

Analysis of Bit Allocation of Temporal Scalability Using Hierarchical B Frames

Ming Li*, Yilin Chang*, Shuai Wan[†] and Fuzheng Yang*

* State Key Lab. of Integrated Service Networks, Xidian University

Email: liming_xidian@yahoo.com.cn, ylchang@xidian.edu.cn, fzhyang@mail.xidian.edu.cn

[†] Shaanxi Key Lab. of Information Acquisition and Processing, Northwestern Polytechnical University

Email: swan@nwpu.edu.cn

Keywords: Bit allocation, temporal scalability, quantization, rate-distortion optimization, Lagrange multiplier.

Abstract

Temporal scalability with dyadic temporal enhancement layers can be very efficiently provided with the concept of hierarchical B frames. In this paper, extensive experiments are carried out to investigate the influence of different bit allocation schemes to the performance of the dyadic hierarchical coding structure with hierarchical B frames. The bit allocation schemes are realized by quantization and rate-distortion optimization. Based on analysis of the experimental results, conclusions are derived with respect to the selections of the quantization step and Lagrange multipliers, respectively, which serve as solid bases for the design of the encoders employing hierarchical B frames to provide temporal scalability.

1 Introduction

Video coding with motion-compensated temporal lifting-based wavelet decompositions was first proposed in [1] and [2] in 2001. Since then, much work has been done in investigating this kind of codecs which is regarded as the temporal sub-band video codec (TSBVC). The motion-compensated temporal filtering (MCTF) extension of H.264/AVC is proposed in JVT-K023 [3]. During the investigation of this extension of H.264/AVC, it is discovered that the major part of the reported gains comes from the hierarchical prediction and coding framework rather than the additional motion-compensated update step [4], and the coding structures with hierarchical B frames are proposed and adopted to further improve the coding efficiency of H.264/AVC [5]. With proper management methods for reference frames and the decoded picture buffer (DPB), the coding structures with hierarchical B frames are employed in scalability video coding (SVC) to support temporal scalability [5] [6].

In the coding structures with hierarchical B frames, the dyadic hierarchical coding structure depicted in Figure 1 has already been integrated in H.264/AVC to improve the coding efficiency, in SVC to provide temporal scalability, and in multiview video coding (MVC) to serve as the basic coding structure [7]. Although the dyadic hierarchical coding structure is widely used, to the best of our knowledge,

efficient bit allocation for it still remains an open issue. This paper mainly investigates the influence of common bit allocation schemes on the rate-distortion (R-D) performance for the dyadic hierarchical coding structure.

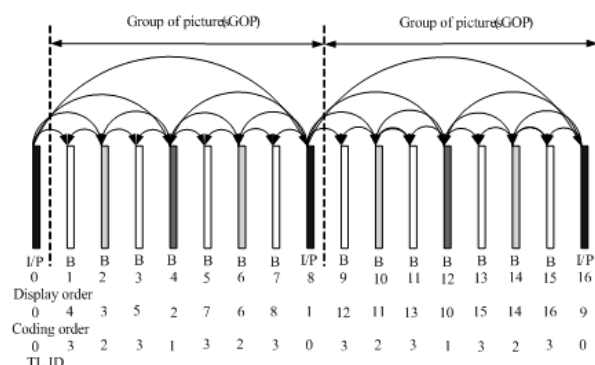


Figure 1: Dyadic hierarchical coding structure with hierarchical B frames, with 4 TLs and a GOP size of 8.

Bit allocation, which distributes the available bit budget among a set of coding units (e.g. frame, macroblock, etc.), is an essential part of coder control. A classical framework for bit allocation falls in R-D theory, which deals with the minimization of source distortion for a given bit budget. Bit allocation has great influence on the entire coding efficiency, and an appropriate bit allocation scheme can bring significant improvement to the encoder performance. In video coders, both the quantization and the R-D optimization (RDO) are the two main ways to implement bit allocation. Therefore, both of them are of great importance to the performance of the encoder. In this paper, extensive experiments are carried out to investigate the performance of the encoder using different designs of the quantization and various methods to set the Lagrange multipliers used in RDO [8]. All the experiments are performed on the H.264/AVC reference software JM10.2 [9]. The dyadic hierarchical coding structure, which serves as the basis for temporal scalability in standard SVC, is employed to provide temporal scalability. By analyzing the influence of different designs of quantization and RDO, conclusions and suggestions are given for optimal coder control in the coders with temporal scalability realized by hierarchical B frames.

The remainder of this paper is organized as follows. The introduction to the dyadic hierarchical coding structure and

some analysis of this coding structure are presented in section 2. The impact of bit allocation to the entire coding efficiency is studied with extensive experiments in section 3 and 4. In section 3, different designs of the quantization are investigated with RDO on and off. Various methods used to set the Lagrange multipliers for RDO are evaluated in section 3 with the quantization chosen according to the coding efficiency recorded in section 2. Conclusions are derived by analyzing the experimental results in the two sections. Section 4 concludes this paper and summarizes the suggestions for the design of quantization and RDO for the encoders which employ hierarchical B frames to provide temporal scalability.

2 Introduction and Analysis of the Dyadic Hierarchical Coding Structure

The dyadic hierarchical coding structure with hierarchical B frames is depicted in Figure 1, in which the display order, the coding order and the temporal layer (TL) identifier (TL ID) are also shown. The TL ID starts from 0 for the base layer and is increased from one TL to the next. The first frame of a video sequence is intra-coded as an instantaneous decoding refreshing (IDR) frame. So-called key frames (black in Figure 1) are coded in regular (or even irregular) intervals and build the base layer. As illustrated in Figure 1, the set of B frames between two successive frames of the temporal base layer together with the succeeding base layer frame is referred to as a group of pictures (GOP). Note that the IDR frame at the beginning of a video sequence is also a key frame. The key frames are either intra-coded or inter-coded using previous coded key frames as reference for motion-compensated prediction. The remaining frames of a GOP are B frames which are hierarchically predicted, and hierarchical B frames with a larger TL ID refers to a higher TL, while a smaller TL ID refers to a lower TL, as depicted in Figure 1. Obviously, this coding structure supports temporal scalability. That is, for each natural number k , the bit stream which is obtained by removing all access units of all TLs with a TL ID larger than k forms another valid bit stream for the given decoder.

In the dyadic hierarchical coding structure, key frames and B frames of different TLs are of different importance to the entire coding quality. That is, the key frames are used as reference frames for successive two GOPs except the first frame in the sequence for only one GOP, and the B frames of smaller TL IDs are used as reference frames directly or indirectly by many B frames of larger TL IDs within the same GOP, as illustrated in Figure 1. Generally, from the view of temporal scalability, the compression efficiency of the sequence can be improved by lowering the relative coding quality of the enhancement layer [10]. So the key frames which build the temporal base layer should be coded with the highest fidelity, and the B frames with smaller TL IDs should be coded with higher fidelity than the B frames with larger TL IDs. That is, from the view point of bit allocation, enough bits should be allocated to the key frames and the B frames with smaller TL IDs to guarantee their high coding quality, so that the entire coding quality could benefit from the improvement in the coding quality of the frames which

contributes more to the entire performance than other frames. In section 3 and 4, various bit allocation schemes carried out by different quantization and RDO parameter sets will be investigated, and some conclusions will also be derived.

3 Analysis of Quantization

Quantization is introduced into video coding to achieve a proper tradeoff between coding bit rate and distortion. That is, different quantization steps lead to different quantization fidelities and different amounts of actual coding bits. Generally, the smaller the quantization step, the better the quantization fidelity and the more the coding bits. This is because when a quantizer with a relatively small quantization step employed, many details would be reserved in the quantized signal and thus a relatively large quantity of bits would be used to code. By choosing the quantization step for each coding unit, the available bits, determined by considering the channel bandwidth and delay/buffer constraints, are allocated to each coding unit. That is, the bit allocation scheme can be implemented by properly choosing a set of the quantization steps used for the coding units. Therefore, the overall R-D performance of the encoder can be significantly improved by appropriately setting the quantization steps.

Extensive experiments have been carried out to evaluate the overall R-D performance of various bit allocation schemes, which are realized by different designs of the quantization used for the encoder with temporal scalability provided by the dyadic hierarchical coding structure. The number of inserted B frames is set to 7 and 15 (denoted as “7HB” and “15HB”), respectively. The RDO scheme recommended by the Joint Model of H.264/AVC [9] is employed in the encoder. According to the experimental results, some analysis will be presented and some conclusions will be derived. In video coding, a parameter called quantization parameter (QP) is used to index the quantization step. So in the remainder of this paper, QP will be used to refer to the quantization step. Lack of space forbids presenting all the experimental results. Therefore, only the experimental results of three test sequences are given in this paper. For other test sequences, similar experimental results could be obtained.

Here, two designs of the quantization will be evaluated. The first one uses fixed QP (denoted as “Fixed QP”). That is, the same QP will be used to quantize all the frames, including the key frames and all the hierarchical B frames. The second one uses cascading QP proposed by HHI in [5]. In this method, the B frames with the same TL ID use the same QP, and B frames with different TL IDs use different QPs. Let $QP(k)$ denote the QP in the TL with TL ID of k . Based on the QP for the key frames (denoted as “ $QP(0)$ ”), the QPs for B frames in different TLs are determined as follows:

$$QP(k) = QP(0) + \Delta QP + k, \quad k = 0, 1, 2, \dots, \quad (1)$$

where ΔQP represents the increment methods of the QP from the temporal base layer to the enhancement layer with TL ID of 1. Different values of ΔQP lead to various cascading strategies, denoted as “Cascading QP ($\Delta QP+1$)”, where

$(\Delta QP+1)$ is the actual increment of the QP from $QP(0)$ to $QP(1)$. In our experiments, ΔQP is set to 0 and 3, respectively. HHI recommends the value of ΔQP equal to 3 [5]. The cascading QP strategies aim to select the QP based on the TL, considering the fact that the key frames heavily impact the quality of the prediction signal for all hierarchical B frames, and the B frames with smaller TL IDs impacts the quality of the prediction signal for the B frames with larger TL IDs.

The experimental results of “Fixed QP”, “Cascading QP 1” and “Cascading QP 4” with RDO turned on are given in Figure 2. It can be observed that for both “7HB” and “15HB”, “Cascading QP 1” can always outperform “Fixed QP”. Using “Cascading QP 4” can improve the R-D performance in the case of “15HB”, but for “7HB”, its performance would be no better or even worse than “Fixed QP”. For the sequences with simple spatial contents or low motions (e.g. *Claire* and *News*), the number of B frames inserted between two adjacent key frames prefers 15 to 7, while for the sequences with complex spatial details or high motions (e.g. *Coastguard*), 7 B frames inserted is better than 15.

The experimental results “Fixed QP”, “Cascading QP 1” and “Cascading QP 4” with RDO turned off are given in Figure 3. Similar to the results with RDO on, “Cascading QP 1” outperforms “Fixed QP” in all the tests. Note that without RDO, “Cascading QP 4” can achieve the same and even slightly better performance compared with “Cascading QP 1”. Similar to the test results with RDO turned on, it is also observed that the most favourable number of B frames inserted between two adjacent key frames is sequence dependent.

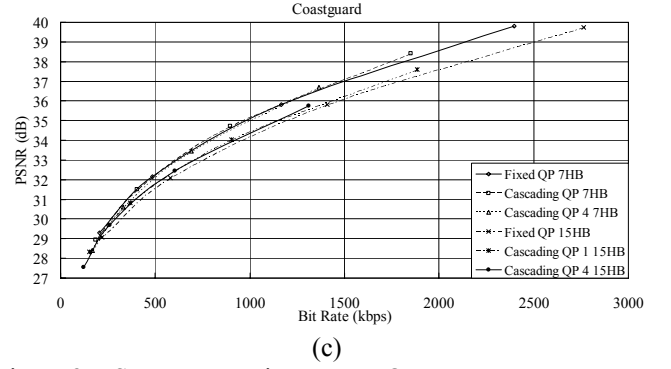


Figure 2: PSNR versus Bit Rate, RDO on.

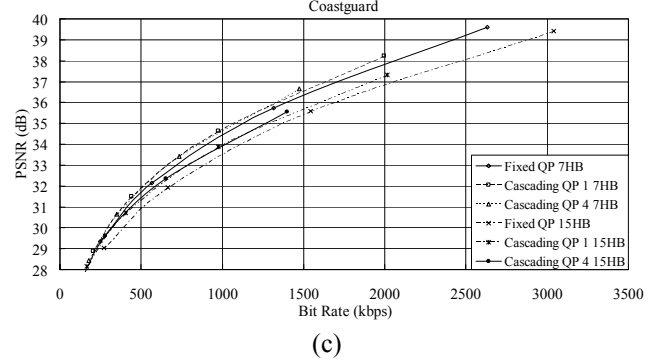
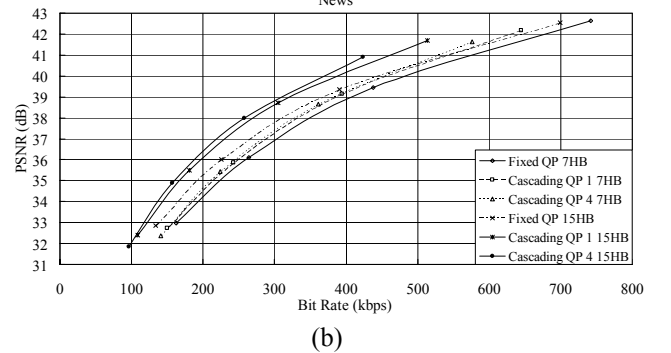
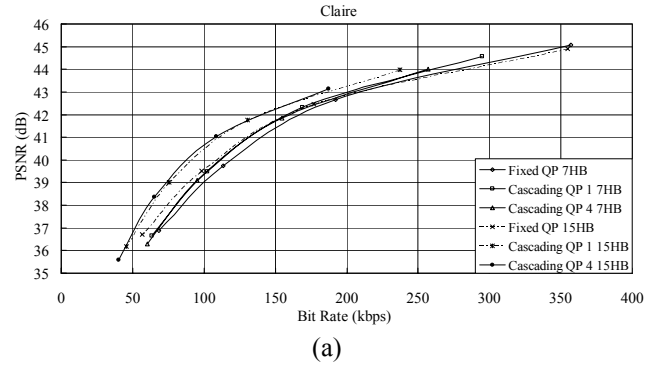


Figure 3: PSNR versus Bit Rate, RDO off.

Based on these experimental results, several conclusions can be derived. First, the number of B frames inserted is sequence dependent. That is, the coding efficiency will be improved if the encoder with temporal scalability provided by the dyadic hierarchical coding structure integrates an adaptive selection scheme for the GOP size by considering the temporal image characteristics of the video sequence. Here, the term “GOP

size” is employed because it is directly related to the number of B frames inserted between adjacent two frames. In the dyadic hierarchical coding structure, “GOP size” is calculated by the number of B frames inserted plus one.

Second, the performance of the encoder with temporal scalability provided by the dyadic hierarchical coding structure could be improved by using the cascading QP strategies over the TLs. By employing cascading QP strategies, more bits are allocated to code the key frames and the B frames with smaller TL IDs, which are directly or indirectly used as reference frames by many B frames in motion estimation. Therefore, the coding quality of the entire sequence would be improved. However, it should be guaranteed that enough bits must be reserved to code the B frames with larger TL IDs, since they are the most part in a GOP. Excessively depressing the coding quality of these B frames will degrade the entire performance. Also note that the cascading QP strategies would cause relatively large peak signal-to-noise ratio (PSNR) fluctuations. Therefore, for the cascading QP strategies, the optimal value by which the QP should be increased from one TL to the next is sequence dependent. The optimal selection of the QP should be done based on a comprehensive analysis of the R-D properties of different TLs.

Third, comparing the performance of “*Cascading QP 4*” with “*Fixed QP*” in the two scenarios with and without RDO, it can be observed that “*Cascading QP 4*” outperforms “*Fixed QP*” in all the tests without RDO. But with RDO turned on, “*Fixed QP*” could perform better than “*Cascading QP 4*” in some cases. It should also be noted that the performance gaps between cascading QP strategies and “*Fixed QP*” become narrow when RDO is on. The reason may be that the frames in the dyadic hierarchical coding structure are of very different importance to the entire performance, and their dependent relations are quite different from the traditional coding structures (e.g. IPPP and IBBP). From this point of view, it can be concluded that the RDO scheme is required to be modified and even re-designed for the coding structures with hierarchical B frames. Bearing this in mind, experiments on RDO are carried out, and the analysis of the RDO will be presented in the next section.

3 Analysis of RDO

The operational control of the source encoder is a key problem in video compression [8]. For the encoding of a video source, many coding parameters are determined by the coder control, such as macroblock modes, motion vectors, reference frames and transform coefficient levels, *etc.*, so that the available bits would be properly allocated to the coding units. Accordingly, the chosen values determine the R-D performance of the produced bitstream of a given encoder.

RDO performs operational control of the video encoders [8]. Generally, Lagrange multipliers are employed for the encoder to determine the appropriate coding parameters under certain source/channel constraints. By properly setting the Lagrange

multipliers according to available coding sources (e.g. the available bit rates), a well tradeoff can be achieved between the coding bit rate and the quality (represented by the distortion). Therefore, combined with the selection of quantization steps, the bit allocation schemes can be realized by appropriately setting the Lagrange multipliers used in the coder control.

In this section, the RDO scheme recommended by the Joint Model of H.264/AVC [9] will be employed for evaluation. In this RDO scheme, the Lagrange multiplier for mode decision (denoted as “ λ_{MODE} ”) used in the dyadic hierarchical coding structure is determined based on the QP of the current coding unit shown as follows:

$$\lambda_{MODE} = 0.68 \times 2^{\theta}, \quad (2)$$

where θ equals to $(QP-12)/3$, and QP is the QP for the current coding unit. Then for hierarchical B frames, the λ_{MODE} calculated by Equation (2) will be adjusted by a weighting factor obtained according to the QP for the current coding unit as below:

$$\lambda'_{MODE} = \begin{cases} 4 \cdot \lambda_{MODE}, & \theta > 4 \\ \theta \cdot \lambda_{MODE}, & 2 \leq \theta \leq 4, \\ 2 \cdot \lambda_{MODE}, & \theta < 2 \end{cases} \quad (3)$$

If this weighting process is turned off, λ'_{MODE} is set equal to λ_{MODE} . After that, λ'_{MODE} will be weighted by the factor determined by the TL which contains the current coding B frame:

$$\lambda''_{MODE} = w_k \cdot \lambda'_{MODE}, \quad (4)$$

where w_k is the weighting factor for the B frames with TL ID of k . w_k is calculated by

$$w_k = \begin{cases} 1.0, & k = TL_{MAX} \\ w_{k+1} - 0.2, & 0.6 \leq w_{k+1} < 1 \text{ and } k \neq TL_{MAX} \\ 0.6, & w_{k+1} < 0.6 \text{ and } k \neq TL_{MAX} \end{cases} \quad (5)$$

where TL_{MAX} is the maximum value of TL ID. If Equation (4) is not enabled, λ''_{MODE} is set equal to λ'_{MODE} .

Obviously, by using Equation (3) and/or Equation (4), the Lagrange multiplier for mode decision would be cascaded. In this paper, it is called “*Cascading λ* ”. Note that “*Cascading λ* ” can also be carried out by only using the cascading QP strategies according to Equation (2), without using Equation (3) and (4). Finally, if the sum of squared differences (SSD) is used in mode decision and the sum of absolute differences (SAD) is used in motion estimation, the Lagrange multiplier for motion estimation (denoted as “ λ_{MOTION} ”) is calculated by

$$\lambda_{MOTION} = \sqrt{\lambda''_{MODE}}. \quad (6)$$

In this section, the methods for weighting the Lagrange multipliers will be evaluated combined with two QP strategies, i.e. “*Fixed QP*” and “*Cascading QP 1*”. “*Cascading QP 1*” is chosen for it outperforms “*Fixed QP*” in all the tests of section 2. The experiments are carried out with various combinations of the two QP strategies and the methods for weighting the Lagrange multipliers, as presented in Table 1.

In all the tests of this section, 7 hierarchical B frames are inserted between two adjacent key frames.

	Equation (3)	Equation (4)
Weighting 1	√	√
Weighting 2	×	√
Weighting 3	√	×
Weighting 4	×	×

Table 1: Methods for weighting the Lagrange multipliers.

Figure 4 presents the experimental results. Note that “Cascading QP 1” is marked as “CasQP” for short in Figure 4. Obviously, the methods using “CasQP” perform better than those using “Fixed QP”. In the four methods using “Fixed QP”, “FixedQP_Weighting_1” and “FixedQP_Weighting_4” outperform the other two, and “FixedQP_Weighting_3” is the worst one in performance of these four methods. “FixedQP_Weighting_3” aims to set the Lagrange multipliers according to the TL, but it weights the Lagrange multipliers manually without taking the R-D properties of each TL into consideration, which is also the problem for Equation (3) because “FixedQP_Weighting_4” outperforms “FixedQP_Weighting_2”. It should be noted that simultaneously using Equation (3) and (4) could improve the entire performance. Comparing the results of the methods using “CasQP”, it can be seen that the performance gaps among “CasQP_Weighting_1”, “CasQP_Weighting_2” and “CasQP_Weighting_4” are very narrow, and these three methods all achieve better performance than “CasQP_Weighting_3”. Note that “Cascading λ ” can also be carried out by only using “CasQP”. The comparisons among “CasQP_Weighting_3”, “CasQP_Weighting_4” and the methods using “Fixed QP” show that “CasQP_Weighting_4” can achieve the best coding efficiency. The above results all demonstrate that the entire performance can benefit from “Cascading λ ”, and the “Cascading λ ” carried out by cascading QP strategies contributes the most to the entire performance. And the methods to weight the Lagrange multipliers should be carefully designed with full considerations for the different R-D properties of different TLs so that the target bits could be reasonably allocated.

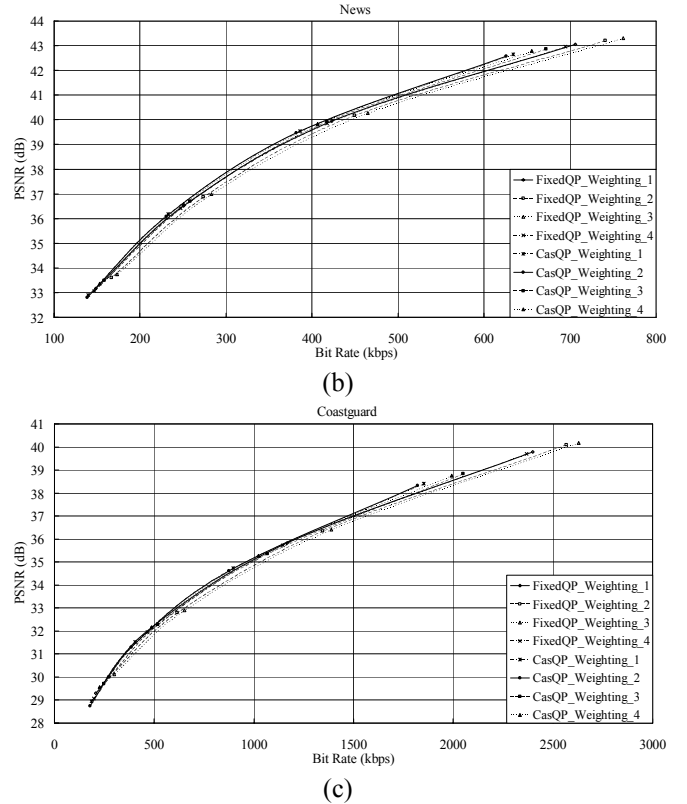
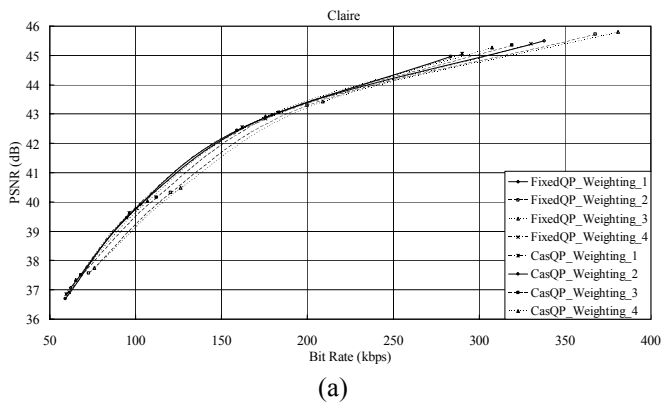


Figure 4: PSNR versus Bit Rate.

Moreover, the experimental results show that the performance of “Cascading λ ” is also related to the available coding bit rates. In the scenarios of low bit rate, the performance gaps between “FixedQP_Weighting_4” and the methods using “CasQP” are very narrow. The reason may be that for low bit rates, the bits reserved from the B frames by adjusting the Lagrange multipliers are quite limited, and re-allocating the reserved bits to the key frames and the B frames in lower TLs could bring little improvement to the coding quality of these frames. Therefore, the improvement of the entire performance is also little. As opposed to low bit rates, for high bit rates, there will be enough bits reserved from the B frames so that the entire performance could benefit from the improvement in the coding quality of the key frames and the B frames in lower TLs. This suggests that the available bit rates should be taken into consideration in determining the Lagrange multipliers. It can also be observed that the “Cascading λ ” methods carried out by “CasQP” can more efficiently reserve bits from the B frames than “Fixed QP” and Equation (4). This is because the Lagrange multipliers should be different for each TL. Equation (2) comes from theoretical analysis and extensive experiments based on the R-D theory. Accordingly, the “Cascading λ ” realized only by “CasQP” is more reasonable than the methods using Equation (3) and/or (4), and “CasQP_Weighting_4” can achieve a good performance. Therefore, the weighting methods for the Lagrange multipliers should be designed based on the theoretical analysis, considering the available sources (e.g. the available bit rates) and the source characteristics.

Let's focus on “FixedQP_Weighting_2” and “FixedQP_Weighting_4”. “FixedQP_Weighting_2” employs Equation (3) to weight the Lagrange multipliers, so that large Lagrange multipliers will be used when the QP for the current B frame is large. This would result in the reduction of the number of bits allocated to B frames, and then more bits could be reserved for the key frames. It seems that by using “FixedQP_Weighting_2”, the overall coding efficiency would be improved, because key frames are the most essential frames in the dyadic hierarchical coding structure. But the fact is that “FixedQP_Weighting_4” outperforms “FixedQP_Weighting_2”, which means that Equation (3) is not appropriate for the hierarchical coding structure. The reason is that in the dyadic hierarchical coding structure, most frames in the sequence are B frames and different B frames in different TLs have unequal influences on the entire coding efficiency. Therefore, only coding the key frames with good quality without full considerations for the B frames could not bring improvement to the entire coding efficiency.

Based on these experimental results and according analysis, several conclusions can be derived. First, from the view point of RDO, “CasQP” can achieve better entire performance than “Fixed QP”, and contributes the most to the improvement of the entire performance brought by “Cascading λ ”. Second, weighting the Lagrange multipliers according to the TL may improve the coding efficiency, but it should be modified or even re-designed with full considerations for the R-D properties of different TLs as well as their correlations. Third, the influence of the available bit rate should be taken into consideration in determining the weighting factors for the Lagrange multipliers. Especially for low bit rate, the number of bits reserved from B frames should be cautiously considered from the view point of the entire performance.

4 Conclusions

In this paper, extensive experiments are carried out to investigate the performance of different bit allocation schemes realized by various designs of the quantization, methods to set the Lagrange multipliers, and combinations of the two. Based on the experimental results and the according analysis, some conclusions are derived for the encoder with temporal scalability provided by the dyadic hierarchical coding structure. A summary of the suggestions are listed as follows.

- (1) The optimal size of a GOP is sequence dependent. The entire performance of the encoder will be improved if GOP size could be adaptively determined by considering the temporal image characteristics of the video sequence.
- (2) Cascading the QP over the TLs could improve the R-D performance of the encoder. However, the optimal value by which the QP should be increased from one TL to the next should be determined based on a comprehensive analysis of the R-D properties of different TLs.
- (3) Cascading the Lagrange multipliers over the TLs could also improve the R-D performance of the encoder. However, the methods to carry out the cascading Lagrange multipliers should be designed with full considerations for the R-D properties of different TLs and their correlations. Moreover, the available bit rate should also be considered in determining the Lagrange multipliers.

Acknowledgements

We would like to acknowledge the supports provided by the National Science Foundation of China (60772134) and the 111 Project (B08038).

References

- [1] Béatrice Pesquet-Popescu, *et al*, “Three-dimensional lifting schemes for motion compensated video compression”, in *Proc. ICASSP*, Vol. 3, pp. 1793-1796, (2001).
- [2] A. Secker, *et al*, “Motion-compensated highly-scalable video compression using an adaptive 3D wavelet transform based on lifting”, in *Proc. ICIP*, Vol. 2, pp. 1029-1032, (2001).
- [3] H. Schwarz, *et al*, “Subband Extension of H.264/AVC”, *ISO/IEC JTC1/SC29/WG11 and ITU-T Q6/SG16, Doc. JVT-K023, 11th Meeting*, Munich, DE, (2004). [Online] Available: <http://ftp3.itu.ch/av-arch/jvt-site/>
- [4] H. Schwarz, *et al*, “Comparison of MCTF and Closed-loop Hierarchical B Pictures”, *ISO/IEC JTC1/SC29/WG11 and ITU-T Q6/SG16, Doc. JVT-P059, 16th Meeting*, Poznan, (2005). [Online] Available: <http://ftp3.itu.ch/av-arch/jvt-site/>
- [5] H. Schwarz, *et al*, “Hierarchical B Pictures”, *ISO/IEC JTC1/SC29/WG11 and ITU-T Q6/SG16, Doc. JVT-P014, 16th Meeting*, Poznan, July 2005. [Online] Available: <http://ftp3.itu.ch/av-arch/jvt-site/>
- [6] H. Schwarz, *et al*, “Overview of the Scalable Video Coding Extension of the H.264/AVC Standard”, *IEEE Tran. on CSVT*, Vol. 17, No. 9, pp. 1103-1120, (2007).
- [7] P. Merkle, *et al*, “Efficient Prediction Structures for Multiview Video Coding”, *IEEE Tran. on CSVT*, Vol. 17, No. 11, pp. 1461-1473, (2007).
- [8] T. Wiegand, *et al*, “Rate-Constrained Coder Control and Comparison of Video Coding Standards”, *IEEE Tran. on CSVT*, Vol. 13, No. 7, pp. 688-703, (2003).
- [9] JVT Reference Software Version JM10.2 [Online]. Available: http://iphone.hhi.de/suehring/tml/download/old_jm/jm10.2.zip.
- [10] M. Flierl, *et al*, “Generalized B Pictures and the Draft H.264/AVC Video-Compression Standard”, *IEEE Tran. on CSVT*, Vol. 13, No. 7, pp. 587-597, (2003).