Tao ZHU[1], Xiong-wei ZHANG[2]

Institute of Communications Engineering, PLA Univ. of Sci. & Tech. (1), Institute of Command Automation, PLA Univ. of Sci. & Tech. (2)

Frame distortion estimation and its application to extraction priority assignment for MGS of H.264/SVC

Abstract*. Based on research of frame distortion estimation, an extraction priority assignment method is proposed for the MGS bit-stream. Firstly, extraction priority is derived from rate distortion optimization method. Then, frame distortion estimation is performed from estimation of coefficients dropping distortion and drift distortion. At last, depending on quality layer concept, extraction priorities are assigned efficiently. Experimental results show that the proposed method outperforms the priority assignment method and bit-stream extraction of H.264/SVC.*

Streszczenie. *Zaproponowano metodę ekstrakcji pierwszeństwa dla MGS bit-stream bazującą na określaniu zakłócenia ramki. Najpierw ekstrakcja pierwszeństwa jest wyprowadzona z metody optymalizacji zniekształcenia szybkości. Następnie określa się zniekształcenie ramki ze współczynników zniekształceń kroplowych i zniekształceń zera. Wreszcie zależnie od koncepcji pasm pierwszeństwo jest określane ostatecznie. Rezultaty eksperymentu wykazały, że proponowana metoda może być zastosowana do określania pierwszeństwa i ekstrakcji strumienia bitów w H.264/SVC. (****Określanie zniekształceń ramki i jego wykorzystanie do ekstrakcji priorytetu w H.264/SVC****)*

Keywords: scalable video coding; extraction priority assignment; drift distortion; Medium Grain Scalability (MGS)
Słowa kluczowe: skalowane kodowanie video, określanie priorytetu, MSG, H.264 .

1 Introduction

The rapid development of modern communication system promotes various video application deployments ranging from mobile phones, peer to peer video streaming, to high-end video processing workstations. In these complex communication environments with heterogeneous network and time-varying bandwidth, the desire for flexible video adaptation device of once-encoded bit-stream, as depicted in Fig. 1, produces the newest scalable video coding specification, the scalable extension of H.264 advance video coding standard (H.264/SVC for short).

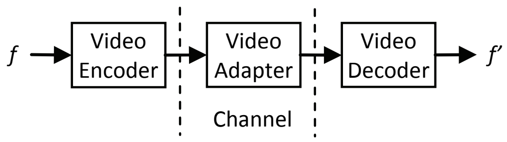


Fig. 1 Adaptive video transmission system

As an important technology to achieve video adaptation, bit-stream extraction is a post-encoding processing essentially, which adapt the scalable video bit-stream to the target rate and/or spatial and temporal resolutions on the basis of syntax and semantics in scalable video coding standard. Besides the scalability information in Network Abstraction Layer Unit (NALU) header, extraction priorities can be delivered through the syntax element priority_id in NALU header or the Supplemental Enhancement Information (SEI) message. Before extraction NALUs in bit-steam would be ordered according to the extraction priority and the extracted bit-steam would be produced. Hence, the calculation and assignment of extraction priorities play a key role in bit-steam extraction.

Initially, extraction priority was assigned to Coarse Grain Scalability (CGS) bit-stream to increase its rate extraction points [1,2,3,4]. Then, priority assignment can be performed for SVC bit-stream with multi-spatial layers [5]. Lim et al. [6] treated multi-spatial layers SVC bit-stream as a tree structure and assigned extraction priority. Truong et al. [7] proposed a frame of controlling quality scalability in multi-spatial layers and focused on the priority assignment to Fine Grain Scalability (FGS) bit-stream. Thomas et al. [8] gave out a backward drift estimation method and apply it to the priority assignment. Jun et al. [9] assigned extraction priority based a rate distortion model for FGS bit-stream. In [10], the child frame drift distortion of MGS bit-stream extraction is described by the distortion of its parent frames accurately to be used to calculate priority, but the parameters of this method need to be decided by fitting with lots of empirical data.

Up to now, H.264/SVC provides three kinds of scalability: spatial, temporal, and quality scalability to support bit-stream extraction. MGS is an important type of quality scalability. With the key picture concept and coefficient partitioning mechanism, MGS can provide flexible video adaptation operation while keeping acceptable coding complexity increase. Using coefficient partitioning mechanism, MGS bit-stream video adaptation can be performed by extracting transform coefficient slices with small Zig-Zag order from bit-stream and dropping other slices. Obviously, mismatch of motion-compensated prediction loops at encoder and decoder, namely drift, takes place and drift distortion would be propagated to next few frames after MGS bit-stream extraction. Consequently, frame distortion has to be taken into consideration during rate distortion optimized priority assignment.

The rest of this paper is organized as follows. In section 2, MGS with key picture concept and coefficient partitioning is described in detail and its extraction priority assignment is analyzed. Then, frame distortion estimation of MGS coefficient slice extraction and its application to extraction priority assignment are presented in Section 3. In Section 4, simulation results are illustrated to verify our method. Finally, we give out the conclusion.

2 MGS and existing priority assignment in H.264/SVC

The key picture and coefficient partitioning are two important concepts in MGS of H.264/SVC. Combined with hierarchical prediction structures of temporal scalability, the key picture concept can achieve reasonable trade-off between enhancement layer coding efficiency and drift as depicted in Fig. 2. All pictures of the lowest temporal layer are encoded as key pictures and the motion compensated prediction of them uses the base quality reconstruction of the reference pictures as reference data. And all non-key pictures in a Group of Pictures (GOP) employ the highest available quality reconstruction for their motion compensated prediction. Generally speaking, the base quality packets of target spatial layers should be reserved after extraction. In this wise, drift distortion propagation is constrained in a GOP due to the resynchronization of key pictures between encoder and decoder reconstruction.

Coefficient partitioning mechanism can be illustrated as Fig. 3. Originally, all coefficients of every 4x4 block in a slice constitute a quality refinement layer. After coefficient partitioning, one quality refinement layer is split into several quality refinement layers. Apparently, the granularity of quality scalability is increased.

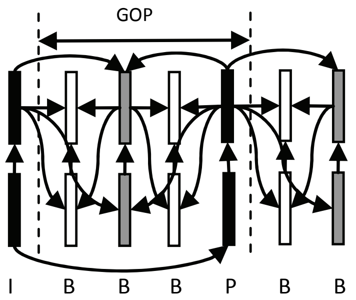


Fig. 2 The key pictures concept of MGS

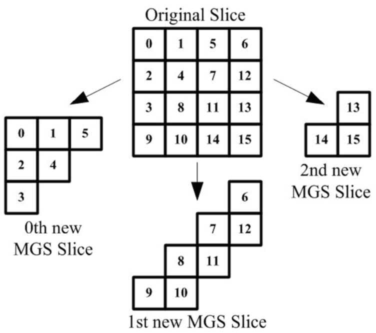


Fig. 3 Coefficient partitioning mechanism

Joint Software Video Model (JSVM) 9.18, which is the reference software implement of H.264/SVC, offers a software bit-stream extractor “BitStreamExtractorStatic” to support two kinds of extraction priority information [11]. One is based on the basic scalability information, namely scalability identifier (dependency_id, temporal_id, quality_id), which is short for (DId, TId, QId), carried in NALU header; the other one employs the rate distortion characteristic of NALU. The former information assumes the uniform statistical distribution of distortion for coded picture, can not reflect the actual features of coded picture and accordingly produces poor extraction performance. However, the latter information can express the significance of every NALU in the rate distortion sense, which can be used to sort the order of NALUs. And then, NALUs with close significance are organized to constitute a Quality Layer (QL), which an extraction priority is assigned to. As stated in this paragraph, the keys of rate distortion optimized assignment are the rate and the distortion effect for bit-stream of every NALU. Rate can be attained from the bits of NALU easily. Nevertheless, the calculation of distortion effect for bit-stream has complicated relation with the dependency between coding units and hence is a tough work.

The key picture concept can limit the drift distortion into a GOP and makes the calculation of distortion effect easier. Because the distortion effect is defined as the difference of distortion before and after dropping a few of NALUs during extraction, the easiest idea for calculating distortion effect is decoding this two bit-streams and gains their distortions. Due to key picture concept, this kind of decoding can be performed in a GOP. At present, JSVM 9.18 uses this method on the unit of a Quality Level instead of a NALU. As illustrated in Fig. 4, on the basic reconstruction quality, namely reconstruction using packets with QId = 0, enhancement packets with same QId are increased to frames in the same temporal level and decoding is performed in the dashed line box of every GOP. At last distortion effect for a quality level is calculated.

Accordingly, above calculation method has two disadvantages. The first is that the priority is calculated for a Quality Level, not for the basic unit of bit-stream, NALU, leading to coarser extraction granularity probably. The second is the highly computationally intensive decoding operation.

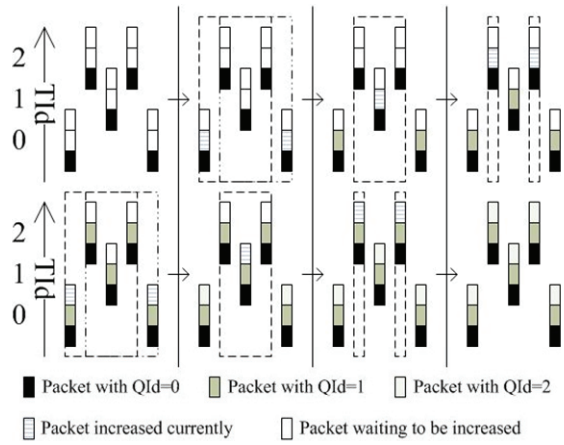


Fig. 4 Distortion effect calculation in JSVM

3 Extraction priority assignment based on distortion estimation

3.1 Rate distortion optimized extraction priority calculation

Rate distortion optimized bit-stream extraction is a constrained optimization essentially. Assume that *S* represents the set of NALUs that constitutes the whole bit-stream; *s* represents a decodable subset that belongs to *S*; *D*(*s*) and *R*(*s*) are reconstruction distortion and rate of *s*. Rate distortion optimized bit-stream extraction can be modeled as

(1) $$\min_{s\in S} D(s) \quad s.t. \quad R(s)\le R_T .$$

Using the Langrange multiplier method, above problem can be converted to an unconstrained optimized problem as next equation.

(2) $$\begin{cases} s^* = \min\limits_{s\in S} J(s) \\ J(s) = D(s) + \lambda R(s) \end{cases} .$$

Let the derivative of *J*(*s*) with respect to *s* equal to 0 and we can get

(3) $$|\lambda| = \left|\frac{\partial D/\partial s}{\partial R/\partial s}\right| .$$

The meaning of equation (3) is that when the subset *s* is changing the amount of distortion corresponds to every bit. Hence, the significance of variational NALUs set ə*s* is given out by equation (3). So, equation (3) can be used as the extraction priority. The number of ə*R* can be calculated from the bits of variational NALUs set. However, the calculation of ə*D* is a complicated work, which needs a deep investigation about the frame distortion during bit-stream extraction.

3.2 Extraction distortion analysis and estimation

If researching into MGS bit-stream decoding operation deeply, without taking the inter-layer prediction in MGS into consideration, the procedure of introducing corrupted error for every frame can be expressed as

(4) $$\begin{aligned} e_c = &\operatorname{Clip}\left\{\operatorname{Rnd}\left[I_{prd}\right]+\operatorname{Rnd}\left[\operatorname{IT}\left(X_{res}\right)\right]\right\} \\ &-\operatorname{Clip}\left\{\operatorname{Rnd}\left[I'_{prd}\right]+\operatorname{Rnd}\left[\operatorname{IT}\left(X'_{res}\right)\right]\right\} \end{aligned},$$

where I_{prd} and I'_{prd} are predicted value before and after drift, X_{res} and X'_{res} are rescaled residual coefficient before and after extraction, Clip( ), Rnd( ), IT( ) represent the operation of clip, round and inverse transform. Actually, the probability for the happening of clip and round is very small and the error magnitude aroused by them is also very small. Hence, for simplicity in the approximation of (4), the effect of clip and round can be ignored as

(5) $$\begin{aligned} e_c \approx \tilde{e}_c &= I_{prd} - I'_{prd} + \operatorname{IT}\left(X_{res} - X'_{res}\right) \\ &= e_{dft} + e_{drp} \end{aligned}.$$

From (5), it can be concluded that if coefficient extraction happens at every frame total error mainly contains drift error from predicted data and the coefficient dropping error. Generally e_q, e_{dft} and e_{drp} can be thought as error with mean of zero; hence, the total distortion of Mean Square Error (MSE) for a corrupted frame can be approximated as

(6) $$\begin{aligned} D_{Tot} &\approx E\left(e_q^2\right) + E\left(e_{dft}^2\right) + E\left(e_{ext}^2\right) \\ &= D_q + D_{dft} + D_{drp} \end{aligned},$$

where D_q, D_{dft} and D_{drp} are quantized distortion, drift distortion and coefficient dropping distortion respectively.

In H.264, block with 4x4 pixels (short for 'block') is the minimum unit for motion estimation, Intra-prediction and transform coding. Hence, the dropped coefficient for a block is our first focus, and then we turn our attention to two propagation behaviors of Intra-coding and Inter-coding.

3.2.1 Coefficient slice dropping distortion

Because Discrete cosine transform (DCT) is a normalized orthogonal transformation and coefficients with same Zig-Zag locations are dropped for all block in a frame during extraction, according to Parseval's Theorem, coefficient dropping distortion of a frame can be estimated as

(7) $$D_{drp} = \frac{1}{MN}\sum_{i=0}^{M-1}\sum_{j=\mathrm{minZZId}}^{\mathrm{maxZZId}} X_{i,j}^2,$$

where minZZId and maxZZId are indices of dropped coefficient in Zig-Zag order, $X_{i,j}$ is the rescaled coefficient, N is the size of block, M is the number of block in frame. To avoiding overlap with quantized error the rescaled coefficient is used to compute coefficient dropping distortion.

3.2.2 Inter-frame distortion propagation

Due to the elaborate procedure of motion compensation and the non-uniform distribution of error, the Inter-frame propagation distortion can't be treated over-macroscopically or over-microcosmically. If we only research the average frame distortion macroscopically, detailed traits of propagation would be missed definitely; and if the detail of propagation is only researched, the complexity is very high because of the tiny unit for estimation and the behavior of propagation is hard to model. Consequently, these two aspects must be trade-off and considered synthetically. In this subsection, we examine the average distortion change between two frames at first, and then take a deep insight into the detail.

Since motion compensated Inter-frame coding can be assumed as an AR (1) source model in temporal domain [12], taking the prediction structure of B frame into consideration the distortion is propagated to subsequent frame in a linear manner, i.e.

(8) $$D_{dft}(n) = \eta\left(\omega_1 D_{Tot1} + \omega_2 D_{Tot2}\right),$$

where $D_{dft}(n)$ is drift distortion of nth frame, D_{Tot1}, $_{\mathrm{Tot2}}$ are total distortion of the parents frames of nth frame, $\omega_{1,2}$ are prediction weights, η is propagation parameter with the range $\eta \geq 0$. Obviously, although (8) is a very simple model, η is hard to be decided because of its dependency on the characteristics of motion compensation and Intra-prediction for different video codec. However, the distortion propagation at frame level can be considered as the average effect of all block in the frame. Therefore, η can be approximated by the average propagation parameter of all blocks in a frame. Hence, the Inter-propagation distortion can be expressed as

(9) $$D_{dft}(n) \approx \left(\frac{1}{N}\sum_{i=0}^{N-1}\eta_i\right)\cdot\left(\omega_1 D_{Tot1} + \omega_2 D_{Tot2}\right),$$

where N is block number in a frame. Next, through analysis of the error elimination of leaky prediction an approximation of η for Inter-block is made. And the estimation of η for Intra-block is discussed.

3.2.3 Propagation parameter setting for Inter-block

Using reconstructed data as reference data achieves synchronization between encoder and decoder; moreover, essentially, it is an implement of well-known leaky prediction (LP), which optimizes motion compensation by minimizing prediction residual. LP has an important characteristic of drift error elimination which increases the robustness of video codec. LP in motion compensation can reduce drift error during Inter-propagation from frame to frame. Hence, error elimination of LP is one of the key factors that affect the error propagation. In H.264 LP is introduced by spatial filtering, which constructs reference frame with a fractional interpolator or a de-block loop filter. According to the analysis in [13], spatial filtering operation has been described by a Gaussian shape model, which behaves like a bandwidth time-varying low pass filter whose transfer function can be expressed by

(10) $$H_t\left(\omega_x, \omega_y\right) = \exp\left(-\frac{\left(\omega_x^2 + \omega_y^2\right)\cdot t\cdot\sigma_f^2}{2}\right).$$

From above transfer function, it can be deduced that: when t = 0, it's an all pass filter and has no effect on the bandwidth of input signal; after every time step, it becomes an low pass filter with gradually narrowed bandwidth; when t approach infiniteness its bandwidth becomes zero and no signal would be permitted to pass through. This feature matches the leaky prediction nicely. Further, its distortion elimination can be derived as

(11) $$\sigma_v^2[t] = \frac{\sigma_u^2}{1+\gamma\cdot t}.$$

For every frame to be predicted, the time difference is set to 1, i.e. t = 1, in above equation. Therefore, distortion between two successive frames can also be delivered by (11). And if the block is treated as the research object, the propagation parameter for block at location (*m*, *n*) can be defined as

(12) $$\eta_{m,n} = \frac{1}{1+\gamma_{m,n}},$$

where $\gamma = \sigma^2_f / \sigma^2_g$ is a parameter describing the efficiency of spatial filter to reduce the introduced error, σ^2_f indicates the strength of loop filter and σ^2_g equals to 1.3 according to [14]. σ^2_f can be interpreted as the energy effect to signal, hence, it can be approximated by the average energy of the filter as

(13) $$\sigma_f^2 \approx \frac{1}{2\pi}\int_{-\pi}^{+\pi} \left|H_f(\omega)\right|^2 d\omega \cdot$$

Because of the fractional accuracy of motion compensation in H.264, the result reference block may be constructed by different kinds of interpolation. When motion vector points to an integer pixel point, no interpolation happens. When motion vector points to a quarter-pixel point, reference block may be constructed by half-pixel interpolation followed by quarter-pixel interpolation. As in Figure 5, taking some example, for integer referent point G, H, M, N, reference block undergoes no filter operation; for quarter reference point j, reference block undergoes one horizontal and one vertical filter operations.

According to [14] and [15], the filtering operations on the same direction can be treated as cascaded filtering. Consequently, (12) can be re-written as

(14) $$\eta = \frac{\left(1+\sum_{k=1}^{N_V}\sigma_{f_k}^2 \Big/ \sigma_g^2\right)^{-1}}{1+\sum_{k=1}^{N_H}\sigma_{f_k}^2 \Big/ \sigma_g^2},$$

where N_V and N_H are number of horizontal and vertical filtering that reference block undergoes.

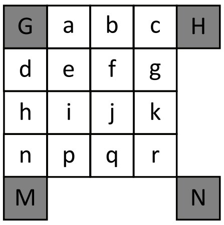


Fig. 5 Different referent point for interpolation

3.2.4 Propagation parameter setting for Intra-block

In contrast to previous standards, where Intra-coding is performed in transform domain, Intra-coding in H.264 is always conducted in spatial domain. This change improves the quality of predicted value, but when error occurs in neighbor block, distortion may be propagated from neighbor corrupted blocks as depicted in Figure 6. Due to various Intra prediction modes, the propagation behavior exhibits very differently from block to block. Fortunately, all the Intra-predictions are linear operation, and if the mean of error assumed to be 0, the distortion Intra-propagation for a block can be modeled as

(15) $$D_i^{ita} \approx w_A^2 D_i^A + w_B^2 D_i^B + w_C^2 D_i^C + w_D^2 D_i^D,$$

where $D_i^{A,\,B,\,C,\,D}$ are the distortions of neighbor block A, B, C, D of block *i*, and $w_{A,\,B,\,C,\,D}$ are prediction weights for block A, B, C, D.

Due to error propagation can't be considered over-microcosmically, the distortion of a block can't be obtained. But we can look this Intra-coded block as an Inter-coded block and recast (15) to calculate the propagation parameter as

(16) $$\eta_i^{ita} \approx w_A^2\eta_i^A + w_B^2\eta_i^B + w_C^2\eta_i^C + w_D^2\eta_i^D.$$

To this end, when video frames undergo compression operation, the needed information, such as motion vector and coding mode of every block, quantization distortion and coefficient slices of every frame, needs to be restored. And then, distortion effect for a NALU and its priority assignment can be performed.

3.3 Extraction priority assignment

During MGS bit-stream extraction, coefficient packets may be dropped. At decoder, using highest available quality packets frame reconstruction would produce drift. Although drift can be limited to a GOP, its incidence for every frame is different. As illustrated in Fig. 7, the drift incidence for different pictures is represented. Taking the frame with TId = 0 as a example, when a few of its packets is dropped during bit-stream extraction all the non-key pictures in corresponding GOPs is included into its drift incidence as shown in the dashed line box.

Based on above analysis, we can obtain the extraction priority assignment method as follows.

1) During the encoding of MGS bit-stream, lots of data has to be stored. They are bits number of every quality enhancement slice NALU, the distortion of every frame, the coding mode and motion vector of every 4x4 block, and the coefficients of every quality enhancement slice.
2) For every quality enhancement slice NALU, equation (7) is used to calculate its dropping distortion D_{drp}.
3) The quantized distortion D_q of related frames and dropping distortion D_{drp} are added together as the total distortion D_{Tot}.
4) According to the drift incidence as illustrated in Fig. 7, drift distortion D_{dft} is computed as section 3.2.
5) The distortion effect ∂D of a NALU can be estimated from the sum of related frame's D_{drp} and D_{dft} in the drift incidence. And the extraction priority can be calculated for every quality enhancement slice NALU according to equation (3).
6) Sort the extraction priority and merge the close priorities to fit for the number of Quality Layer that the application demands. The method of merging can be found in [11].
7) Assign Quality Layer extraction priority to NALUs.

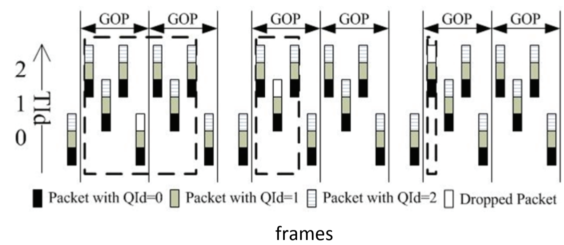


Fig. 7 Drift incidence of dropping coefficient packets for different frames

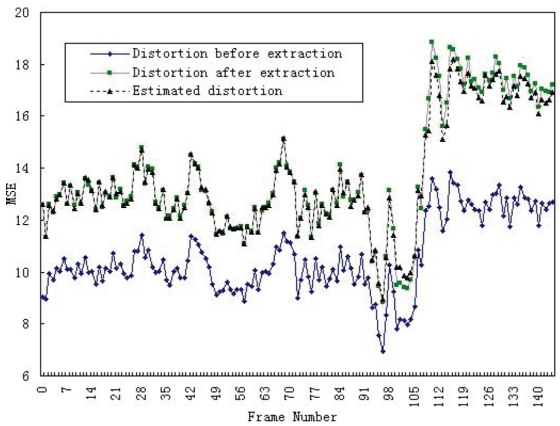
4 Experiment result

On the experimental platform of JSVM 9.18, lots of simulations are carried out to verify the proposed method. MGS bit-stream is produced by the encoder configuration of single spatial layer, 4 temporal levels, 3 quality levels and the latter 2 quality levels corresponds to two coefficients partitioning slices with Zig-Zag index of 0-7 and 8-16 respectively. Many video sequences usually selected by Joint Video Team (JVT) are chosen as the test sequences. The same encoding configuration is employed for all sequences and the experimental results of bus, foreman and ice in QCIF size and mobile in CIF size are illustrated in this section.

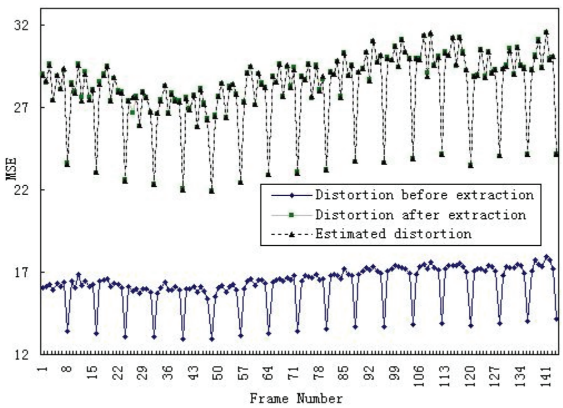
4.1 Distortion estimation experiments

During encoding, information needed to estimate the distortion is stored. Using the bit-stream extractor of ‘BitStreamExtractorStatic’, packets with QId≤1 are extracted to produce another bit-stream. And then, decoder of JSVM is performed to decode original and extracted bit-stream. Finally, the distortions of this two bit-streams are recorded. After estimating with the proposed method, the estimated and actual distortion curves are depicted in Fig. 8 for two selected test sequences, foreman and mobile.

In Fig. 8, distortion curves of two test sequences behave much differently. For mobile the distortion is regularly distributed according to GOP structure, which demonstrates that the motion of objects in mobile is not complicated. While for foreman distortion fluctuation is more drastic than that of mobile. Obviously, these two sequences exhibit distinct coding characteristic. Sequence foreman has high motion which may result in more Intra-blocks. Hence, due to the estimation of Intra-propagation parameter from neighbor blocks, more estimation error would be produced in foreman as the estimated result for frames after the 100th frame. In a nutshell, however, the proposed method can capture the drift distortion propagation characteristic of the test sequences and give out a proper estimation of distortion for every frame in a MGS bit-stream.



(a) foreman



(b) mobile

Fig. 8 Frame distortion estimation

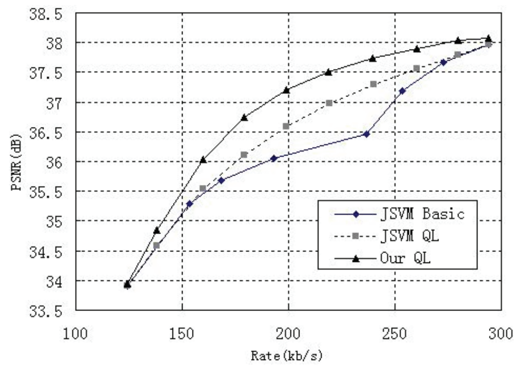
4.2 Extraction performance tests

At first, ‘QualityLevelAssignerStatic’, which is the software offered by JSVM to assign extraction priority, and our method are respectively performed to compare the efficiency of them. The test results are list in Table 1. Obviously, ‘QualityLevelAssignerStatic’ needs more time to assign priority than the proposed method. The need for decoding related frames consumes much more calculation time of ‘QualityLevelAssignerStatic’.

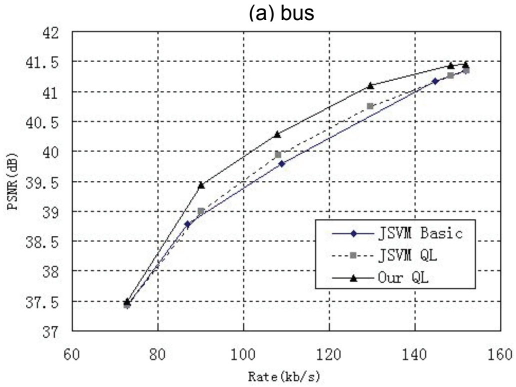
Table 1 Comparison of assignment time consume

Methods	The needed time (s)			
	bus	foreman	ice	mobile
QualityLevel Assigner	25.023	47.970	43.524	191.371
Proposed method	3.156	6.582	5.873	23.721

After assignment of Quality Layer respectively by ‘QualityLevelAssignerStatic’ and the proposed method, the extraction performances of the basic extraction of JSVM (referred to as ‘JSVM Basic’), the Quality Layer extraction of JSVM (referred to as ‘JSVM QL’) and the extraction based on our assignment method (referred to as ‘Our QL’) are compared as Fig. 9. As expected, ‘JSVM Basic’ performs the worst since it only uses the high level syntax elements of the NALUs to order them, therefore, it can not be conscious of their impact on the quality of the sequence. Due to the accurate estimation of the distortion, ‘Our QL’ outperforms ‘JSVM QL’. Consequently, our priority assignment method can identify the contribution of NALUs to video quality and assign proper rate distortion significance to every NALU more accurately. Accordingly, more excellent extraction performance is presented in the experiment.



(a) bus



(b) ice

Fig. 9 Comparison of extraction performance

5 Conclusion

In this paper, the introduced distortion of extracting MGS coefficient slice is analyzed. And then, a frame distortion estimation method for the MGS bit-stream with coefficient partitioning slices is proposed. The rate distortion priority assignment is conducted by the frame distortion estimation method and achieves a better extraction performance more efficiently with the same Quality Layer concept. Video adaptation application may be performed more simply and efficiently with the proposed method.

Acknowledgement

The authors wish to acknowledge the financial support of Natural Science Foundation of Jiangsu Province in 2009 (BK2009059) and Pre-research Foundation of PLA Univ. of Sci. & Tech. in 2009 (2009TX08).

REFERENCES

[1] Isabelle Amonou, Nathalie Cammas, Sylvain Kervadec, et al.. Layered quality optimization for SVM. *MPEG M111704*, Hong Kong, 2005.1.

[2] Isabelle Amonou, Nathalie Cammas, Sylvain Kervadec, et al.. Layered quality optimization of JSVM3 when considering closed-loop encoding. *ITU-T VCEG JVT-Q081*, Nice, 2005,10.

[3] H. Schwarz, D. Marpe, and T. Wiegand, Closed-loop coding with quality layers. *ITU-T VCEG JVT-Q030*, Nice, 2005.10.

[4] Isabelle Amonou, Nathalie Cammas, Sylvain Kervadec, et al.. Enhanced SNR scalability for layered CGS coding using quality layers. *ITU-T VCEG JVT-S044*, Geneva, 2006.

[5] Manu Mathew, Kyohyuk Lee, Woo-Jin Han. Multi Layer Quality Layers. *ITU-T VCEG JVT-S043*, Geneva, 2006.4.

[6] J. Lim, M. Kin, S. Hahm, et al.. An optimization-theoretic approach to optimal extraction of SVC Bitstreams. *ITU-T VCEG JVT-U081*, October 2006.

[7] Truong Cong Thang, Jung Won Kang, Jeong-Ju Yoo, et al.. Optimal Multilayer Adaptation of SVC Video over Heterogeneous Environments. *Journal of Advances in Multimedia*, vol. 2008: 1-8.

[8] Thomas Rusert, Jens-Rainer Ohm. Backward drift estimation with application to quality layer assignment in H.264/AVC based scalable video coding. *ICASSP 2007*, vol. 1, pp. 653-656.

[9] Jun Sun, Wen Gao, Debin Zhao, Weiping Li, “On Rate-Distortion Modeling and Extraction of H.264/SVC Fine-Granular Scalable Video,” *IEEE Trans. Circuits Syst. Video Technol.*, Vol. 19, No. 3, Mar. 2009.

[10] Ehsan Maani, Aggelos K. Katsaggelos. Optimized Bit Extraction Using Distortion Modeling in the Scalable Extension of H.264/AVC. *IEEE Trans. on image processing*, Vol. 18, No. 9, pp. 2022-2029, September 2009.

[11] Isabelle Amonou, Nathalie Cammas, Sylvain Kervadec, et al.. Optimized rate-distortion extraction with quality layers in the scalable extension of H.264/AVC. *IEEE Trans. on Circuits and Syst. Video Technol.*, 2007: 1186-1193.

[12] Batra, P. and Eleftheriadis, A. “Analysis of optimal dynamic rate shaping of Markov-1 sources”, *Imaging Processing, Proc. of IEEE Int. Conf. on*, vol. 3, October 2003, pp. 789~792.

[13] N. Färber, K. Stuhlmüller, and B. Girod, “Analysis of error propagation in hybrid video coding with application to error resilience”, *Imaging Processing, Proc. of IEEE Int. Conf. on*, vol. 2, October 1999, pp. 550-554.

[14] B. Girod, N. Färber, “Error-resilient standard-compliant video coding”, *Recovery Techniques for Image and Video Compression*, Kluwer Academic Publishers, P244, 1998.

[15] J. -G. Kim, J. Kim, and C. -C. J. Kuo, “Corruption model of loss propagation for relative prioritized packet video”, *Proc. of SPIE, Application of Digital Image Processing XXIII*, vol. 4115, January 2000, pp. 214-224.

***Authors**:*
Tao Zhu, Postgraduate Team 2, Institute of Communications Engineering, Biaoyin 2, Yudao Street, Nanjing, China, 210007, E-mail: zhutao007@gmail.com;
Xiong-wei Zhang, Institute of Command Automation, PLA Univ. of Sci. & Tech.