for QP for each part of the video sequence. R-Q curves can be modeled with a logarithmic equation as follows.

$$QP = a \cdot \ln R_c + b \qquad (6)$$

where a and b are sequence dependent constants, R_c is the rate of current layer individually. For a given rate, we can get the QP value using the R-Q model.

Although this model is not proposed for H.264/SVC, it still can be used in our work because of the rate independent characteristic. The rate of each layer is independent and is dependent on its own QP in H.264/SVC temporal scalability. Figure 5 shows a simulation result of rate dependency of two temporal layers, and similar results can be observed in different layers and videos. From Fig. 5 we can find that the bit rate of layer TL-1 is mainly determined by its own QP and it is independent of the bit rate of its reference layer, TL-0. In this work, R-Q model is determined by encoding the first 16 frames (2 GOPs) with recommended method. Fig. 6 shows the modeled result and each equation QPx in Fig.6 denotes the R-Q model of xth temporal layer.

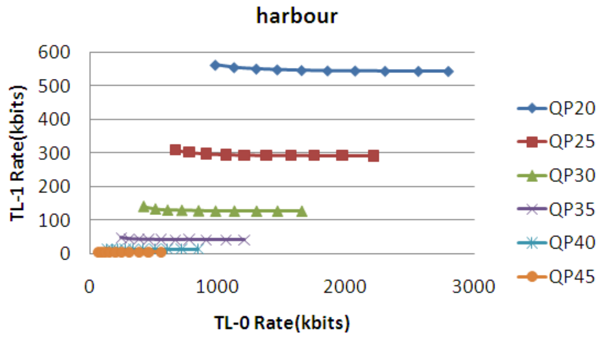

Fig. 5 Rate independent illustration of rate dependency, where the x-axis is the bit rate of layer TL-0 (reference layer) and the y-axis is the bit rate of layer TL-1 (dependent layer).

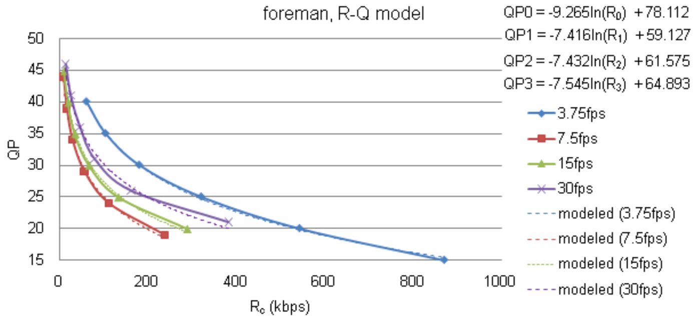

Fig. 6 Results of R-Q model.

C. *Rate-QM (R-D) model*

R-D curves can be modeled with a logarithmic equation by curve fitting.

$$QM = c \cdot \ln R_t + d \qquad (7)$$

where c and d are sequence dependent constants, R_t includes the rates of current layer and all previous layers. The same as R_0 ratio and R-Q model, R-D model is determined by encoding the first 16 frames (2 GOPs) with recommended method. The modeled result is shown in Fig. 7 and each equation QMx denotes the R-D model of xth temporal layer. We encode 2 GOPs to get all the training parameters (R_0 ratio、a、b、c、d).

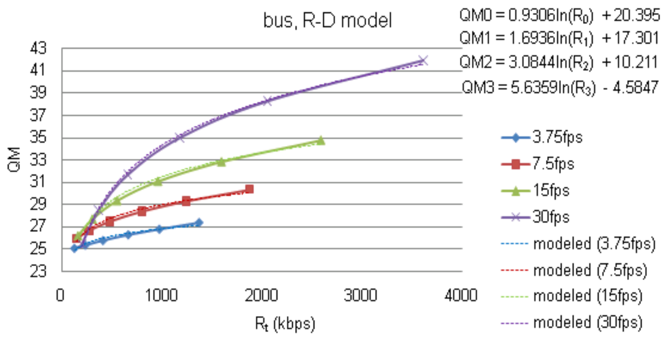

Fig. 7 Results of R-D model.

TABLE II
SIMULATION ENVIRONMENT

Software	JSVM 9_16
Sequence Name	Harbour、Soccer、Bus、Foreman
Resolution	CIF
Sequence Type	IBBBB…I
GOP Size	8
Intra Period	8
Frame Rate	30、15、7.5 、3.75fps
Number of Reference Frame	1
Number of Encoder Frames	150

V. SIMULATION RESULTS

Our scheme is implemented on JSVM 9.16. The test condition is shown in TABLE II. In our experiments, four standard test sequences including Soccer, Bus, Foreman and Harbour have been tested. The performance assessments in our experiments include the bitrate, *QM*-sum and △*QM* which are defined as follows.

$$\text{QM-sum} = QM_{7.5}+QM_{15}+QM_{30}$$

$$\Delta QM\,(15\text{-}7.5)=\text{Proposed}(QM_{15}\text{-}QM_{7.5})\text{-Recom.}(QM_{15}\text{-}QM_{7.5})$$

$$\Delta QM\,(30\text{-}15)=\text{Proposed}(QM_{30}\text{-}QM_{15})\text{-Recom.}(QM_{30}\text{-}QM_{15})$$

The performance of our proposed algorithm compared with recommended method is shown in TABLE III and TABLE IV. Soccer and Foreman also have the similar results. First we can observe that the bitrate of the original recommended method and our proposed method will exceed the target bitrate in some cases. This is because JSVM 9.16 does not have the function of rate control for enhancement layers, thus we choose the encoded result with bitrate closest to the target bitrate. And in our proposed algorithm, we determine the QP setting for the encoder by the R-D and R-Q models before encoding the total frames. Although the selected QP setting leads to target bitrate exceeding in some cases, the amount is very small. This method avoids encoding video multiple times and is more appropriate for real applications.

TABLE III
SIMULATION RESULTS OF BUS

Sequence	Frame rate	Target bitrate (Kbit/s)	Recom. Total Rate, R_t (Kbit/s)	Recom. QP	Recom. QM	Recom. QM-sum	Proposed Total Rate, R_t (Kbit/s)	Proposed QP	Proposed QM	Proposed QM-sum	compare two method ΔQM
Bus	3.75	1400	761.045	23	26.656	97.578	893.94	21	26.875	97.615	15-7.5
	7.5		913.669	27	29.076		1166.422	23	29.620		-0.437
	15		1123.784	28	32.085		1266.638	33	32.192		30-15
	30		1396.498	29	36.417		1390.862	34	35.804		-0.72
Bus	3.75	1250	701.953	24	26.553	96.392	830.368	22	26.774	96.577	15-7.5
	7.5		834.319	28	28.89		1034.708	25	29.352		-0.001
	15		1015.04	29	31.738		1186.379	30	32.199		30-15
	30		1249.251	30	35.764		1248.946	39	35.026		-1.2
Bus	3.75	800	499.437	28	26.158	91.703	592.671	26	26.347	92.109	15-7.5
	7.5		573.073	32	28.159		708.864	29	28.576		-0.13
	15		670.962	33	30.364		773.531	36	30.651		30-15
	30		801.235	34	33.18		829.285	40	32.882		-0.585
Bus	3.75	650	410.054	30	25.965	89.397	449.231	29	26.057	90.126	15-7.5
	7.5		465.64	34	27.804		537.842	31	28.282		-0.178
	15		538.698	35	29.692		600.368	36	29.991		30-15
	30		638.302	36	31.901		664.104	39	32.052		-0.148

TABLE IV
SIMULATION RESULTS OF HARBOUR

Sequence	Frame rate	Target bitrate (Kbit/s)	Recom. Total Rate, R_t (Kbit/s)	Recom. QP	Recom. QM	Recom. QM-sum	Proposed Total Rate, R_t (Kbit/s)	Proposed QP	Proposed QM	Proposed QM-sum	compare two method ΔQM
Harbour	3.75	1700	847.562	22	27.325	99.882	912.196	21	27.433	99.704	15-7.5
	7.5		1109.133	26	30.152		1360.085	22	30.642		-0.738
	15		1415.168	27	33.411		1435.538	34	33.162		30-15
	30		1715.429	28	36.32		1713.258	29	35.900		-0.172
Harbour	3.75	1200	670.674	25	27.003	96.136	716.642	24	27.092	96.796	15-7.5
	7.5		826.817	29	29.533		1023.12	25	29.988		-0.116
	15		992.101	30	32.264		1125.763	32	32.602		30-15
	30		1158.794	31	34.34		1186.696	36	34.207		-0.471
Harbour	3.75	700	459.25	29	26.539	91.378	510.677	28	26.663	91.873	15-7.5
	7.5		528.499	33	28.691		644.021	30	29.065		-0.245
	15		599.419	34	30.796		675.536	39	30.925		30-15
	30		680.803	35	31.891		723.491	38	31.882		-0.138
Harbour	3.75	550	386.477	31	26.339	89.26	418.62	30	26.430	89.514	15-7.5
	7.5		429.676	35	28.32		461.548	35	28.430		-0.022
	15		479.237	36	30.14		502.829	37	30.229		30-15
	30		536.952	37	30.8		550.291	38	30.855		-0.034

We can also observe that the QP settings obtained by our method tend to have lower QP in lower layers and have higher QP in higher layers compared with the recommended ones. And from the last columns we can see the proposed method reduces the subjective quality gaps between all different temporal layers and different target bit rates. Although our proposed method results in larger QP fluctuation, it is hardly to be perceived from the encoded video. The proposed algorithm efficiently allocates the rate for each temporal layer with closer subjective video quality in frame rate switching. In addition, the proposed method achieves similar or even better R-D performance compared with the recommended method, as shown in Fig. 8.

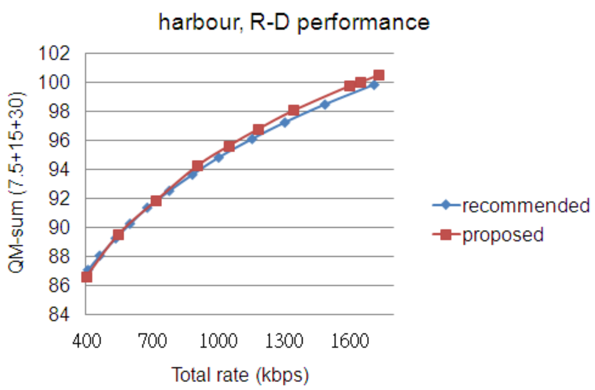

Fig. 8 R-D performance of harbour sequence.

VI. CONCLUSION

In this paper, we proposed a rate allocation method for SVC temporal scalability. We utilize the subjective quality metric, instead of the conventional objective measurement to measure the video quality. Under various bit rate constraints, we can achieve closer subjective visual quality in terms of QM for different frame rates. In the future, spatial scalability can be integrated in our proposed scheme.

REFERENCES

[1] H. Schwarz, D. Marpe, and T. Weigand, "Overview of the Scalable Video Coding Extension of the H.264/AVC Standard," *IEEE Trans. Circuits Syst. Video Tech.*, vol. 17, no. 9, pp. 1103-1120, Sep. 2007.

[2] H. Schwarz, D. Marpe, and T. Wiegand, "Hierarchical B Pictures," Poznan, Poland, Doc. JVT-P014, July 2005.

[3] Y. Cho, J. Liu, D. K. Kwon and C. C. Jay Kuo, "H.264/SVC Temporal Bit Allocation with Dependent Distortion Model," in *Proc. ICASSP 09*, Taipei, Taiwan, Apr. 2009, pp. 641 – 644.

[4] ISO/IEC JTC1, "Joint Scalable Video Model JSVM-12," ISO/IEC JTC1/SC29/WG11 and ITU-T SG16 Q.6, Doc. JVTY202, Oct. 2007.

[5] D. Wang, F. Speranza, A. Vincent, T. Martin, and P. Blanchfield, "Towards optimal rate control: a study of the impact of spatial resolution, frame rate, and quantization on subjective video quality and bit rate," in *Proc. Visual Communications and Image Processing 2003*, Lugano, Switzerland, July 8-11, 2003, pp. 198-209.

[6] R. Feghali, F. Speranza, D. Wang, and A. Vincent, "Video quality metric for bit rate control via joint adjustment of quantization and frame rate," *IEEE Trans. on Broadcasting*, vol. 53, no. 1, pp. 441–446,Mar. 2007.

[7] L. Czuni, G. Csaszar, and A. Licsar, "Estimating the Optimal Quantization Parameter in H.264," *IEEE International Conference on Pattern Recognition (ICPR)*, Sep. 2006, pp. 330-333.