Hindawi Publishing Corporation
EURASIP Journal on Advances in Signal Processing
Volume 2007, Article ID 70914, 13 pages
doi:10.1155/2007/70914

Research Article

Multiple Adaptations and Content-Adaptive FEC Using Parameterized RD Model for Embedded Wavelet Video

Ya-Huei Yu, Chien-Peng Ho, and Chun-Jen Tsai

Department of Computer Science and Information Engineering, National Chiao Tung University, Hsinchu 30010, Taiwan

Received 12 September 2006; Revised 16 February 2007; Accepted 16 April 2007

Recommended by Anthony Vetro

Scalable video coding (SVC) has been an active research topic for the past decade. In the past, most SVC technologies were based on a coarse-granularity scalable model which puts many scalability constraints on the encoded bitstreams. As a result, the application scenario of adapting a preencoded bitstream multiple times along the distribution chain has not been seriously investigated before. In this paper, a model-based multiple-adaptation framework based on a wavelet video codec, MC-EZBC, is proposed. The proposed technology allows multiple adaptations on both the video data and the content-adaptive FEC protection codes. For multiple adaptations of video data, rate-distortion information must be embedded within the video bitstream in order to allow rate-distortion optimized operations for each adaptation. Experimental results show that the proposed method reduces the amount of side information by more than 50% on average when compared to the existing technique. It also reduces the number of iterations required to perform the tier-2 entropy coding by more than 64% on average. In addition, due to the nondiscrete nature of the rate-distortion model, the proposed framework also enables multiple adaptations of content-adaptive FEC protection scheme for more flexible error-resilient transmission of bitstreams.

1. INTRODUCTION

Multimedia distribution over heterogeneous networks and devices has become the mainstream enabling technology for new generations of services. For distribution and playback of a video content on various devices under different network conditions, scalable video coding schemes are usually used. A typical approach for scalable coding is to use a layered coding approach such as that of MPEG-4 simple scalable profile [1] or FGS [2]. In these approaches, the video bitstream quality is optimized for certain bitrate conditions. Adaptation of such content to a new target bitrate after the encoding process usually results in suboptimal bitstreams.

A different approach from the layered coding schemes is to design a scalable codec that produces embedded scalable bitstreams without inherent layered structures. The wavelets-based video codecs belong to this category [3–5]. Because there is no inherent layer structure for wavelet video bitstreams, video parameters such as resolution, frame rate, and bitrate can be dynamically adapted with fine granularity after the encoding procedure. If the rate-distortion (R-D) tradeoff information is embedded in the bitstream, the adaptation process can produce an R-D optimal bitstream at runtime for the target application. One major advantage of wavelet codecs over coarse-granularity layer-based codecs is that wavelet bistreams facilitate multiple adaptations. For example, in Figure 1, the video server transmits dynamically adapted scalable bitstreams to two different devices, namely the notebook and the cellular phone. Upon reception of the embedded bitstreams, the notebook plays the high-quality bitstream on its screen. In addition, it truncates (adapts) the received bitstream further and sends it to another device (the PDA) with tighter channel and device constraints. For the other distribution chain in Figure 1, the cellular phone first receives an adapted bitstream from the server and plays it on its internal large screen. Later, when the user decides to watch the video on the small external screen to conserve power, the video decoder can extract and decode only part of the received bitstream and displays a smaller video.

Although multiple adaptations can be achieved using layer-structured embedded bitstreams as well, they are not desirable because each layer of such bitstreams is preoptimized for certain target bitrate by the encoder. Take the scenario in Figure 1 for example; in order to adapt and transmit the received bitstream to the PDA, the notebook can only extract the embedded layers which do not exceed the channel

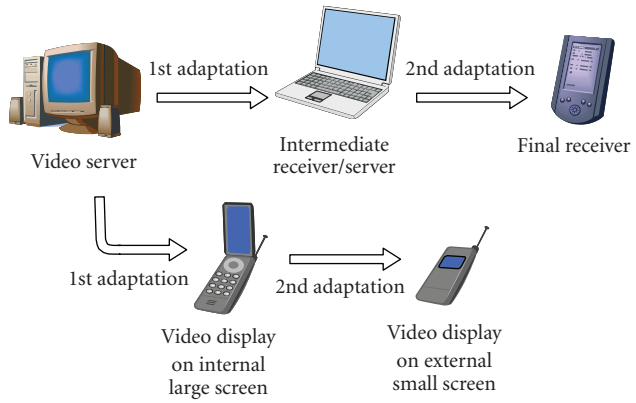

Figure 1: Two examples of multiple-adaptation applications where the same video content is adapted several times down the distribution chains.

and device constraints of the PDAs. This approach is quite simple but the bitstream cannot achieve the best quality possible since the runtime constraints may not meet the preoptimized layers embedded in the scalable bitstream. On the other hand, with a fully embedded bitstream where both R-D information and the wavelet video data are transmitted to the notebook, the notebook can extract an R-D optimized bitstream according to the runtime constraints of the target device. This approach achieves better quality than the layer-structured scheme, but the side information, namely the R-D information, is required and the complexity of the bitstream adaptor is higher. The issue is especially true for resource critical systems, like PDAs or cellular phones. Therefore, a low-complexity bitstream adaptation mechanism which can extract embedded R-D optimized bitstream is very important.

Many rate adaptation schemes have been proposed for embedded image/video codecs [6–8]. The basic idea behind these rate control techniques is similar. In general, the rate control scheme for embedded coders is composed of two parts. The first part is to model the rate-distortion characteristics of a group of input image/video data, and the second part is the bit allocation mechanism that assigns proper number of bits to various parts of the input data according to their importance. For wavelet video codecs, the most popular rate adaptation scheme is the 3D-ESCOT proposed by Xu et al. [4]. In this approach, R-D information is computed from real data points and is encoded into the bitstream for later adaptation. Bisection search is applied at runtime to determine the optimal truncation point. Although the adapted bitstream achieves optimality given certain rate constraint, the size of the side information and the complexity of the adaptation are not trivial for small devices.

In addition to multiple adaptations of video data, R-D side information is also very useful for content-adaptive forward error correction (FEC) protection of video data. Several frameworks for wavelet-based video streaming have been proposed in the literature recently. However, none of the existing work allows for multiple-adaptation of content-adaptive FEC protection data. Chu and Xiong [9] introduced a packetization scheme for combined wavelet video coding and FEC for video streaming and multicasting. However, data interleaving is not used in this work and the FEC protection degree is not adaptive to coefficients of different coding passes, which makes the system less robust. Dong and Zheng [10] proposed a content-based retransmission framework for wavelet video streaming. Nevertheless, retransmission-based error control requires longer jitter buffer and may consume too much extra bandwidth in high error rate channels [11]. In addition, fixed degree of FEC protection consumes considerable overhead which is wasted if there are less channel errors than estimated. Ho and Tsai [12] proposed a content-adaptive FEC protection/packetization mechanism of wavelet video data, but multiple adaptations of FEC codes are not considered because transmission of the side information was a nonnegligible overhead.

In this paper, a parameterized R-D model-based approach for R-D optimized multiple adaptations of video bitstream and content-adaptive FEC protection is proposed. The major achievement of the proposed framework is to reduce both the size of the R-D side information embedded in the bitstreams and the computational complexity of the runtime rate adaptor. The organization of the paper is as follows. Section 2 introduces the problem of multiple-adaptation problem for embedded codecs and content-adaptive FEC protection to the granularity of coding pass level. Section 3 discusses a parameterized rate-distortion model for more efficient R-D side information representation. The proposed multiple-adaptation schemes for both video data and FEC protection data based on the parameterized R-D model are presented in Section 4. The experimental results will be shown in Section 5. Finally, the conclusion and discussions are given in Section 6.

2. MULTIPLE-ADAPTATION PROBLEM OF FEC-PROTECTED WAVELET VIDEO DATA

The functional diagram of the wavelet-based embedded video codec with 3D-ESCOT [4] is shown in Figure 2. The input YC_BC_R frame data is first transformed into frequency domain via temporal and spatial subband decompositions. The transform process is followed by the quantization and the entropy coding processes with rate allocation mechanism. Popular wavelet-based image and video coders typically use discrete wavelet transform (DWT) for spatial subband decomposition and motion-compensated temporal filtering (MCTF) for temporal subband decomposition. Context-adaptive arithmetic coding is used for entropy coding. Finally, the rate allocation procedure 3D-ESCOT is used to explore bitrate (quality) scalability of the embedded bitstreams. For wavelet-based codecs, video data is partitioned into coding units, which could be a frame, a frequency band, or a coding block. The function of rate allocation is to extract a smaller subbitstream from a compressed bitstream that meets some application constraints.

During the rate allocation process, the frame rate, resolution, and bitrate can all be changed to form the target bitstreams. This is done in the tier-two process of the

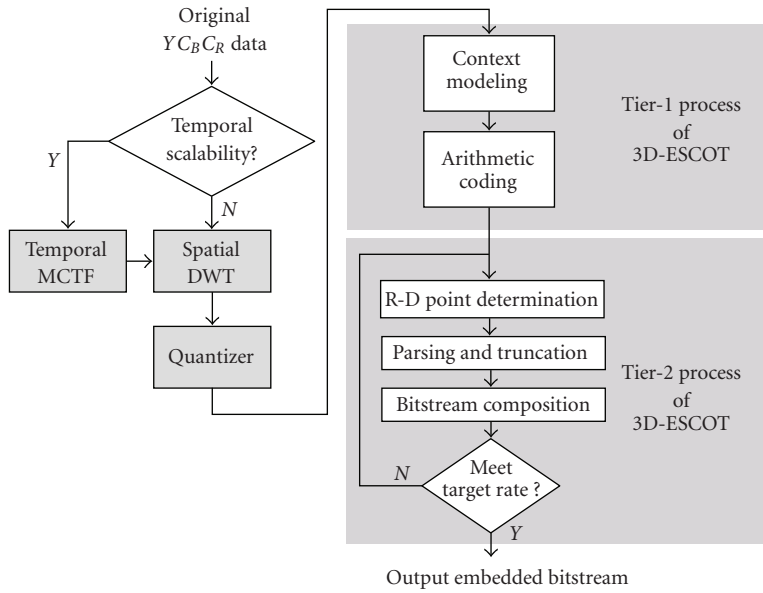

Figure 2: Wavelet video coding framework. The shaded areas illustrate the two-stage 3D-ESCOT rate adaptation process.

3D-ESCOT algorithm. As shown in Figure 2, the tier-two process is composed of three modules, namely, R-D point determination, parsing and truncation, and bitstream composition. For each candidate R-D point selected by the rate allocation algorithm (in the R-D point determination module), the parse-and-truncation operation and the bitstream composition operation must be performed in order to get the actual bitrate associated with the candidate R-D point. It is important to point out that the parsing-and-truncation module requires a lot of bit-level manipulations and the bitstream composition module requires many memory copy operations. Therefore, reducing the number of search iterations is particularly crucial for a mobile decoder such as a handset or a PDA since theses devices uses RISC processors with slow memory subsystems which are less efficient for these operations.

For multiple-adaptation applications, in order to achieve R-D optimal truncation of the bitstream and generation of content-adaptive FEC protection codes, R-D side information must be embedded into the bitstream throughout the distribution chain. Therefore, the size of the side information must be as small as possible to reduce transmission overhead. In addition, the intermediate adaptation of the bitstream is very likely to be performed by mobile devices. Therefore, a mechanism to reduce the complexity for the nonlinear R-D optimization problem is also crucial.

2.1. R-D side information and R-D optimized rate allocation

Several R-D models have been proposed to establish the tradeoff between rate and distortion for each coding unit [4, 8, 13]. An R-D model represents the degree of degradation of a coding unit when the size of the compressed data is constrained by the available bandwidth. The R-D models of the coding units can be used by the bit allocation algorithm to sort out the priority of the coding units. There are two typical ways to build the R-D characteristics model. The first method computes discrete R-D relationship data points from the real image data for model construction. The other method is to use a parameterized close-form model.

In wavelet-based embedded codecs, bitrate scalability is achieved by fractional bitplane coding. Inclusion of an additional fractional bitplane in a coding unit to the bitstream contributes to both increment of bits (rate) and reduction of quality loss (distortion). Recording of the rate and distortion data point of each fractional bitplane provides a precise, yet discrete, R-D model of the embedded bitstream [4]. However, storing all the discrete R-D values for each fractional bitplane in each coding unit is expensive. Even worse, for multiple adaptations, this R-D information must be embedded into the bitstream throughout the distribution chain. Furthermore, in order to find the best truncation point which matches the rate constraint, nonlinear optimization techniques must be used for bit allocation.

Different from the discrete R-D model approach, some literatures [8, 13] use close-form models to describe the R-D characteristic of the video data. In the closed-form R-D equation, content-dependent information is summarized in a few parameters. In general, the parameters can be estimated from the content statistics and/or by curve fitting of sparse data points. By using a closed-form R-D model, memory consumption of the rate control process can be substantially reduced, but the accuracy of bit allocation may decrease, depending on the accuracy of the R-D model.

The goal of the bit allocation procedure is to achieve maximal quality for a given bitrate or minimal bitrate for a given distortion. Giving the R-D characteristics models for each coding unit, nonlinear optimization techniques can be applied to distribute the coding bits among all coding units

in an optimal way. A popular approach is to use the Lagrange multiplier to transform constrained optimization problem into unconstrained optimization problem [4, 8, 13]. During this process, some truncation points will be deleted from the candidates of optimal solutions since they do not fall on the convex hull of R-D curves. Among the optimal truncation point attributes, the λ values represent the tradeoff parameters between rate and distortion at those truncation points. By applying a specific λ_c to all coding units, the collective set of all truncation points with their λ values closest to λ_c builds an optimal bitstream with the given constraint. An iterative search method, such as bisection search, can be used to iteratively select different λ_c until the composed bitstream meets the target constraint. The weakness of the iterative search method is that the convergence rate may be slow. Further improvement can be achieved if the search process takes advantage of the R-D characteristics of the content.

Besides the iterative search method, some studies [14, 15] designed special data structure to record R-D tradeoff points of all coding units. For example, a heap-based structure has been proposed to process rate allocation for embedded image coding in [14]. One major disadvantage of fast search algorithm with special data structure is that the required memory may be extremely large in order to build the complete data structure to store all coding unit information; therefore they are not suitable for small mobile devices.

2.2. R-D side information and content-adaptive FEC protection

For streaming of scalable video over lossy IP networks, FEC coding is a very practical error-resilience technique for unequal error protection of video data. However, previous FEC techniques only allow for coarse layer-based unequal error protection [16–18], or unequal protection between different types of syntax elements [19, 20]. Ho and Tsai [12] propose a new method for fine-level adaptive FEC protection of wavelet coefficients. In [12], the R-D side information of wavelet codecs is used to calculate the degree of importance of the wavelet coefficients given estimated packet loss rate of the channel. The granularity of the protection level can be fine-tuned for different wavelet video coefficient coding passes. Although the proposed technique performs very well in practice, it does not allow for multiple adaptations since the side information will be discarded after packetization due to its nontrivial overhead.

3. THE PROPOSED R-D SIDE INFORMATION FOR MULTIPLE-ADAPTATION APPLICATIONS

In this section, the parameterized R-D model and the way the model is encoded in the wavelet bitstream are presented. Although the fundamental R-D model used in the proposed framework is well known for video codec researchers, some modifications must be exercised in order to facilitate tier-two of the 3D-ESCOT rate adaptation algorithm. In particular, two R-D models (one for coding block-level modeling and another one for GOP-level modeling) must be used together in order to speed up the nonlinear bitrate adaptation process.

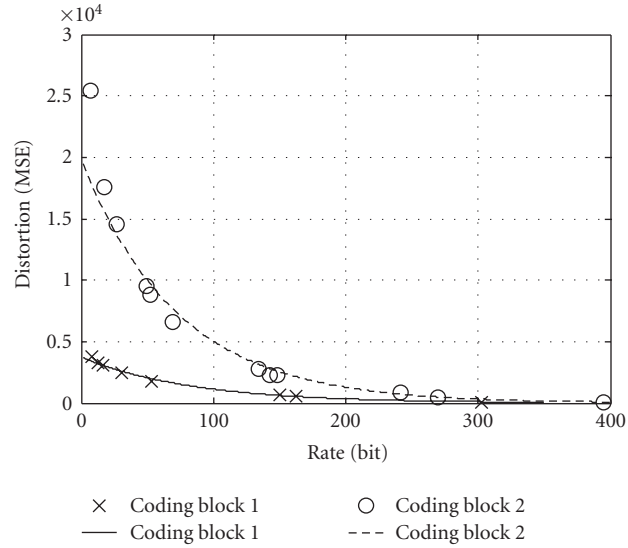

Figure 3: R-D models for coding blocks in a wavelet video codec.

3.1. Parameterized coding block-level R-D models

The application of the rate distortion theory [21] to video codecs is investigated in many literatures [12, 19, 20]. Some literatures [8, 15] apply the function to embedded wavelet coder and make a little empirical adjustment on the parameters. A general R-D model for embedded wavelet coder with square-error distortion measure is as follows:

$$R(D) = \gamma \ln \frac{\omega}{D}, \tag{1}$$

where γ and ω are source-dependent parameters of the logarithmic R-D model. In particular, ω is related to the signal variance of the source.

To verify the accuracy of (1) for wavelet coded sources, we conducted some experiments using the MSRA wavelet video codec reference implementation [5]. The test sequence is stefan in CIF resolution. The results for two coding blocks are shown in Figure 3. Each point in the figure represents an available truncated point in a coding block, and each curve represents the characteristic model for a coding block. The models are calculated by solving the parameters γ and ω in (1) using least-squares-error curve-fitting method. The experiment shows the precision and the reliability of the rate distortion function when applying to coding blocks with different characteristics. Obviously, the R-D information of a coding block can be represented using simply two parameters, γ and ω, instead of 12 or 8 data points as shown in Figure 3.

Although this model fits the R-D characteristics of a single coding block well, it cannot be directly used to represent the R-D model of a complete GOP without losing its accuracy. To reduce the complexity of the tier-two rate adaptation

algorithm of 3D-ESCOT, we still need a better model that represents the R-D information of a GOP of coding blocks.

3.2. GOP-level model and the proposed side information encoding mechanism

To apply the well-known R-D model (1) to efficient multiple adaptations of wavelet video bitstreams, two issues must be addressed first. First of all, an R-D model must be derived for a GOP of coding blocks. Second, the model should facilitate the Lagrange multiplier-based iterative optimization algorithm of 3D-ESCOT. In order to achieve the second goal, the closed-form R-D model (i.e., the γ-ω model in (1)) must be changed to a closed-form R-λ model.

3.2.1. R-lambda model and the model for a GOP of coding blocks

Recall that in (1), the parameter γ depends on the distribution of the source, and the parameter ω is related to the signal variance. For a given value λ, the Lagrange cost function $J(R) = D + \lambda R$ is minimized when $dJ(R)/dR = 0$, that is,

$$\lambda = -\frac{dD(R)}{dR}. \tag{2}$$

Taking the inverse of (1), we have $D(R) = \omega e^{-R/\gamma}$. Substituting $D(R)$ into (2), we obtain the relationship between the Lagrange multiplier and the rate. The R-λ model in coding block level can be written as

$$\lambda = \alpha e^{\beta R}, \tag{3}$$

where the parameters α and β are source-dependent. For each coding block, a parameter pair of (α, β) will be estimated by curve fitting to real R-λ data points.

The GOP-level R-λ model can be extended from the coding block model. First, define $R = \max((1/\beta)\ln(\lambda/\alpha), 0)$ as a nonnegative R-D model. For $\alpha > 0$ and $\beta < 0$, the R-λ model at GOP level is derived as follows:

$$\begin{aligned} R_{\text{GOP}} &= \sum_i R_{\text{block}\,i} = \sum_i \max\left(\frac{1}{\beta_i}\ln\frac{\lambda}{\alpha_i}, 0\right) \\ &= \sum_j \frac{1}{\beta_j}\ln\frac{\lambda}{\alpha_j}, \quad \text{where } \{j \in S \mid \alpha_j > \lambda \text{ in } S\}, \\ &= \left(\sum_j \frac{1}{\beta_j}\right)\ln\lambda - \left(\sum_j \frac{1}{\beta_j}\ln\alpha_j\right). \end{aligned} \tag{4}$$

It is straightforward that the rate of a GOP is the sum of the rates of a group of coding blocks; and the size of the group is related to the λ value. We define the two summation terms in (4) as follows:

$$p_{\text{GOP}} = \sum_j \frac{1}{\beta_j}, \qquad q_{\text{GOP}} = \sum_j \frac{1}{\beta_j}\ln\alpha_j. \tag{5}$$

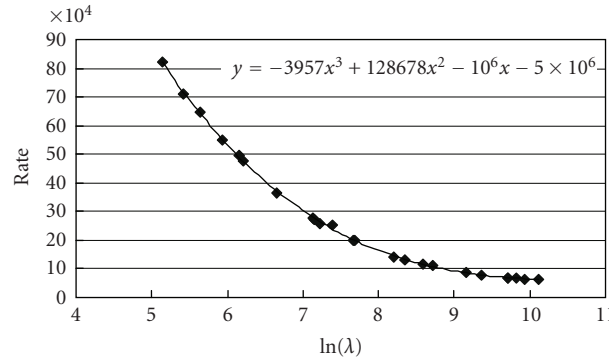

Figure 4: Example of GOP-level R-λ model and real R-D data points.

In order to keep the model simple, we assume that these two summations can be modeled by polynomials as follows:

$$\begin{aligned} P_{\text{GOP}} &= a_1(\ln(\lambda))^{n-1} + a_2(\ln(\lambda))^{n-2} + \cdots + a_n, \\ q_{\text{GOP}} &= b_1(\ln(\lambda))^{n-1} + b_2(\ln(\lambda))^{n-2} + \cdots + b_n. \end{aligned} \tag{6}$$

Finally, the relationship of the GOP-level R-λ model is established:

$$\begin{aligned} R_{\text{GOP}} &= p_{\text{GOP}}\ln\lambda - q_{\text{GOP}} \\ &= \gamma_1(\ln\lambda)^n + \gamma_2(\ln\lambda)^{n-1} + \cdots + \gamma_{n+1}. \end{aligned} \tag{7}$$

Figure 4 illustrates the accuracy of the GOP-level R-λ model for a GOP of the stefan sequence. The order of the function is determined empirically. In general, a cubic function can be used to fit the data points quite well for a wide range of rates.

3.2.2. Proposed rate-distortion side information coding mechanism

In order to allow for multiple-adaptation applications, we must embed the R-D information into the bitstream so that a terminal receiving the bitstream can perform another adaptation with R-D optimality. In addition, we must minimize the size of the R-D information so that it will not consume too much bandwidth. In the following discussions, we assume that the input to the R-D information embedding algorithm is the original full wavelet bitstreams generated by the MSRA encoder. That is, all the R-λ data points for all the fractional bitplane coding pass truncation points are embedded in the bitstream. Although it is not necessary for an embedded wavelet bitstream to assume a layer structure, it is a common practice for the MSRA codec to generate bitstreams with preoptimized quality layers (one for each potential target bitrate). Note that this structure is only for application convenience and is not a necessary feature of wavelet-based scalable video. However, we still preserve this structure in the proposed algorithm.

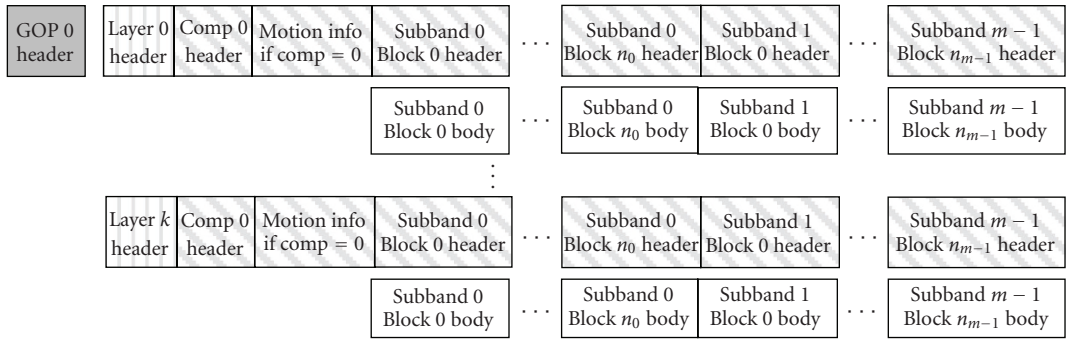

Figure 5: MSRA wavelet bitstream format (please note that there is no need to enforce layer structure for MCTF-based wavelet bitstreams).

The coding block-level model (3) is used as an adaptive model since the source-dependent parameters α and β are estimated based on the input data. Given n pairs of numerical data (λ_i, R_i), $i = 0, \ldots, n-1$, the parameters α and β can be calculated as follows. First, (3) can be rewritten as $\ln\lambda = \ln\alpha + \beta \cdot R$. Therefore, for $n > 2$, we have an overdetermined system of

$$\begin{pmatrix} \ln\lambda_0 \\ \ln\lambda_1 \\ \vdots \\ \ln\lambda_{n-1} \end{pmatrix} = \begin{pmatrix} 1 & R_0 \\ 1 & R_1 \\ 1 & \vdots \\ 1 & R_{n-1} \end{pmatrix} \begin{pmatrix} \ln\alpha \\ \beta \end{pmatrix}. \tag{8}$$

The system can be solved using least-squares estimation. Once the parameters α and β are determined, the relationship between the Lagrange multiplier and rate is directly established. In a similar manner, the GOP-level R-λ model (equation (7)) is adaptively built by the least-squares curve-fitting method. For certain GOP, assume that

$$Y = \begin{pmatrix} R_{\mathrm{GOP1}} \\ R_{\mathrm{GOP2}} \\ \vdots \end{pmatrix},$$

$$A = \begin{pmatrix} (\ln\lambda_1)^n & (\ln\lambda_1)^{n-1} \cdots & 1 \\ (\ln\lambda_2)^n & (\ln\lambda_2)^{n-1} \cdots & 1 \\ \vdots & \vdots & \vdots \end{pmatrix}, \tag{9}$$

$$X = \begin{pmatrix} \gamma_1 \\ \gamma_2 \\ \vdots \\ \gamma_{n+1} \end{pmatrix},$$

where the parameters $\gamma_1, \gamma_2, \ldots, \gamma_{n+1}$ are solved by computing the pseudo inverse $X = (A^T A)^{-1} A^T Y$. As the whole GOP-level R-λ model is established, the λ value can be solved using closed-form solutions for $n < 5$ (typical n is 3).

The algorithm used to embed R-D information into an MSRA encoded bitstream is summarized as follows (note that the original discrete R-D information will be removed).

(1) Search for the optimal Lagrange multiplier at GOP level:
   (a) find the first n pairs of (λ, R) in a quality layer of the input wavelet bitstream (encoded by the original MSRA encoder), and n is typically 4 if cubic model is used in GOP level;
   (b) solve for the parameter $(\gamma_1, \gamma_2, \ldots, \gamma_{n+1})$;
   (c) given the target bitrate, solve the R-λ model for λ. Use the estimated λ to form a bitstream quality layer and obtain another (λ, R) data point;
   (d) add the new (λ, R) pair to the data set;
   (e) iteratively doing the (b)–(d) steps until the R value is close enough to the target bitrate within a tolerable error range TR;
   (f) repeat the procedure for other quality layers.

(2) Embed R-D property of each coding block. In procedure (d), a bitstream quality layer is formed given a GOP-level Lagrange multiplier value. The truncation point of each coding block is determined at the fractional bitplane pass with the nearest Lagrange multiplier value using the R-λ model of the coding block. The parameters α and β are stored for each coding block, and the coding block-level rate allocation can be easily done by computing the inverse R-λ model with a given Lagrange multiplier.

It must be emphasized again that storing a wavelet bitstream in multiple precomputed quality layers is not necessary, but can facilitate adaptation if the target rate happens to match exactly the quality layer rate. If this is not the case, new quality layers must be formed at runtime (e.g., for the second adaptation and above).

4. PROPOSED MULTIPLE-ADAPTATION FRAMEWORK FOR CONTENT-ADAPTIVE FEC-PROTECTED WAVELET BIRSTREAMS

In this section, we present the proposed multiple-adaptation scheme and content-adaptive FEC protection for streaming applications for wavelet codec using the parameterized R-D model introduced in Section 3. The implementation is based on the MSRA wavelet codec [5]. The bitstream of a GOP

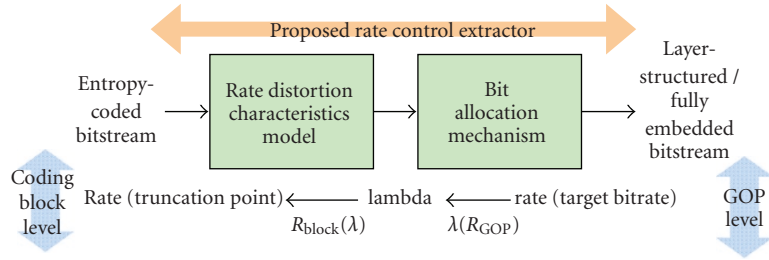

Figure 6: The framework proposed rate control extractor.

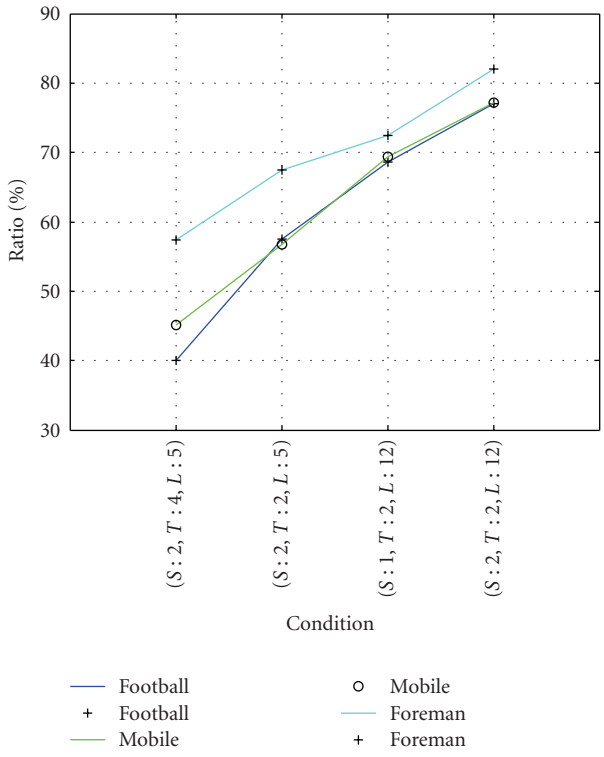

Figure 7: Computation reduction ratio of the proposed method.

encoded using the MSRA codec is organized in the format shown in Figure 5. In Figure 5, m is the total number of temporal and spatial subbands and n_i is the number of coding blocks in subband i.

To prepare a bitstream for multiple-adaptation application over lossy channels, the side information will be used to determine the video data truncation point as well as the level of FEC protection for different fractional bitplanes. Note that the problem of adapting the bitstream to a specific bitrate is not related the quality layer structure of the original bitstream mentioned in Section 3.2.2. If the target rate happens to match one of the preencoded quality layers, the adaptation process is as simple as extracting that quality layer as the output bitstream. However, preencoded quality layer only provides you with coarse-granularity scalability. In this section, it is assumed that the target bistream does not match any of the quality layers in the original wavelet bitstream. Therefore, the adaptation process becomes much more complex.

A bitstream parser extracts the information for the truncated candidates from the headers. After all, the required data are collected, the subband data parsing-and-truncation procedure begins without entropy decoding involved. The parsing-and-truncation module is referred to as the tier-two process of 3D-ESCOT (see Figure 2), and it decides the truncation point in order to meet the resolution, frame rate, and bit rate criteria. The bitstream is then composed again with new header information and truncated body bits. Note that in order to obtain an R-D optimized solution, the parsing-and-truncation process and bitstream composition process will be executed repeatedly until the quality layer converges to the target rate.

4.1. Rate adaptation procedure

R-D optimized adaptation of bitstreams is a complex process. Take the tier-two process of 3D-ESCOT for example. On a PC platform, according to a software profiler, the parsing-and-truncation process of the MSRA reference software accounts for 72% of the computation while the bitstream composition process accounts for another 23% of the load. Note that the implementation of the MSRA reference software is not optimized, therefore this profile may be a rough indication of the computation distribution of the algorithm. The proposed framework (see Figure 6) tries to build a closed-form R-λ relationship for each coding block and each GOP. The rate of each coding block corresponds to the truncation point, and the rate of each GOP corresponds to the target bit rate. These two values are related to each others by the λ value. Therefore, the truncated point for each coding block can be selected given the target bit rate.

Runtime adaptation to a target bitrate becomes a question of searching for a λ value that marks all the truncation points to form a target bitstream that follows the rate constraint. For discrete R-D information used by the original MSRA codec, bisection search is used for determining the λ value. The search process starts from the initial maximum and minimum λ value estimates. By half-eliminating the search range at each iterative step, the search results converge and the λ value which meets the target bitrate is obtained at the end.

For the proposed algorithm, the λ value is estimated in a different way. Because the GOP-level model is a cubic function, the procedure begins with four evenly spaced initial guesses. Then the model is fitted to these data points. The closed-form model is then solved to determine the λ value. If this λ value results in a bitstream that meets the target rate, the process stops, otherwise, the process will be repeated with the new (R, λ) pair replacing the first data point. Usually, the λ estimation process can meet the target bitrate in two steps.

4.2. Adaptation of content-adaptive FEC protection

For video streaming applications, a source-coded video bitstream is first protected by FEC codes, packetized into data packets, and then mapped to IP datagrams. If multiple adaptations are required for a packetized bitstream, recalculation of the FEC codes may be required. In [12], we have proposed a fine-granularity unequal error protection mechanism for wavelet-based video. The mechanism uses the original MSRA R-D side information to fine-tune the protection level of coefficients of different fractional bitplanes. The approach maximizes the use of protection bit budget to achieve better performances than existing approaches of unequal error protection based on different syntax element types. However, multiple adaptations are not possible in [12] since side information were considered too expensive to protect and transmit.

In this section, the adoption of the proposed side information coding mechanism is incorporated into the content-adaptive FEC framework to facilitate multiple adaptations. For each group of video bitstream data, an (n, k) Reed-Solomon (RS) code can be applied to add resiliency to the data. For (n, k) RS code, n is the codeword length, k is the number of video data symbols (e.g., a symbol is composed of 8 bits of bitstream data). The number of parity symbols is $2s$, where $2s = n - k$. This means that if burst errors occur during transmission, the RS decoder can correct up to s errors and detect up to $2s$ errors per codeword.

Note that for content-adaptive FEC protection, the degree of protection level s should be based on the importance of the video data. In a wavelet video bitstream, the importance of the coefficients within a coding block in a particular subband can be ranked based on the R-D side information of the coding block. After wavelet decomposition, the subbands can be arranged and indexed from low to high frequencies. The smaller the index is, the lower the frequency is. Therefore, each coding block in subband i has a temporal subband index ω_i and a spatial subband index τ_i. The importance of the coefficients in a coding pass is first determined by the importance of the coding block it is located in. The importance of a coding block is in turn determined by the subband it is located in. The importance factor W_i of a coding block is computed by

$$W_i = \exp\left[(-1)\cdot\left(\frac{(T-\omega_i)\cdot U_1}{T} + \frac{1}{(S-\tau_i)}\right)\right], \quad (10)$$

where T is the maximum temporal-level index, S is the maximum spatial subband index, and U_1 is a weighting factor.

The level of FEC protection is defined by the value s, the number of correctable symbols. Without loss of generality, assume that the bitstream of a coding block j is divided into m codewords. The protection level $s_{j,x}$ of the coefficients in coding pass x of coding block j is computed by

$$\hat{s}_{j,x} = \left\lfloor \left(\frac{\alpha_j \cdot \exp\left(\beta_j \cdot \sum_{k=0}^{x} R_{j,k}\right)}{\omega}\right)\cdot n_{pl}\cdot W_j \right\rfloor,$$
$$s_{j,x} = \hat{s}_{j,x} + o, \quad o = \begin{cases} 0 & \text{if } \hat{s}_{j,x} \text{ is even,} \\ 1 & \text{if } \hat{s}_{j,x} \text{ is odd,} \end{cases} \quad (11)$$

where $x = 0, 1, \ldots, m-1$, the parameters α_i and β_i are the close-form R-λ model (3) parameters for the coding block j, $R_{j,x}$ is the length of the xth RS codeword in coding block j, n_{pl} denotes the estimated number of packet losses per second, and ω is a scale factor determined empirically. Equation (11) is designed so that $s_{i,0} \geq s_{i,1} \geq \cdots \geq s_{i,m-1}$, that is, the level of protection decreases following fractional bitplane coding pass order. Note that the operation $\lfloor \cdot \rfloor$ stands for "taking the largest integer that is smaller than or equal to the parameter."

For some multiple-adaptation applications, the second (and above) adaptations may be due to the change of device capabilities instead of channel conditions. For such case, there is no need to recompute the FEC codes since the level of protection does not change. However, repacketization may still be necessary for efficient transmission of the readapted data.

5. EXPERIMENTAL RESULTS

In this section, some experiments on the proposed algorithm are conducted using the MSRA scalable video codec, with the MPEG test sequences, Stefan, Foreman, Mobile, and Football in CIF resolution.

5.1. Computational cost reduction for runtime bitstream adaptation

In this section, the number of iterations of the tier-two 3D-ESCOT nonlinear R-D optimization process is used as the measure for complexity analysis. This is a reasonable complexity measure since, as mentioned in Section 2, each iteration of the nonlinear optimization must perform three things: R-D point determination, parsing and truncation of fractional bitplane coding passes, and bitstream composition. A software profiler was used to estimate the ratio of required machine instructions for these modules for Pentium instruction sets. On average, for each iteration, the parsing and truncation and bitstream composition together account for more than 95% of the complexity while the R-D point determination accounts for less than 1% of the complexity. Therefore, the overhead of R-D point determination is negligible.

The number of iterations required before the solution converges for the proposed method and the bisection search

Table 1: Number of iterations for the MSRA and proposed approach. S is Number of spatial scalabilities, T is Number of temporal transform, L is Number of bitstream layers.

Sequence	*MSRA bisection*	*R-λ model*	*Complexity saving ratio*
Mobile ($S:2, T:4, L:5$)	9.67	5.30	45.17%
Mobile ($S:2, T:2, L:5$)	9.67	4.18	56.77%
Mobile ($S:1, T:2, L:12$)	14.83	4.55	69.32%
Mobile ($S:2, T:2, L:12$)	14.83	3.39	77.14%
Foreman ($S:2, T:4, L:5$)	10.68	4.55	57.41%
Foreman ($S:2, T:2, L:5$)	10.68	3.48	67.43%
Foreman ($S:1, T:2, L:12$)	14.35	3.95	72.47%
Foreman ($S:2, T:2, L:12$)	14.92	2.68	82.04%
Football ($S:2, T:4, L:5$)	7.84	4.70	40.05%
Football ($S:2, T:2, L:5$)	7.67	3.26	57.50%
Football ($S:1, T:2, L:12$)	13.56	4.26	68.58%
Football ($S:2, T:2, L:12$)	13.62	3.12	77.09%

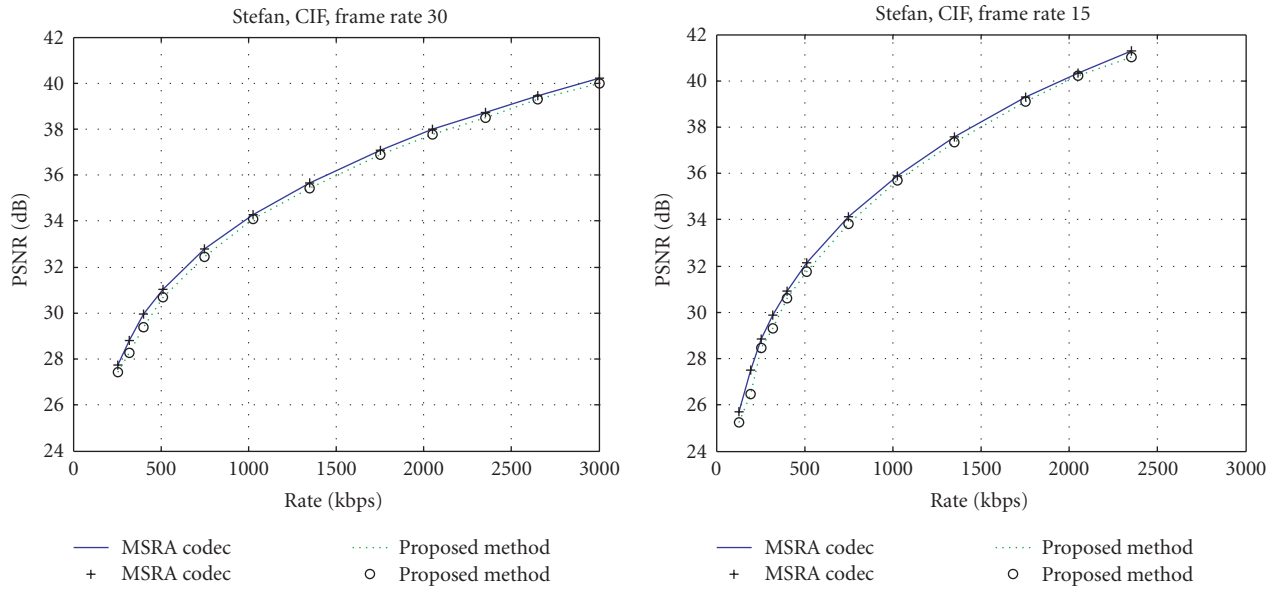

Figure 8: PSNR performance comparison of Stefan.

used in the MSRA codec are shown in Table 1. The coding parameters used in the experiments are as follows. The GOP size is 64 and the frame rate is 30 fps. A cubic polynomial is used for the proposed GOP-level model, and the bitrate error threshold is set to 3% of the target bitrate. When the number of layers for each resolution and frame rate setting increases, the proposed search procedure can converge even faster by taking advantage of the R-λ model from the previous layer. According to the experiments, the average complexity saving ratio is over 64%. The saving ratio of iteration times is about 60% when the layer number is 5, and up to 80% when the layer number is 12 (see Figure 7).

Since the proposed mechanism allocates rate for each coding block differently from that of the MSRA codecs, the rate distribution (and quality) in a GOP is different from that of the MSRA codecs. The coding efficiency is shown in Figures 8, 9, and 10. The test sequences are Stefan, Football, and Foreman in CIF resolution and are truncated at frame rates 30 and 15. The figures show that the proposed rate adaptation mechanism achieves similar PSNR performance in comparison with that of the MSRA codecs at any rates. The average PSNR degradation is less than 0.25 dB.

5.2. Side information saving for multiple adaptations

The experimental result in Table 2 shows the saving ratio in different resolutions and frame rates for different sequences in a multiple-adaptation scenario. The average saving ratio of the side information is about 54.73%, and the side information percentage in the bitstream is reduced from 3.39%

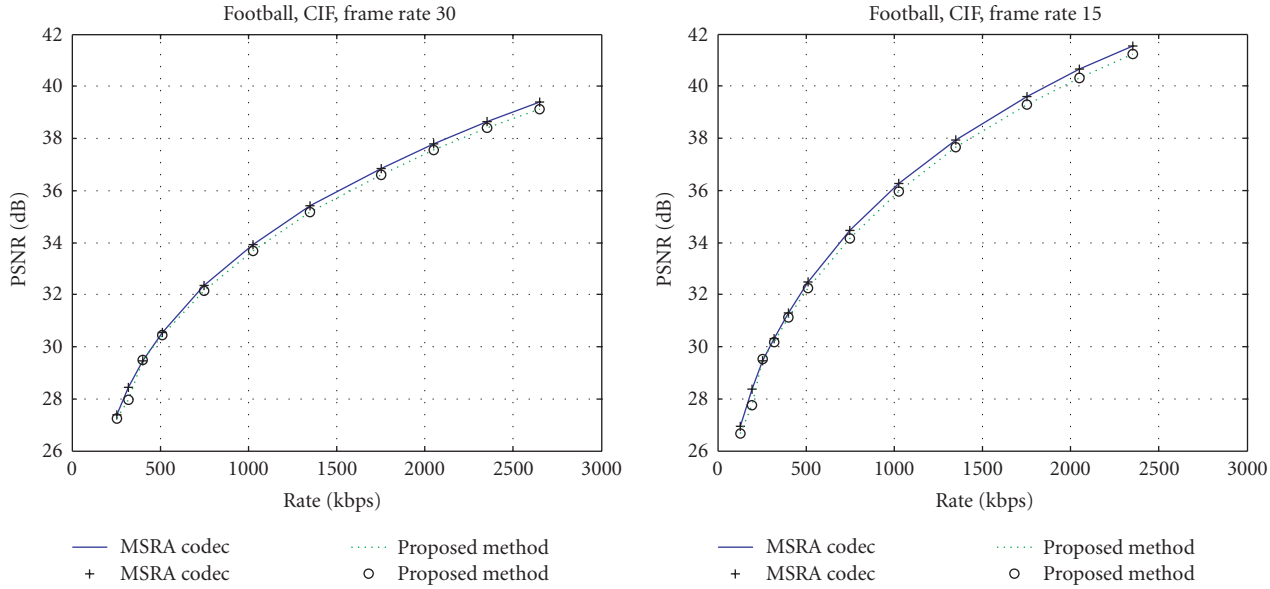

Figure 9: PSNR performance comparison of Football.

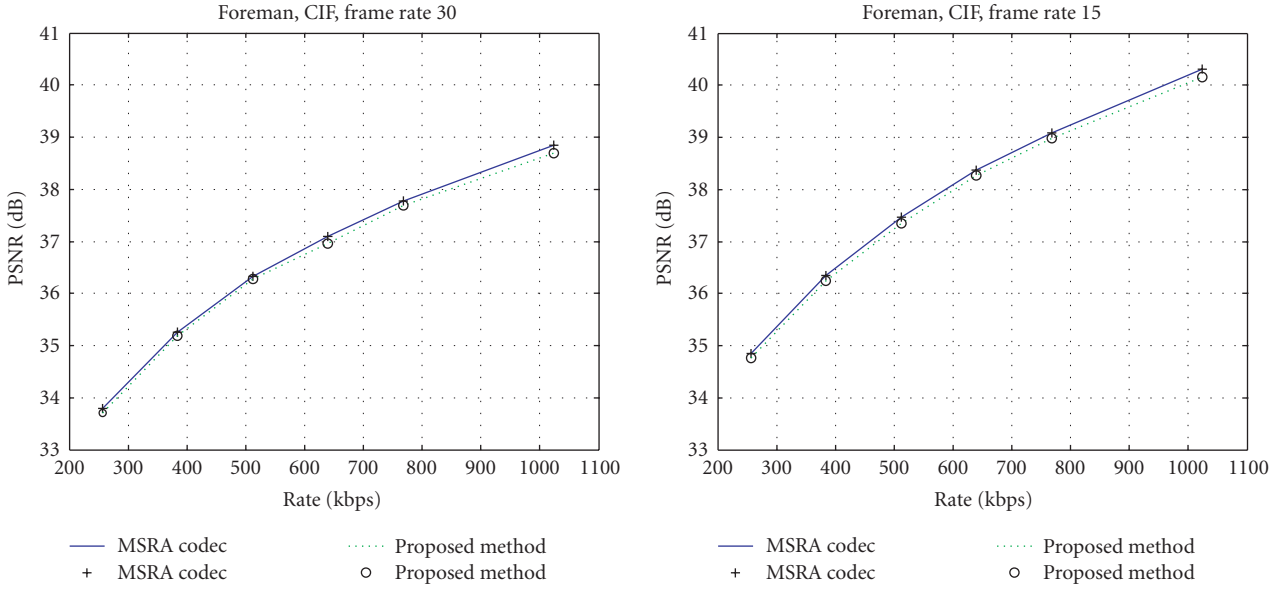

Figure 10: PSNR performance comparison of Foreman.

to 1.6%. Table 3 illustrates the saving ratio for different GOP sizes. One can observe that the proposed method can properly adapt for a variety of GOP lengths. In these experiments, the video sequences are encoded at 15 fps (150 frames) with temporal level 2 and single quality layer.

It is important to know that the original MSRA side information is already in compressed format. Therefore, it is not possible to simply use a lossless compression technique to compress it. To demonstrate this point, two popular lossless compression utilities, WinZIP and WinRAR, are used to compress the side information of the original MSRA bistreams. The results are shown in Table 4 (the same encoding settings as those for Table 3). From Table 4, one can see that the average saving ratio using lossless compressor is about 2% while the proposed approach is more than 50%.

5.3. Content-adaptive FEC protection experiments

For the evaluation of the performance of the content-adaptive FEC protection, the CIF version of the standard

Table 2: Side information saving ratio.

Sequence name	*Resolution/frame rate/bit rate(kbps)*	*Side information bits (% in bitstream)*		*Saving ratio*
		MSRA	*Proposed method*	
Mobile	CIF/15/4096	898,872 (2.14%)	307,120 (0.73%)	65.89%
	CIF/15/2048	557,560 (2.66%)	245,696 (1.71%)	56.02%
	CIF/15/512	163,944 (3.13%)	76,780 (1.47%)	53.04%
	CIF/30/2048	600,648 (2.86%)	297,500 (1.42%)	50.35%
	CIF/30/512	164,864 (3.15%)	77,734 (1.48%)	53.02%
Foreman	CIF/15/1152	433,112 (3.67%)	227,360 (1.93%)	47.41%
	CIF/15/640	290,000 (4.43%)	140,963 (2.15%)	51.47%
	CIF/15/256	127,120 (4.86%)	58,581 (2.24%)	53.91%
	CIF/30/1152	505,224 (4.28%)	245,089 (2.08%)	51.40%
	CIF/30/640	308,480 (4.71%)	142,710 (2.18%)	53.72%
Stefan	CIF/15/4096	805,856(1.92%)	278,960 (0.67%)	65.10%
	CIF/15/3072	711,888(2.26%)	312,435 (0.99%)	56.19%
	CIF/15/1024	350,584 (3.34%)	156,217 (1.49%)	55.39%
	CIF/15/512	218,112 (4.16%)	103,215 (1.97%)	52.64%
	CIF/30/3072	883,672 (2.81%)	380,304 (1.21%)	56.94%
	CIF/30/1024	404,744 (3.86%)	190,152 (1.81%)	53.11%
Average		3.39%	1.60%	54.73%

Table 3: Side information bit overhead versus GOP size.

Sequence name	*Resolution/frame rate/bit rate(kbps)*	*Side information bits (% in bitstream)*		*Saving ratio*
		MSRA	*Proposed method*	
Mobile	CIF/64/4096	898,872 (2.14%)	307,120 (0.73%)	65.89%
	CIF/32/4096	1,011,720 (2.41%)	307,440 (0.73%)	69.71%
	CIF/16/4096	890,696 (2.12%)	306,720 (0.73%)	65.57%
Foreman	CIF/64/1152	433,112 (3.67%)	227,360 (1.93%)	47.41%
	CIF/32/1152	552,744 (4.69%)	284,200 (1.95%)	58.85%
	CIF/16/1152	431,088 (3.66%)	227,200 (1.93%)	47.27%
Stefan	CIF/64/4096	805,856 (1.92%)	278,960 (0.67%)	65.1%
	CIF/32/4096	1,005,192 (2.40%)	278,720 (0.66%)	72.5%
	CIF/16/4096	759,384 (1.81%)	278,560 (0.66%)	63.54%
Average		2.76%	1.11%	61.76%

MPEG test sequences Stefan and Mobile are used. Those sequences are encoded using the MSRA codec at 15 frames per second and the GOP size of 64 frames. Four levels of 5/3 MCTF temporal decomposition and three levels of 9/7 wavelet spatial decomposition are used for subband decomposition. The number of luma coding blocks is 1024 and the number of chroma coding blocks is 608.

Based on the reports in [22, 23], we have applied 5% packet loss rate to the IP packets in order to evaluate the performance of the proposed content-adaptive FEC protection system. Adaptive FEC protection using the proposed side information is compared against that using the original MSRA side information. The PSNR of the luma channel of the reconstructed video sequences is shown in Figures 11 and 12. In either case, the maximal packet loss protection level can only recover up to 4% packet losses on average so that we can evaluate the differences in quality degradation using different side information. As one can see from the figures, the proposed side information (using closed-form R-D model) is as efficient as the original side information (using discrete R-D data points) for content-adaptive FEC protection.

6. CONCLUSIONS AND FUTURE WORK

In this paper, we have proposed a framework for wavelet video multiple adaptations and content-adaptive FEC protection. The proposed framework uses two closed-form R-λ models to reduce the size of the R-D side information

Table 4: Lossless compression of side information.

Sequence name	*Resolution/GOP size/bit rate(kbps)*	*Side information bits (% for saving ratio)*		
		MSRA	*WinZIP* 8.1	*WinRar* 3.42
Mobile	CIF/64/4096	898,872	869,288 (3.29%)	869,256 (3.29%)
	CIF/32/4096	1,011,720	982,696 (2.87%)	983,392 (2.80%)
	CIF/64/512	163,944	162,464 (0.90%)	162,280 (1.01%)
	CIF/32/512	164,144	162,464 (1.02%)	162,336 (1.10%)
Foreman	CIF/64/1152	433,112	421,032 (2.79%)	421,432 (2.70%)
	CIF/32/1152	552,744	542,904 (1.78%)	543,008 (1.76%)
	CIF/64/640	290,000	284,064 (2.05%)	283,784 (2.14%)
	CIF/32/640	290,488	284,600 (2.03%)	284,384 (2.10%)
Stefan	CIF/64/4096	805,856	780,752 (3.12%)	780,584 (3.14%)
	CIF/32/4096	1,005,192	979,080 (2.60%)	980,032 (2.50%)
	CIF/64/1024	350,584	343,536 (2.01%)	343,312 (2.07%)
	CIF/32/1024	350,504	343,568 (1.98%)	343,088 (2.12%)
Average			2.20%	2.23%

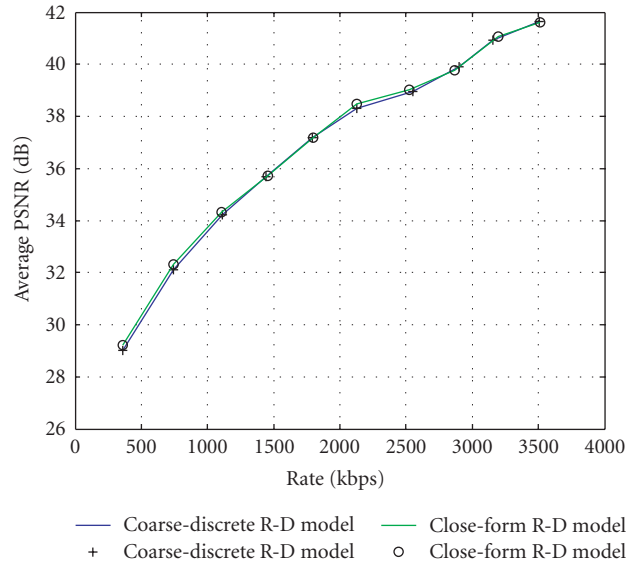

Figure 11: Content-adaptive FEC test for the Stefan sequence (5% losses).

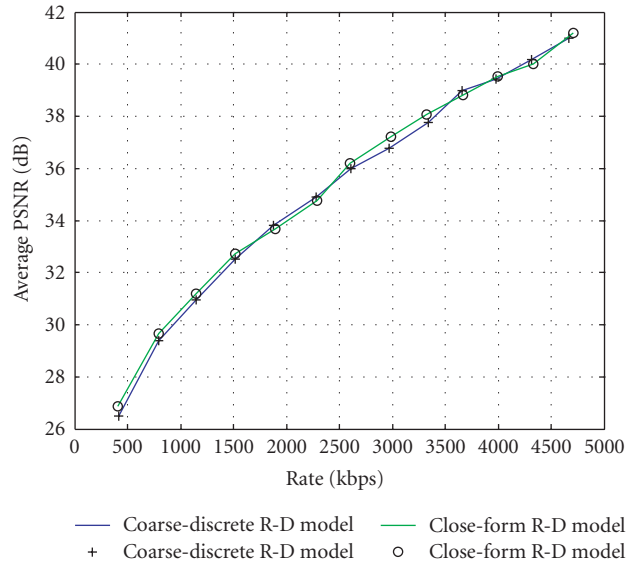

Figure 12: Content-adaptive FEC test for the Mobile sequence (5% losses).

embedded in the coded bitstream by more than 50% on average while maintaining the accuracy of the rate-distortion information of the video data. In addition, the proposed technique can reduce the computational complexity of the tier-two 3D-ESCOT wavelet adaptation process by more than 64% on average. Although the existing model achieves good performance, there are still rooms for improvement in the future. For example, at high resolution and high bitrate, the motion vector information is quite large and is not covered by existing R-D model. There have been some researches on scalable motion vector coding. Similar ideas can be applied to the construction of an R-D model for motion vector bits to further increase the performance.

ACKNOWLEDGMENT

This research is partly funded by National Science Council, Taiwan, under Grant no. NSC 95-2221-E-009-073-MY3.

REFERENCES

[1] ISO/IEC JTC 1/SC 29/WG 11, 14496-2: 2002 Information Technology - Coding of Audio-Visual Objects—Part 2: Visual 3rd Edition, March 2003.

[2] W. Li, "Overview of fine granularity scalability in MPEG-4 video standard," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 11, no. 3, pp. 301–317, 2001.

[3] S.-J. Choi and J. W. Woods, "Motion-compensated 3-D subband coding of video," *IEEE Transactions on Image Processing*, vol. 8, no. 2, pp. 155–167, 1999.
[4] J. Xu, Z. Xiong, S. Li, and Y.-Q. Zhang, "Three-dimensional embedded subband coding with optimized truncation (3-D ESCOT)," *Applied and Computational Harmonic Analysis*, vol. 10, no. 3, pp. 290–315, 2001.
[5] ISO/IEC MPEG Video Group, *Wavelet Codec Reference Document and Software Manual V1.0*, MPEG Document N7573, July 2005.
[6] A. Said and W. A. Pearlman, "A new, fast, and efficient image codec based on set partitioning in hierarchical trees," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 6, no. 3, pp. 243–250, 1996.
[7] D. Taubman, "High performance scalable image compression with EBCOT," *IEEE Transactions on Image Processing*, vol. 9, no. 7, pp. 1158–1170, 2000.
[8] P.-Y. Cheng, J. Li, and C.-C. J. Kuo, "Rate control for an embedded wavelet video coder," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 7, no. 4, pp. 696–702, 1997.
[9] T. Chu and Z. Xiong, "Combined wavelet video coding and error control for Internet streaming and multicast," *EURASIP Journal on Applied Signal Processing*, vol. 2003, no. 1, pp. 66–80, 2003.
[10] J. Dong and Y. F. Zheng, "Content-based retransmission for 3-D wavelet video streaming on the Internet," in *Proceedings of International Symposium on Information Technology: Coding and Computing (ITCC '02)*, pp. 452–457, Las Vegas, Nev, USA, April 2002.
[11] W.-T. Tan and A. Zakhor, "Real-time Internet video using error resilient scalable compression and TCP-friendly transport protocol," *IEEE Transactions on Multimedia*, vol. 1, no. 2, pp. 172–186, 1999.
[12] C.-P. Ho and C.-J. Tsai, "Content-adaptive packetization and streaming of wavelet video over IP networks," *EURASIP Journal on Image and Video Processing*, vol. 2007, Article ID 45201, 12 pages, 2007.
[13] A. Aminlou and O. Fatemi, "Very fast bit allocation algorithm, based on simplified R-D curve modeling," in *Proceedings of the 10th IEEE International Conference on Electronics, Circuits and Systems (ICECS '03)*, vol. 1, pp. 112–115, Sharjah, United Arab Emirates, December 2003.
[14] W. Yu, "Integrated rate control and entropy coding for JPEG 2000," in *Proceedings of Data Compression Conference (DCC '04)*, pp. 152–161, Snowbird, Utah, USA, March 2004.
[15] J. Li, C.-C. J. Kuo, and P.-Y. Cheng, "Embedded wavelet packet image coder with fast rate-distortion optimized decomposition," in *Visual Communications and Image Processing*, vol. 3024 of *Proceedings of SPIE*, pp. 1077–1088, San Jose, Calif, USA, February 1997.
[16] Y. Shan, S. Yi, S. Kalyanaraman, and J. W. Woods, "Two-stage FEC scheme for scalable video transmission over wireless networks," in *Multimedia Systems and Applications VIII*, vol. 6015 of *Proceedings of SPIE*, pp. 173–186, Boston, Mass, USA, October 2005.
[17] W.-T. Tan and A. Zakhor, "Video multicast using layered FEC and scalable compression," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 11, no. 3, pp. 373–386, 2001.
[18] J. Goshi, A. E. Mohr, R. E. Ladner, E. A. Riskin, and A. Lippman, "Unequal loss protection for H.263 compressed video," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 15, no. 3, pp. 412–419, 2005.
[19] J. T. H. Chung-How and D. R. Bull, "Robust H.263+ video for real-time Internet applications," in *Proceedings of IEEE International Conference on Image Processing (ICIP '00)*, vol. 3, pp. 544–547, Vancouver, BC, Canada, September 2000.
[20] J.-R. Chen, C.-S. Lu, and K.-C. Fan, "A significant motion vector protection-based error-resilient scheme in H.264," in *Proceedings of the 6th IEEE Workshop on Multimedia Signal Processing (MMSP '04)*, pp. 287–290, Siena, Italy, September-October 2004.
[21] C. Shannon, "A mathematical theory of communication," *Bell System Technical Journal*, vol. 27, pp. 379–423, 623–656, 1948.
[22] J. M. Boyce and R. D. Gaglianello, "Packet loss effects on MPEG video sent over the public Internet," in *Proceedings of the 6th ACM International Conference on Multimedia*, pp. 181–190, Bristol, UK, September 1998.
[23] K. Lai, M. Roussopoulos, D. Tang, X. Zhao, and M. Baker, "Experiences with a mobile testbed," in *Proceedings of the 2nd International Conference on Worldwide Computing and Its Applications (WWCA '98)*, pp. 222–237, Tsukuba, Japan, March 1998.

Ya-Huei Yu was born in Taipei, Taiwan, in 1980. In 2003 and 2005, she received the B.S. and M.S. degrees in computer science and information engineering from National Chiao Tung University, Hsinchu, Taiwan. She joined MediaTek Inc. in 2005. Her research interests include image and video compression techniques and rate-distortion modeling of video contents.

Chien-Peng Ho received the M.S. degree in electrical engineering from National Taiwan University of Science and Technology, Taipei, Taiwan, in 1995. He is currently pursuing the Ph.D. degree in the Department of Computer Science, National Chiao Tung University, Hsinchu, Taiwan. His research interests include video compression and transmission.

Chun-Jen Tsai received the B.S. degree in mathematics from Fu-Jen Catholic University, Taiwan, in 1989, the M.S. degree in computer science and information engineering from National Taiwan University, Taipei, in 1992, and the Ph.D. degree in electrical engineering from Northwestern University, Evanston, Ill, in 1998. From 1999 to 2002, he was with PacketVideo Corporation, San Diego, CA, where he was working on video codec for embedded systems and wireless multimedia streaming system design. Since 2000, he has been a US National Body Delegate for ISO/IEC MPEG Organization. In 2002, he joined the Department of Computer Science and Information Engineering, National Chiao Tung University, Hsinchu, Taiwan, where he is currently an Assistant Professor. His current research interests are in multimedia embedded systems hardware/software codesign, theory and optimization of video compression technologies, and distributed multimedia systems.