Rate Control for Scalable Video Model

Long Xu[*a], Siwei Ma[a], Debin Zhao[a,b], Wen Gao[a,b]

[a]Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100080, China;
[b]Department of Computing Science, Harbin Institute of Technology, Harbin, 150001, China

ABSTRACT

Scalable video coding (SVC) has become more and more important with the enrichment of multimedia data and the diversification of network and terminal devices. In current MPEG SVC activities, a scalable extension of H.264/AVC, called scalable video model (SVM)[1-4], is proposed by HHI and has shown further coding efficiency improvement and scalability functionality. However, the SVM model doesn't provide an efficient rate control scheme now, and rate control is achieved through a full search for selecting a suitable quantization parameter (QP). That is very inefficient and much time-consuming. In this paper, an efficient rate control scheme is proposed for the SVM, which is derived from the state-of-the-art hybrid rate control schemes of JVT[5,6] with some considerations for scalable video coding. In the proposed rate control scheme, the rate distortion optimization (RDO) involved in the step of encoding temporal subband pictures is only implemented on the low-pass subband pictures, and rate control is independently applied to each spatial layer. For each spatial layer, the rate control is implemented at GOP, picture and basic unit levels. Furthermore, for the temporal subband pictures obtained from the motion compensation temporal filtering[7-11] (MCTF), the target bit allocation and quantization parameter selection inside a GOP could make full use of the hierarchical relations inherent from the MCTF. The proposed rate control scheme has been implemented into SVM3.0[1] and experiment results show that the proposed algorithm can achieve the target bit rate with little bit rate fluctuation and keep fine image quality at the same time, but the computation complexity is reduced heavily.

Keywords: rate control; scalable video coding; motion compensation temporal filtering; rate distortion optimization

1. INTRODUCTION

With the fast convergence of computer, communications, and entertainment industries, various digital media applications are connected through the communications networks now. In a networking environment such as Internet, the end systems may have a wide variety of capabilities and requirements. The compressed bit streams created for one particular application may not be satisfactory, efficient, or even useful for servicing the users with different resource capacities. An efficient approach to dealing with the heterogeneity issue is scalable coding, where the coded bitstreams for low-end applications are just embedded as subsets of the coded bitsstreams for high-end applications. As such, a single coded bitstreams can be applied to diverse application environments by selectively transmitting and decoding related sub-bitstreams. Some desirable scalable functionalities including SNR scalability, bit rate scalability, spatial scalability, temporal scalability, and complexity scalability, allow decoding at different image quality, coding bit rate, pixel resolution, frame rate, and decoder complexity respectively. Scalable coding is also demonstrated to be more robust for transmission over an error-prone environment combined with a prioritized data protection strategy.

Recently, with the enhancing demands for scalable video coding, in MPEG-21, a scalable video coding group has been founded to establish a standard to meet this requirements. In the just past MPEG meeting, a scalable extension of H.264/AVC proposed by HHI, has adopted as the reference software of the scalable video coding, called SVM in SVC group. To achieve an efficient scalable bit-stream representation of a video sequence, the temporal correlations between pictures are exploited using an open-loop prediction incorporated in the MCTF. The related temporal analysis-synthesis filter bank is generalized to facilitate an adaptive block-based choice between the motion-compensated lifting representations[12,13] of the Haar filter (uni-directional prediction) and the 5/3 filter (bi-prediction), both coupled with multiple-reference frame capabilities. Employing the MCTF, an ordered set of temporal subband pictures with reduced temporal frequency are produced, and coded subsequently in different approaches. The first picture is independently

[*] Corresponding author. Tel: +86-10-58858300
E-mail addresses: {lxu, swma, dbzhao, wgao}@jdl.ac.cn (L. Xu, S. Ma, D. Zhao, W. Gao).

Visual Communications and Image Processing 2005, edited by Shipeng Li,
Fernando Pereira, Heung-Yeung Shum, Andrew G. Tescher, Proc. of SPIE Vol. 5960
(SPIE, Bellingham, WA, 2005) · 0277-786X/05/$15 · doi: 10.1117/12.631424

coded as an instantaneous decoding refresh (IDR) picture, and all remaining pictures are coded in "B...BP" or "B...BI" format. In addition, forward motion estimation and mode selection is applied between the temporal low-pass subband pictures. For the temporal high-pass pictures, except for intra MBs, there aren't further processing involved. To achieve spatial scalability, a multiple layers stream with different spatial resolution is coded together in SVM, and each layer is coded with that for H.264/AVC. In addition, a new quantization strategy based on the hierarchical relations inherent from MCTF, is employed for the temporal subband pictures. Furthermore, intra mode would be selected when no inter mode is efficient. These new tools further improve coding efficiency compared with the single-layer coding. On the other hand, these new features make it difficult for rate control on SVM.

Rate control holds an important position in video coding, although it's not a normative tool for any video coding standard. In video communications, rate control must ensure the coded bitstream can be transmitted successfully and make full use of the limited bandwidth. As a consequence, a proper rate control scheme is usually recommended by a standard during the development, e.g. TM5[14] for MPEG-2, TMN8 and TMN12[15,16] for H.263, and VM8[17] for MPEG-4, etc. H.264/AVC is the newest international video coding standard, and some work about rate control has been done for H.264/AVC too. In the contribution[18], a rate control scheme based on VM8 has been proposed to and adopted by H.264/AVC test model. In another contribution[19], an improved rate control for H.264/AVC is provided with rate distortion optimization (RDO) and hypothetical reference decoder (HRD) jointly considered, part of which has also been adopted by H.264/AVC test model.

Rate control involves not only quantization parameter selection but also optimal mode decision and optimal bit allocation. Rate control also should guarantee the decoder never suffers from underflow or overflow. Firstly, the quantization parameter for each coding unit (e.g. the macroblock) is determined so as to fulfill the target bit rate constraint, which can be fulfilled by the rate-distortion (R-D) model. For example, in TM5, a simple linear rate-distortion model is introduced. In TMN8 and VM8, a more accurate quadratic R-D model is used, which can reduce rate control error and provide better performance but have relatively higher computational complexity. In a different way, the relation between rate and QP is indirectly represented with the relation between rate and ρ, where ρ is the percent of zero coefficients after quantization[20,21]; and also, a modified linear R-D model with an offset indication overhead bits is used for rate control on H.261/3/4 in the contributions[22,23]. Secondly, rate control usually incorporate with RDO, which could brings more coding efficiency for optimized mode decision and bit allocation. In order to reduce the temporal correlations among successive frames, inter-frame coding is widely used, which is usually realized by motion compensation prediction (MCP). With block basis motion estimation, the residual texture and motion vectors associated in the current block need to be coded finally. Obviously, for a given bit rate, over-large motion information or residual information wouldn't give the best coding efficiency, so the trade-off between the motion information and the residual information, on which the motion compensated video coding heavily depends, should be considered. The trade-off is usually achieved by a rate distortion optimization (RDO), that is formulated by minimizing the cost J [24], shown as follows,

$$J = D + \lambda_{opt} R \tag{1}$$

here the distortion D representing the residual (texture or prediction error) measured as sum absolute distortion (SAD) or mean absolute distortion (MAD), is weighted against the number of bits R associated with the motion information by using the Lagrange multiplier λ_{opt}. Each λ_{opt} corresponds to a bit rate range and a trade-off between the motion information and the residual information. A large λ_{opt} works well at a low bit rate while a small λ_{opt} works well at a high bit rate[24]. Thirdly, the coded bitstream is usually variable bit rate with variable length coding (VLC). For transmitting the bitstream in the constant bit rate (CBR) or variable bit rate (VBR) channel, buffering mechanism should be considered to smooth the variations between the bit rate of coded bitstream and that of channel bandwidth. Thus, in the past the buffer manage strategy have been fully studied for the video coding standards, which is known as the hypothetical reference decoder (HRD) in H.261/3/4 and the video buffer verifier (VBV) in MPEG-1/2/4. As shown in Figure 1,

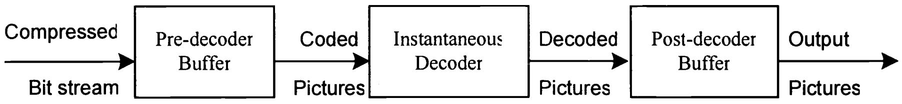

Figure 1: A Hypothetical Reference Decoder

the HRD/VBV contains a pre-decoder buffer (or VBV Buffer) through which compressed data flows with a precisely specified arrival and removal timing. It may be designed for VBR or CBR operation, and for low-delay or delay-tolerant behavior[25]. Compressed data representing a sequence of coded pictures flows into the pre-decoder buffer according to a specified arrival schedule. All compressed bits associated with a given coded picture are removed from the pre-decoder buffer by the instantaneous decoder at the specified removal time of the picture. If the buffer becomes full and more bits are arriving, the pre-decoder will suffer from buffer overflow; on the contrary, it will suffer from buffer underflow if the removal time for a picture occurs before all the bits of the picture have arrived.

In this paper, a rate control scheme for SVM is proposed with some considerations on R-D model, RDO and HRD/VBV. Firstly, a quadratic R-D model similar to that in H.264/AVC is used with adaptive quantization and updating model based on the content of picture; secondly, the RDO is implemented on the temporal low-pass subband pictures coded in "IP...PIP...P" format, while no further processing is involved for the temporal high-pass subband pictures. It should be pointed out that our proposed rate control scheme only works on the temporal subband pictures, i.e. excluding the motion estimation and mode selection evolved in MCTF; thirdly, a simplified buffer management with the occupancy and variation of virtual buffer jointly considered, is presented in the SVM. The rest of the paper is arranged as follows. In section 2, the detailed description about our proposed rate control algorithm is presented. In section 3, the experiment results are shown. Section 4 concludes this paper.

2. RATE CONTROL ALGORITHM

Rate control is used to create a stream to meet with the available bandwidth, and also compliant to HRD/VBV. In this paper, a rate control algorithm based on JVT proposals[5,6] is proposed for SVM with some considerations on virtual decoder buffer occupancy. Generally, it can be divided into three tightly consecutive levels: group of pictures (GOP) level rate control, picture level rate control and the optional basic unit level rate control. The proposed rate control algorithm is applied on the temporal subband pictures. And also, the picture types I/P and B represent temporal low-pass and high-pass subband picture respectively.

Being that hierarchical relations in the subband pictures, a new target bit allocation strategy is employed in the picture level. Meanwhile, a different prediction mechanism for I/P and B pictures is provided. Furthermore, the virtual decoder buffer is considered for the first time in our proposed rate control algorithm. For restructuring base layer pictures in decoder, there is need much buffer buffering at least a GOP of pictures; for decoding enhancement layer, not only a GOP of pictures in the current layer but also corresponding temporal subband and reconstructed pictures in its base layer need to be buffered. However, a relative simplified buffer management is considered if just taking decoding temporal subband pictures into account, i.e. only the section between MCTF and inv-MCTF considered in the whole codec. In this way, similar buffer management of H.264/AVC can be satisfied, only decoding temporal subband pictures.

2.1. GOP level rate control

GOP level rate control calculates the remaining bits for the rest pictures, and initializes the quantization parameter of the first picture (I or P) in the current GOP. When the j^{th} picture in the i^{th} GOP is coded, the number of total bits for the rest pictures in this GOP is computed as follows,

$$B_i(j)=\begin{cases} R_i(j)/f\times N_i-V_i(j) & j=1 \\ B_i(j-1)+\dfrac{R_i(j)-R_i(j-1)}{f}\times(N_i-j+1)-b_i(j-1) & j=2,3,\ldots,N_i \end{cases} \quad (2)$$

where f is the predefined frame rate, N_i is the size of the i^{th} GOP, $R_i(j)$, $b_i(j)$ and $V_i(j)$ are the instant available bit rate, actual generated bits and occupancy of the virtual buffer for the j^{th} picture in the i^{th} GOP, respectively. For the first picture (j = 1) in a GOP, the number of remaining bits calculated from the upper formula in (2) is the allocated bits for the current GOP in fact. Besides, the instant available bit rate $R_i(j)$ can be variable for the different frames or GOPs Considering the VBR case, while in the CBR case, $R_i(j)$ is always equal to $R_i(j-1)$ and (2) can be simplified as:

$$B_i(j) = B_i(j-1) - b_i(j-1) \tag{3}$$

Initially, the virtual buffer is filled by the motion bits generated previously in the MCTF, so the occupancy of virtual buffer is initialized as $m_i(1)$ which presents the motion bits of the j^{th} picture in the i^{th} GOP. Except the first GOP, besides initial motion bits, the virtual buffer's occupancy of the last GOP coded also is considered as upper formula (4) shown. After a picture coded, the $V_i(j)$ is updated as bottom formula (4):

$$\begin{aligned} V_i(1) &= \begin{cases} m_i(1) & i = 1 \\ V_{i-1}(N_{i-1}) + m_i(1) & other \end{cases} \\ V_i(j) &= V_i(j-1) + b_i(j-1) - \frac{R_i(j-1)}{f} \quad j = 2,3,\ldots,N_i \end{aligned} \tag{4}$$

Besides bit allocation, the initial quantization parameter decision is also included in the GOP level rate control. For the first GOP, the predefined quantization parameter specified for motion estimation/mode decision in the MCTF is used as the initial quantization parameter for simplicity. For other GOPs, the initial quantization parameter is predicted as follows,

$$QP_i(1) = \frac{sumQP(i-1)}{1 + N_{i-1}^p} \tag{5}$$

where $sumQP(i-1)$ is the sum of average QP for all I/P pictures in the $(i-1)$th GOP, and N_{i-1}^p is the total number of P pictures in the $(i-1)$th GOP.

2.2. Picture level rate control

Picture level rate control allocates target bits for each picture based on the remaining bits, picture's complexity and virtual buffer's occupancy. Getting the target bits and MAD of current picture, the quantization parameter can be obtained based on the R-D model[15]. The MAD of a block A of size N×N located at (x, y) inside the current picture compared to a block B located at a displacement of (v_x, v_y) relative to A in a previous picture is defined as:

$$MAD(x,y) = \frac{1}{N^2} \sum_{m,n=0}^{N-1} \left| F_i(x+m, y+n) - F_{i-t}(x+v_x+m, y+v_y+n) \right| \tag{6}$$

where F_i is the current picture and F_{i-t} is a previously coded picture. In our proposed rate control algorithm, the picture level rate control consists of two stages: pre-encoding and post-encoding. In the pre-encoding stage, QP decision for each picture are accomplished with virtual buffer considerations, while in the post-encoding stage, the models updating with the statistical results is implemented.

2.2.1. Pre-encoding stage

In this stage, the quantization parameter of each picture is calculated. Firstly, the target bits are allocated for the current picture, and then the quantization parameter for the current picture can be obtained with the pre-defined rate distortion (R-D) model.

The target bit allocation should both consider the occupancy of virtual buffer and remaining bits for the rest pictures. Firstly, smoothing the occupancy of virtual buffer by regulating bit rate arriving, the target bits allocated for the j^{th} picture in the i^{th} GOP based on instant bit rate and the occupancy of virtual buffer are determined as:

$$\widetilde{T}_i(j) = (1 - \frac{V_i(j+1) - V_i(j)}{V_i(j)}) \times \frac{R_i(j)}{f} \tag{7}$$

Secondly, remaining bit allocation for the j^{th} picture in the i^{th} GOP is computed as:

$$\hat{T}_i^t(j) = \frac{\hat{X}_i^t(j) \times B_i(j)}{\sum_{t=p,b} K_t \times \hat{X}_i^t(j) \times N_{t,r}} \tag{8}$$

where $N_{p,r}$ and $N_{b,r}$ are the number of the remaining I/P pictures and the number of the remaining B pictures, respectively, $\hat{X}_i^t(j)$ is the predicted complexity measure for the current coding picture, and K_p/K_b is the ratio of I picture's QP and P/B picture's QP regulated with the selected wavelet function in the MCTF[1]. The complexity measure is the product of target bits and average QP for a picture (basic unit or MB). For pictures with type B, the complexity can be determined beforehand, while for the pictures with type I/P, the complexity only can be predicted from the nearest picture coded previously. After coding a picture in the i^{th} GOP, the actual generated bits and average QP can be obtained, and then, the complexity measure is updated as:

$$X_i(j) = \alpha \times b_i(j-1) \times avgQP_i(j-1) \tag{9}$$

where $avgQP_i(j\text{-}1)$ is the average of quantization parameters of the previously coded picture, α is a constant and set as 0.9 when next picture is P type otherwise set as 1 in our experiments. Specially, in the SVM, the pictures with type of I or P both are the temporal low subband pictures, and also except the first GOP with one I and one P pictures, only one I or P picture is in a GOP, so the complexity of I/P picture in the next GOP shall be predicted from that of I/P picture in the previously coded GOP. In conclusion, the predicted complexity measure is computed as:

$$\hat{X}_i^t(j) = \begin{cases} X_{i-1}(1) & t = i, p \;\&\; i \neq 1 \\ X_i(1) & t = i, p \;\&\; i = 1 \\ X_i(j) & t = b \end{cases} \tag{10}$$

Lastly , the parameter of target bits is determined with a weighted combination of $\widetilde{T}_i(j)$ and $\hat{T}_i(j)$:

$$T_i(j) = \beta \times \hat{T}_i(j) + (1-\beta) \times \widetilde{T}_i(j) \tag{11}$$

where β is a constant and set as 0.9 in our experiments. To conform to the virtual buffer requirement, the target bits are further bounded by:

$$T_i(j) = \min\{U_i(j), \max\{Z_i(j), \mathrm{T_i}(j)\}\} \tag{12}$$

where Zi(j) and Ui(j) are the minimum buffer constraint and maximum buffer constraint for preventing buffer from overflow and underflow. Same as the state-of-the-art hybrid coding, at least a picture needs buffering for decoding successfully. At the same time, the maximum buffer constraint is set as (14) avoiding buffer overflow.

$$U_i(j) = \begin{cases} B_{i-1}(N_{i-1}) + t_{r,1}(1) \times R_i(1) & j = 1 \\ U_i(j-1) + (\frac{R_i(j)}{f} - b_i(j)) & other \end{cases} \tag{13}$$

$$Z_i(j) = \begin{cases} B_{i-1}(N_{i-1}) + \frac{R_i(j)}{f} & j = 1 \\ Z_i(j-1) + \frac{R_i(j)}{f} - b_i(j) & other \end{cases} \tag{14}$$

where $t_{r,1}(1)$ is the removal time of the first picture from the coded picture buffer.

Getting the target bits for a picture, the QP can be obtained with pre-defined R-D model. After motion estimation and mode selection in the MCTF (pre-mode-decision), the MAD of I/P pictures is still unable to be determined, so it is predicted from the closet picture coded previously by a linear model,

$$\tilde{\sigma}_i(j) = a_1 \times \sigma_i(j-1) + a_2 \tag{15}$$

where a_1 and a_2 are two coefficients with initial values 1 and 0. And then, the quantization parameter corresponding to the target bits is computed as:

$$T_i(j) = c_1 \times \frac{\tilde{\sigma}_i(j)}{QP_i(j)} + c_2 \times \frac{\tilde{\sigma}_i(j)}{QP_i^2(j)} - m_{h,i}(j) \tag{16}$$

where $m_{h,i}(j)$ is the total number of header bits and motion vector bits, c_1 and c_2 are two coefficients. Since a drop in peak signal-to-noise ratios (PSNR) among successive pictures will deteriorates the visual quality of the whole sequence, the quantization parameter $QP_i(j)$ is adjusted by:

$$QP_i(1) = \max\{QP_{i-1}(1) - 2, \min\{QP_{i-1}(1) + 2, QP_i(1)\}\} \tag{17}$$

With such modifications, the difference in PSNR is not more than 2 between two successive pictures. And more, considering QP boundary in the SVM, the final quantization parameter is further bounded by 51 and 0. The quantization parameter is then used to perform quantization for each MB in the current picture. Specially, for B pictures, the MAD can be calculated from the current picture except intra block determined in the MCTF. The quantization parameter corresponding to the target bits is then calculated by using the formula (16). But for the intra blocks in B pictures, the MAD can't be obtained, and also is unreasonable predicted from any coded picture; however, only few intra blocks lie in a B picture. When pre-mode-decision is implemented, those intra modes can be recorded. Therefore, the MAD of the current intra block can be calculated approximately based on the recorded information in the pre-mode-decision stage.

2.2.2. Post-encoding stage

After encoding a picture, the parameters *a1* and *a2* of linear prediction model (15), as well as c_1 and c_2 of quadratic R-D model (16) are updated with a linear regression method similar to MPEG-4 Q2[26,27]. Meanwhile, the remaining bits for the rest pictures $B_i(j)$ is updated using (3).

2.3. Basic unit level rate control

Basic unit is defined to be a group of continuous MBs. It is used to obtain a trade-off between the overall coding efficiency and the bits fluctuation. The basic unit level rate control is similar to the picture level rate control, including MAD prediction, bit allocation, and quantization parameter decision in basic unit level. In our simulating system, the spatial layers with different resolutions from QCIF (176×144) to 4CIF (704×576) are coded for a same video clip. Generally, the basic unit level rate control is efficient for the large size pictures (>QCIF) from our experience.

Firstly, the MAD of the l^{th} basic unit is calculated in the current coding picture. In case I/P pictures, the predictive MAD, $\tilde{\sigma}_{l,i}(j)$, is obtained by model (15) using the actual MAD of co-located basic units in the picture coded previously. In case B pictures, the MAD of current basic unit can be calculated directly. Secondly, determining the target bits for the l^{th} basic unit is implemented as follows,

$$\hat{b}_l = T_r \times \frac{\tilde{\sigma}_{l,i}^2(l)}{\sum_{n=l}^{Nunit} \tilde{\sigma}_{n,i}^2(k)} \tag{18}$$

where T_r is the remaining bits for the rest basic units in the current picture, and initialized as the picture target bits $T_i(j)$. Thirdly, the quantization parameter $QP_{l,i}(j)$ for the l^{th} basic unit of j^{th} picture in i^{th} GOP is calculated using the quadratic R-D model (16), and then bounded by:

$$QP_{l,i}(j) = \max\{QP_{l-1,i}(j) - DQuant, \min\{QP_{l.i}(j), QP_{l-1,i}(j) + DQuant\}\} \tag{19}$$

where *DQuant* is a constant, and generally is regulated with the quantization parameter. In our experiments, *DQuant* is 1 if $QP_{l-1,i}(j)$ is greater than 27, otherwise is 2. Meanwhile, to maintain the smoothness of visual quality, (19) is further bounded by

$$QP_{l,i}(j) = \max\{0, \overline{QP_i}(j-L-1)-6, \min\{51, \overline{QP_i}(j-L-1)+6, QP_{l,i}(j)\}\} \tag{20}$$

Specially, for the first basic unit in the current picture, the QP can be derived from average QP of all basic units in the previously coded picture,

$$QP_{1,i}(j) = \alpha \times \frac{sumQP_i(j-1)}{N_{unit}} \tag{21}$$

where N_{unit} is the number of basic unit in this picture, α is a constant as provided in (9). When the number of remaining bits is less than 0, the QP is set as:

$$QP_{l,i}(j) = QP_{l-1,i}(j) + DQuant \tag{22}$$

Similarly, the QP is further bounded by (20) to maintain the smoothness of perceptual quality. Lastly, the QP is used to perform RDO for all MBs in the current basic unit. After coding a basic unit, the number of remaining bits, the coefficients of linear prediction model (15) and quadratic R-D model (16) are updated.

3. EXPERIMENT RESULTS

In order to evaluate the performance of the proposed rate control algorithm, we did encoding experiments using the 4CIF sequence “Crew” (704x576, 60Hz), CIF sequences “Husky” and “Foreman” (352x288, 30Hz) with the proposed rate control algorithm and the original rate control algorithm. The results show that the proposed algorithm controls bit rate successfully with little bit rate fluctuation; meanwhile, the picture’s quality of the proposed algorithm is better than that of original algorithm; furthermore, time-consuming is reduced compared with original method. In our simulating system, four level spatial scalabilities (QCIF, 15Hz, 96kbps), (QCIF, 15Hz, 192kbps), (CIF, 30Hz, 384kbps) and (CIF, 30Hz, 750kbps) is provided for the sequences “Crew” and “Husky”; four level spatial scalabilities (QCIF, 15Hz, 64kbps), (QCIF, 15Hz, 128kbps), (CIF, 30Hz, 256kbps) and (CIF, 30Hz, 512kbps) is provided for the sequence “Foreman”. The proposed rate control algorithm is implemented independently in each layer. For the enhancement layer with same spatial resolution as base layer, the motion information is obtained by using that of base layer; meanwhile, it is only performed in the CBR case although it can be worked in the VBR case; in addition, it is implemented in the basic unit level with basic unit size being the number of MBs in a row of a picture. Table 1-3 in the next page shows the coding results of the proposed rate control algorithm. The sequence format and testing conditions are also shown in Table 1-3. From the tables, we can see that the proposed algorithm can efficiently control the bit rate at each layer which has different combined resolutions of spatial, temporal and SNR. Usually, the mismatch of target bit rate and real bit rate does not exceed 0.5%. Figure 2-3 shows the PSNR per frame for the test sequence “Crew” coded at the same bit rate with the proposed rate control and the original rate control. From these figures, the curves of proposed algorithm almost are above the curves of original algorithm, i.e. the PSNR of proposed algorithm is higher than that of original algorithm at the large part of frames. In average, the PSNR gain is about 0.2-1.5dB.

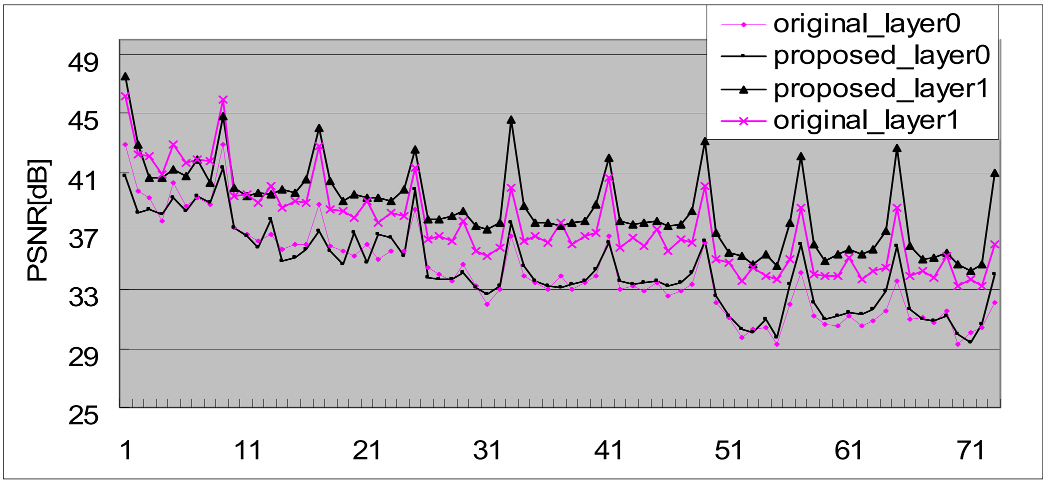

Figure 2: PSNR versus frame curve of sequence "Crew" at 96kbps/192kbps with QCIF, 15Hz

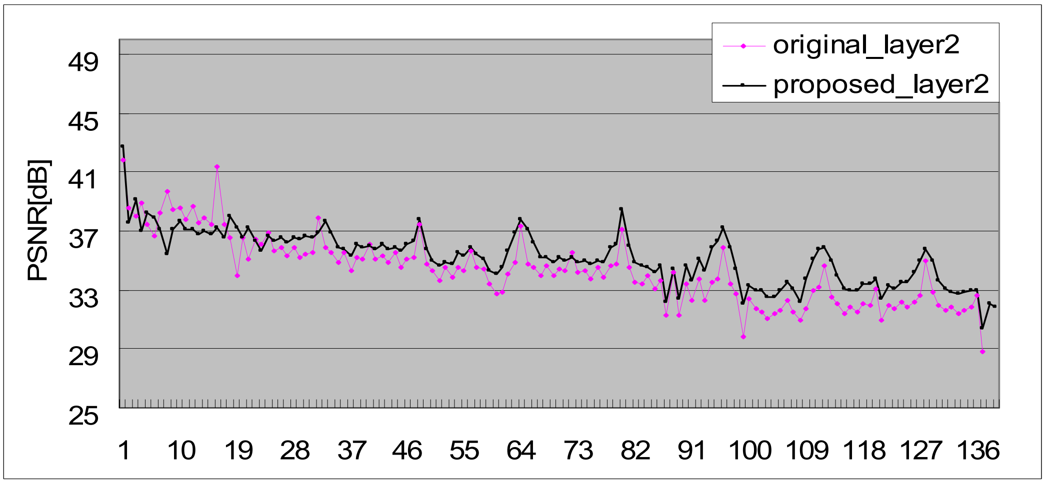

Figure 3: PSNR versus frame curve of sequence "Crew" at 384kbps with CIF, 30Hz

Table 1: The results with proposed rate control for "Crew" sequence

Layer	Target bit-rate(kbps)	Achieved bit-rate(kbps)	Mismatch	Psnr(db)
0(176x144,15)	96	96.28	-0.29%	34.13
1(176x144,15)	192	192.52	-0.27%	38.09
2(352x288,30)	384	383.31	0.18%	34.14
3(352x288,30)	750	748.28	0.23%	37.07

Table 2: The results with proposed rate control for "Foreman" sequence

Layer	Target bit-rate(kbps)	Achieved bit-rate(kbps)	Mismatch	Psnr(db)
0(176x144,15)	64	64.28	-0. 44%	33.86
1(176x144,15)	128	127.76	0.19%	37.20
2(352x288,30)	256	255.44	0.22%	33.96
3(352x288,30)	512	510.52	0.29%	36.99

Table 3: The results with proposed rate control for "Husky" sequence

Layer	Target bit-rate(kbps)	Achieved bit-rate(kbps)	Mismatch	Psnr(db)
0(176x144,15)	96	95. 62	0.39%	33.72
1(176x144,15)	192	192.21	-0.11%	38.38
2(352x288,30)	384	385.46	-0.38%	35.19
3(352x288,30)	750	747.68	0.31%	37.10

4. CONCLUSIONS

In this paper, an efficient rate control scheme for the SVM is presented. In each layer, the bit allocation is scaled down from the low to the high subband pictures owing to the hierarchical wavelet decomposition structure, and subsequently, the QP for each picture (or a basic unit) is determined based on a quadratic R-D model. Experiment results show that the generated bit rate is very close to the target bit rate, and meanwhile keep fine quality of the pictures. The initial parameters decision, optimized quantization mechanism and further inter-layer predictions need to be studied in our future work. Meanwhile, the probability model and rate distortion function specified for the subband pictures should be considered furthermore.

REFERENCES

1. ISO/IEC JTC1/SC29/WG11, Scalable Video Model 3.0, Palma de Mallorca, Spain, October 2004.
2. Heiko Schwarz, Detlev Marpe, and Thomas Wiegand, Further experiments for an MCTF extension of H.264, ITU-T VCEG, Doc. VCEG-v06, 23rd Meeting, Redmond, Washington, USA, July, 2004.
3. H. Schwarz, D. Marpe, and T. Wiegand, Subband Extension of H.264/AVC, ITU-T VCEG, Doc. VCEG-V04, March 2004.
4. Heiko Schwarz, Detlev Marpe, and Thomas Wiegand, Scalable Extension of H.264/AVC ISO/IEC JTC1/SC29/WG11 MPEG04/M10569/S03 Munich, March 2004.
5. Siwei Ma et al, Proposed Draft Description of Rate Control, Joint Video Team (JVT) of ISO/IEC MPEG & ITU-T VCEG, Doc. JVT-F086-L, 6th Meeting, Awaji, Island, JP, 5-13 December, 2002
6. Siwei Ma et al, Proposed Draft Adaptive Rate Control, Joint Video Team (JVT) of ISO/IEC MPEG & ITU-T VCEG, Doc. JVT-H017r3, 8th Meeting,Geneva, 20-26 May, 2003
7. J.-R. Ohm, Complexity and delay analysis of MCTF interframe wavelet structures, ISO/IEC JTC1/WG11 Doc. M8520, July 2002.
8. D. Taubman, Successive refinement of video: fundamental issues, past efforts and new directions, Proc. of SPIE (VCIP 2003), vol. 5120, pp. 649-663, July 2003.
9. M- Flierl and B. Girod, Video coding with motion-compensated lifted wavelet transforms, Proc. of PCS, pp. 59-62, April 2003.
10. M. Flierl, Video Coding with Lifted Wavelet Transforms and Frame-Adaptive Motion Compensation, Proc. of VLBL, pp. 243-251, Sep. 2003.
11. W. Sweldens, A custom-design construction of biorthogonal wavelets, J. Appl. Comp. Harm. Anal., vol. 3 (no. 2), pp. 186-200, 1996.
12. ISO/IEC JTC1, Test model 5, JTC1/SC29/WG11 Coding of Moving Pictures and Associated Audio, MPEG 1994.

13. P. P. Vaidyanathan and P.-Q. Hoang, Lattice structures for optimal design and robust implementation of two-band perfect reconstruction QMF banks, IEEE Trans. Acoust. Speech Signal Process, 36:81.94, 1988.
14. Daubechies I, Sweldens W, Factoring wavelet transforms into lifting steps, Journal of Fourier Analysis and Application, 1998,4(3): 245-267.
15. Jordi Ribas-Corbera, Shawmin Lei, Rate Control in DCT Video Coding for Low-Delay Communications, IEEE Transactions on Circuits and Systems for Video Technology, VOL. 9, NO. 1, FEBRUARY 1999, pp172-185.
16. Appendix II of H.263 Version 3, Test model 12 , Tenth Meeting, Portland, OR, USA, Aug. 22-24, 2000
17. Chiang T, Zhang Y, A new rate control scheme using quadratic rate distortion model, IEEE Trans, 1997, CSVT-7(1):287-311.
18. ISO/IEC TC JTC 1/SC29, Draft ISO/IEC 14496-5:2002/PDAM6:2003, DOC. N5821
19. S.Ma, W.Gao, F.Wu, Y.Lu, rate control for AVC video coding scheme with HRD considerations, IEEE international conference on Image Processing, ICIP 2003, Barcelona, Spain, 14-17 Sept. 2003.
20. S. Milani, L. Celetto, G. A. Mian, A Rate Control Algorithm for the H.264 encoder, Proceedings of the Sixth Baiona Workshop on Signal Processing in Communications, 390-396, Baiona, Spain, Sept. 2003.
21. Z. He, S. K. Mitra, A unified rate-distortion analysis frame work for transform coding, IEEE Trans. on Circuits Syst. Video Technology. Vol. 11. No. 12, Dec. 2001
22. J. Wang, Z. Chen, Y. He, Y. Chen, A MAD-based rate control strategy, Doc. JVT-D070, Klagenfurt, Austria, July 2002
23. C. Wong, O. C. Au, B. Meng, H. Lam, Novel H.26X optimal rate control for low-delay communicaitons, ICICS-PCM 2003, 15–18 Dec. 2003, Sigapore
24. T. Wiegand, M. Lightstone, D. Mukherjee, T. G. Campbell, and S. K. Mitra, Rate-distortion optimized mode for very low bit rate video coding and emerging H.263 standard, IEEE Trans. on Circuits and Systems for Video Technology, Vol. 6, No. 2, April 1996, pp. 182-190.
25. ISO/IEC 14496-10, HRD Editer's Input to Final Committer Draft (FCD) of Joint Video Specification (ITU-T Rec.H.264\ISO/IEC 14496-10 AVC) regarding Hypothetical Refefrnce Decoder, Doc. JVT-D146, 4th Meeting, Klagenfurt, Austria, 22-26 July, 2002.
26. I.-H. Shin, Y.-L. Lee & HW Park, Rate control using linear rate-model for H.264, Signal Processing: Image Com., vol. 19, pp. 341-352, April 2004.
27. H.M. Kim, Adaptive Rate Control Using Nonlinear Regression, IEEE Trans. Circuit and System for Video Technology, vol. 13, pp. 432-439, May 2003.