行政院國家科學委員會專題研究計畫　成果報告

子計畫三:MPEG 多媒體傳輸機制及通訊協定在嵌入式行動平台上的分析設計(III)

計畫類別： 整合型計畫
計畫編號： NSC94-2219-E-009-009-
執行期間： 94 年 08 月 01 日至 95 年 07 月 31 日
執行單位： 國立交通大學資訊工程學系(所)

計畫主持人： 蔡淳仁

計畫參與人員： 何健鵬、高政汗、張文潔、蔡雅婷

報告類型： 完整報告
報告附件： 出席國際會議研究心得報告及發表論文
處理方式： 本計畫可公開查詢

中　華　民　國 95 年 10 月 16 日

行政院國家科學委員會專題研究計畫成果報告

MPEG 多媒體傳輸機制及通訊協定
在嵌入式行動平台上的分析設計

Design and Analysis of MPEG Multimedia Transport Mechanisms and Protocols for Embedded and Mobile Environment

計畫編號：NSC 94-2219-E-009-009,
執行期限：94 年 8 月 1 日至 95 年 7 月 31 日

主持人：蔡淳仁 國立交通大學資訊工程系
參與人員：何健鵬、高政汗、張文潔、蔡雅婷　國立交通大學資訊工程系

中文摘要

在三年的整合計畫中，本子計畫除了配合總計畫團隊，協助 MPEG 標準 ISO/IEC 21000-12：Test Bed for MPEG-21 Resource Delivery 的制定之外，在過去三年共發展了以下幾項在異質性網路及不同能力的終端設備間進行多媒體串流傳輸相關的技術。在第一年的時候，除了開發 Test Bed，主要是研究現有的流量控制和容錯機制的設計。第二年則改進了現有的碼率失真最佳化（rate-distortion optimized）可調式串流傳輸機制並整合進計畫團隊開發的 MPEG Test Bed 標準測試環境中。另外，在第二年的計畫中，也針對未來多媒體嵌入式系統的相當有潛力的多次內容調適(multiple adaptations)的應用，設計出以 wavelet 為基礎之 scalable video codec 的流量控制機制。最後，在第三年的研究中，我們更進一步提出了更好的碼率失真最佳化（rate-distortion optimized）可調式串流傳輸機制。在此系統中，我們會根據封包內容的重要度進行不同的 Reed-Solomon 容錯保護。在研究成果的產出方面，根據第二年和第三年的研究，我們分別寫了兩篇期刊論文，目前正在審核中。

關鍵詞：MPEG-4、MPEG-21、多媒體串流傳輸、嵌入式行動多媒體系統、碼率失真最佳化、多媒體傳輸流量控制、可調式視訊壓縮

Abstract

There are several achievements in this project for the past three years. First of all, the project team participated in the MPEG standard activities and developed the International Standard, ISO/IEC 21000-12: Test Bed for MPEG-21 Resource Delivery. In addition, we have developed several technologies for multimedia streaming over heterogeneous networks and devices. In the first year of the project, in addition to participating in the development of the test bed, we have investigated existing rate control and error control mechanisms for video streaming. During the second year of the project, we have improved one of the existing rate-distortion optimized scalable streaming schemes and integrated it into the test bed. Furthermore, we have looked into the problem of multiple adaptations of scalable video streaming. Multiple adaptations have a lot of important applications for future heterogeneous environment, especially for embedded multimedia devices. Finally, for the third year of the project, we have designed a better rate-distortion optimized scalable streaming system. In this system, we have developed a content-adaptive Reed-Solomon error control mechanism with interleaving packetization to cope with packet losses over IP networks. Both the research results in the second and third years have been submitted to international journals and are currently under the review process.

Keywords: MPEG-4, MPEG-21, multimedia streaming, digital content transport, flow control, embedded multimedia systems, rate-distortion optimization, scalable video coding.

目錄

Part I: 綜合討論

一、前言 1
二、研究目的 1
三、文獻探討 1
四、結果與討論 1
五、計劃成果自評 2
六、與計畫相關之已發表文獻 3

Part II: 研究成果詳細內容

研究成果一：**Multiple Adaptation and Content-Adaptive FEC Using Parameterized R-D Model for Embedded Wavelet Video** ……..…… 4
研究成果二：**Content-Adaptive Packetization and Streaming of Wavelet Video over IP Networks** …….. 18

一、前言

本報告是過去三年的整合計畫的完整報告。在最後一年的計畫中，我們整合了過去三年的研究，寫了兩篇尚未發表的期刊論文，目前已交付國際期刊審查中。本報告內容主要是以這兩篇論文為主，詳見第頁開始的研究成果。

二、研究目的

過去的串流傳輸系統的設計重點是在於同一時間內要能服務最多的客戶端為主，而且多半假設所有連到伺服器的終端機都是透過類似的網路以及具有相似的計算能力。本計畫的主要目的在於配合 MPEG-21 的理念，設計出一個具有實用性系統，可以橫跨不同的網路架構和在不同的客戶端設備上（PC、手機、PDA 等等）提供高品質多媒體傳輸播放服務。因此，這個傳輸系統必須能動態地根據不同的應用環境調整資料包裝和傳輸的方式。簡言之，一個數位多媒體傳輸系統的架構必需包含流量控制和容錯機制。另外，依據客戶端的能力來調整媒體資料流品質的能力也是十分重要的。

為了在頻寬和封包掉失率隨時在改變的環境下給使用者最好的串流傳輸品質，許多人嘗試對傳輸的資料進行碼率失真最佳化(Rate-Distortion)分析，然後根據分析結果來傳輸視訊串流。過去所提出的碼率失真最佳化的網路傳輸的流量控制機制裏，最大的問題在於封包漏失所造成的失真(distortion）沒有辦法量化。許多現有系統的做法在封包漏失所造成的失真的估測都十分不切實際（或完全忽略了）。我們提出了一個較實際的方法，可以合理的把封包失真量化。

要達到碼率失真最佳化的網路串流傳輸，除了要根據網路頻寬和封包漏失率來估測出最好的串流資料量，還必須能在算出的串流資料量限制下，從 scalable video 的位元流中，抽出最好的子位元流。所以我們針對這方面的需求，設計出了一個快速收斂的小波轉換視訊壓縮法(wavelet video coding)的流量控制機制，另外，更進一步設計了可以有效進行多次碼率失真最佳化的子位元流抽取(multiple adaptation)的機制。

另外，多媒體串流資料傳輸之前，一般的系統會進行 Forward Error Correction (FEC)或 Automatic Retransmission reQuest (ARQ)的保護。然而，在目前公開發表的系統中，最多只是針對漸層式編碼可調式編碼的的不同傳輸層做不同程度的 FEC 保護。而我們則改進了這項缺點。在本計畫第三年的成果中，我們提出了一套機制，在碼率失真最佳化的原則下，能較精確地對視訊內容依其重要性進行不同程度的 Reed-Solomon 編碼保護，並配合 data interleaving 的封包包裝法來強化資料對封包掉失的容錯度。

三、文獻探討

本計畫的總合成果主要有兩大方向，首先是以適用於小波轉換視訊壓縮法的多次碼率失真最佳化的子位元流抽取(multiple adaptation)的機制，以及小波轉換視訊串流傳輸的能自動隨視訊內容重要性和即時封包掉失率來調整 FEC 保護強度的機制。關於這兩個研究方向的文獻探討，請參見綜合報告之後的成果報告一和二。

四、結果與討論

前面大致提到了本計畫在過去三年除了參加 MPEG 標準制訂之外的三項主要研究成果，首先是一個較實用的碼率失真最佳化串流傳輸系統、其次是一個較適合嵌入式系統使用的低複雜度可多次內容調適的流量控制機制、最後則是一個根據 IP 封包網路串流傳輸特性所設計的自動隨內容重要性調整的（content-adaptive）FEC 保護機制。

本節首先描述本計畫在碼率失真最佳化串流傳輸系統方面的創新設計。目前在這方面較知名的是由 P. A. Chou 等人所發展的系統。這個系統有兩大缺點。首先是他們使用封包掉失率來估測碼率失真最佳化分析中的失真。這是很不實際的做法。其次，該系統為了避開傳統 ARQ 對封包回傳時間很敏感的問題，因此採用封包預先重覆傳送的方法來降低失真，這也是

很沒有效率的容錯保護機制。因此這套方法目前發表的成果以理論分析為主，在實作上有很多細節並沒有提出解決方案，而且在頻寬變化大的網路環境下，串流傳輸最難達到的平滑播放要求也沒有考量。

在本計畫的碼率失真最佳化串流傳輸系統中，我們把封包漏失所造成的失真，轉化為不同程度的 FEC 保護所造成的失真。舉例而言，10^{-3} 的封包漏失率造成的失真，就相當於 10^{-3} 的 FEC 的 error protection 導致 data rate 降低所造成的失真。整個系統可以分成兩大部份：媒體封包相依性控制：媒體封包相依控制和碼率失真最佳化傳輸控制。在提出的系統中，碼率失真最佳化傳輸模組會根據資料內容的移動量資訊，來調整流量控制時的最佳化決定。簡單地說，一般的碼率失真最佳化控制架構必需在資料單元群組之間利用碼率及失真的 Lagrangian cost function 來算出最小值的解來有效率的分配時間和頻寬的網路資源。而本計畫提出的系統則會同時利用 FEC 的編碼率和每一段視訊資料所含的移動量大小來校正傳統的 cost function 以求得更佳的效果。

在可調式位元串流傳輸中，影像資料可以分成好幾次傳送，每次的傳送都可以幫助解碼端得到更接近於原影像資料的重建訊號，因此可調式位視訊編碼法必須必須能支援多樣化的調適運作(adaptation operations)以針對不同應用產生有效可解碼的位元流。前面所提到的系統設計，主要是以 MPEG-4 FGS 的編碼法（從 2005 年起，FGS 已經不屬於 MPEG-4 標準的一部份）為可調式編碼的核心。但在第一年的研究過程中，我們發現 FGS 的諸多限制，因此在第二年開始的研究中，就開始轉向採用小波轉換視訊壓縮法。

小波轉換視訊壓縮法的流量控制機制，一般系統(e.g. JPEG 2000 及 3-D ESCOT)是利用在編碼階段所產生的碼率失真資料表，配合二分搜尋法，來決定在某個頻寬條件下最佳的視訊位元流子集合。在本計畫中，我們設計了一個比碼率失真表更有效的雙參數碼率失真模型來進行最佳的位元流切割點的快速搜尋。也因為我們所提出的視訊視訊資料模型比較精簡也更有效率，我們可以把它隨著抽取出的位元流一起傳輸到接收端以進行多次的碼率失真最佳化子位元流切割。

在第三年的計畫中，我們開發了一個能自動隨視訊內容重要性和即時封包掉失率來調整 FEC 保護強度的機制。這部份的設計主要是針對 IP 網路封包串流傳輸的應用而開發的。目前所有的 IP 網路的串流傳輸系統都會由 layer-2 以下的通訊協定處理位元錯誤，而在 layer-3 以上往往只能看到封包掉失的錯誤。現有系統多半以 ARQ 封包重傳的方法或是用 Reed-Solomon 編碼配合資料交錯安排（data interleaving）來達到修復掉失的封包的功能。前者往往不適用封包回傳時間（round-trip time）較長的網路，而後者則是很難達到碼率失真最佳化的原則。在本計畫提出的方法，則是利用我們在第二年時所發展的雙參數碼率失真模型，並根據小波轉換的視覺效應，定出一個能夠隨視訊內容和即時網路封包掉失率來調整 Reed-Solomon 保護強度的 FEC 封包編碼保護機制。

五、計劃成果自評

總合三年的成果，和原計畫提出的目標相當吻合。在達成預期目標情況方面有以下數點：

1. 配合總計畫團隊，成功地制訂出 ISO/IEC TR 21000-12：Test Bed for MPEG-21 Resource Delivery 的國際標準。
2. 完成碼率失真最佳化串流傳輸系統的開發，並將其整合到總計畫團隊為 MPEG 開發的所設計出的多媒體傳輸共通測試平台上。
3. 完成小波轉換視訊壓縮法的可進行多次碼率失真最佳化切割的流量控制機制設計。這部份的設計特別適合異質性點對點(p2p)的應用，如從桌上電腦傳到 PDA 再傳到手機的串流傳輸。或者是行動裝置的省電機制。利如手機首先收到適合較大的內螢幕播放的視訊位元流，然後在需要省電時，可以從這個位元流中以碼率失真最佳化的原則抽出較小畫面的子位元流在外螢幕播放。
4. 設計了一套能自動隨視訊內容重要性

和即時封包掉失率來調整保護強度的 FEC 保護強度的機制。這個設計可以提供 IP 網路封包串流傳輸的應用更好的效果。

5. 在人才培育方面本計畫三年來共有一個博士生，七個碩士生參與。參與過的學生畢業後分別加入聯詠、聯發科、IBM、明碁、工研院等單位工作。

在論文發表方面，本計畫已發表一篇國際期刊（IEEE Trans. On Multimedia）、一篇國內期刊（CCL Technical Journal）、另外投稿兩篇國際期刊論文正在審核中。除此尚有四篇國際研討會論文。

六、與計畫相關之已發表文獻

[1] *ISO/IEC 21000-12:2004(E), Information Technology – Multimedia Framework (MPEG-21) – Part 12: Test Bed for MPEG-21 Resource Delivery*, 2004.

[2] C.-W. Tang, C.-H. Chen, Y.-H. Yu, and C.-J. Tsai, "Visual Sensitivity-Guided Bit Allocation for Video Coding," *IEEE Trans. Multimedia*, Vol. 8, No. 1, 2005, pp. 11-18.

[3] C.-P. Ho, C.-J. Tsai, and Y.-F. Hsu, "MPEG-21 Digital Item Adaptation Architecture for Fully Scalable Video Streaming," *CCL Technical Journal*, Vol. 109, Sep. 2004, pp. 55- 64.

[4] C.-P. Ho , W.-C. Chang , K.-C. Lee , C.-J. Tsai, "Rate-Distortion Optimized Video Streaming with Smooth Quality Constraint," *Proc. IEEE Int. Symposium on Circuit and System*, Canada, May 2005, pp.3271 - 3274.

[5] Y.-H. Yu and C.-J. Tsai, "A Model-based Rate Allocation Mechanism for Wavelet-based Embedded Image and Video Coding," *Proc. IEEE Int. Symposium on Circuit and System*, Canada, May 2005, pp.6066 - 6069.

[6] C.-W. Tang, C.-H. Chen, Y.-H. Yu, and C.-J. Tsai, "A Novel Visual Distortion Sensitivity Analysis for Video Encoder Bit Allocation," *Proc. IEEE Int. Conference on Image Processing*, Vol. 5, Singapore, October 2004, pp. 3225-3228.

[7] C.-J. Tsai, C.-W. Tang, C.-H. Chen, and Y.-H. Yu, "Adaptive Rate-Distortion Optimization using Perceptual Hints," *Proc. IEEE Int. Conference on Multimedia and Expo*, Vol. 1, Taipei, Taiwan, June 2004, pp. 667-670.

研究成果一

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

1. Introduction

Data networks for multimedia communications are growing fast nowadays. The network technologies vary from dial-up connections, broadband cable/ADSL networks, to wireless/mobile networks. In addition, the terminal devices for multimedia distribution systems are different in many aspects, including storage capacity, computational power, and screen sizes, etc. 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]. For layered coding approaches, the content quality is optimized for certain bitrate conditions. Adaptation of such content to a new target bitrate after encoding process usually results in sub-optimal 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 belongs to this category [3][4][5]. Although it is not necessary for an embedded wavelet video bitstream to assume a layered structure, video parameters such as resolution, frame rate, and bitrate of the bitstream can still be dynamically adapted with fine granularity after the encoding procedure. If the R-D tradeoff information is also embedded in the bitstream, the dynamic bitstream adaptation process can produce an R-D optimal bitstream at run-time for the target application. One of the advantages of embedded wavelet bitstreams is for multiple adaptation applications. For example, in Fig. 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 send it to another device with tighter channel and device constraints (the PDA). For the other distribution chain in Fig. 1, the cellular phone received a good quality bitstream 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 decode only part of the received bitstream and display a smaller video.

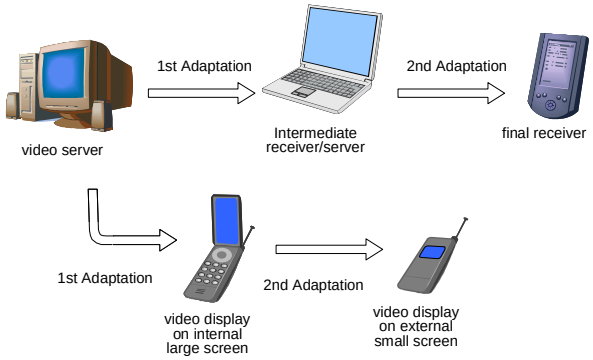


Fig. 1. Two examples of multiple adaptation applications where the same video content is adapted several times down the distribution chains

Although multiple adaptations can be achieved using layer-structured embedded bitstreams as well, it is not desirable because each layer of such bitstreams is pre-optimized for certain target bitrate by the encoder. For the same example in Fig. 1, in order to adapt the bitstream for the PDA, the notebook extracts the embedded layers which do not exceed the channel and device constraints of the PDA's. This approach is quite simple but the bitstream can not achieve the best quality possible since the run-time constraints may not meets the pre-optimized layers of 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 run-time 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 PDA's 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 are computed from real data points and encoded into the bitstream for later adaptation. Bisection search is applied at run-time 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, R-D side information is also very useful for video streaming applications. Several frameworks for wavelet based video streaming have been proposed in the literature recently. Chu and Xiong [9] introduced a packetization scheme for combined wavelet video coding and FEC for video streaming and multicasting. The packetized wavelet video coder marks the truncation points of the bitstream at the nearest packet boundaries (instead of the end of each fractional bit-plane). In the FEC-based error protection scheme, it applies Reed-Solomon (RS) coding to produce parity packets. The scheme broadcasts all source packets to one multicast group and parity packets to different multicast groups. Hence, for each client, the optimal number of layers and error protection to subscribe to can be determined by the packet loss ratio and the available channel bandwidth. However, data-interleaving is not used in this work, which makes the system less robust to burst errors. Dong and Zheng [10] proposed a content-based retransmission framework for wavelet video streaming. The compression module adopt dynamical grouping and bounded coding scheme for improving compression efficiency and removing unnecessary dependency to each coefficient subband. In the transmission module, a video packet includes one or more subbands, and a content-based retransmission is used to provide robustness against transmission errors. The content-based retransmission scheme is based on the importance of packet content which is computed by the square-sum of coefficients for each wavelet subband. Nevertheless, retransmission-based error control requires longer jitter buffer and may consume too much extra bandwidth in high error rate channels [11]. On the other hand, fixed level of FEC protection consumes considerable overhead which are wasted if there is not channel error. If the R-D side information is available to the streaming server, it can estimate the importance of subband data more accurately. Hence, it is possible to design a content-adaptive FEC protection scheme that has lower channel distortion with lower FEC overhead.

In this paper, a parameterized R-D model-based approach for multiple adaptation applications and for content-adaptive FEC protection for streaming video is proposed. For multiple adaptations, the goal is to reduce both the size of the R-D side information embedded in the bitstreams and the computational complexity of the run-time rate adaptor. For content-adaptive FEC protection applications, the goal is to improve the accuracy of importance estimation of various subband data. The organization of the paper is as

follows. Section 2 introduces some previous work of the rate adaptation scheme for embedded codecs and discusses their strengths and weaknesses. Section 3 discusses a parameterized rate-distortion model for wavelet video. The proposed multiple adaptation scheme and content adaptive FEC protection scheme 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 will be given in section 6.

2. Rate Adaptation Problem of Wavelet Video Coding

A general framework for wavelet-based embedded video coding [3][4] is shown in Fig. 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 is used to explore bitrate (quality) scalability of the embedded bitstreams. Note that, in addition to being an encoder module, the rate allocation module can be used in a video server as a standalone module with an entropy-coded embedded bitstream as the input and its subset bitstream as the output.

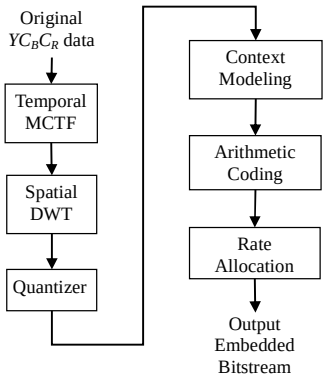


Fig. 2. General wavelet coding framework

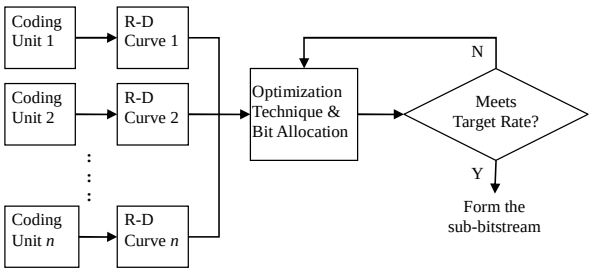


Fig. 3. Basic rate allocation scheme

For wavelet-based codecs, video data is partitioned into coding units, which could be a frame, a frequency band, or a coding block. A basic scheme of the rate allocation module is shown in Fig. 3. The function of rate allocation is to extract a smaller sub-bitstream 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 called full-dimensional scalability. Many rate control schemes have been proposed for embedded image/video codecs. In general, the rate control scheme for embedded codecs is composed of two parts. The first part is to model the R-D 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.

2.1. R-D Model Construction

Several R-D models have been proposed to establish the tradeoff between rate and distortion for each coding unit [4][8][12]. 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 bit plane coding. As the available bandwidth of

the target applications goes from low to high, more and more fractional bit planes could be included into the target bitstreams according to their significance. In other words, an embedded bitstream is composed of fractional bit planes. The more fractional bit planes the bitstream contains, the higher quality it would be. Therefore, inclusion of an additional fractional bit plane in a coding unit contributes to both increment of bits (rate) and reduction of quality loss (distortion). Recording of the rate and distortion data point of each fractional bit plane provides a precise yet discrete R-D model of the embedded bitstream [4]. By using real data points to represent the R-D model, the tradeoff between rate and distortion at each truncation point can be precisely determined. However, storing all the discrete R-D values for each fractional bit plane in each coding unit during the bit allocation procedure requires a lot of memory space. Even worse, for multiple adaptations, this R-D information must be embedded in the bitstream throughout the distribution chain. Furthermore, in order to find the best truncation point which matches the rate constraint, some search techniques (possibly time-consuming) must be applied while doing bit allocation.

Different from the discrete R-D model approach, some literatures [8][12] use close-form models to describe the R-D characteristic of the video data. This approach first applies information theory on a simplified source model and a codec model to calculate the relationship between coding rate and distortion. In the closed form R-D equation, content-dependent information is summarized in a few parameters. With the parameterized R-D model, the R-D characteristic of each coding unit will be estimated at runtime by solving the content-dependent 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.

2.2. Bit Allocation

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 unit in an optimal way. A popular approach is to use the Lagrange multiplier to transform constrained optimization problem into unconstrained optimization problem [4][8][12]. 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.

After establishing the R-D characteristic model and the optimization process using Lagrange multiplier, each optimal truncation point contains three attributes including rate, distortion, and the Lagrange multiplier value (refer to as the λ value hereafter). The next step is to form an optimal target bitstream given a rate or distortion constraint. Some literatures use iterative search method to achieve this goal [4][8][12]. Among the optimal truncation point attributes, the λ values represent the trade off 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 selecting 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 [13][14] 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 [13]. The heap structure which contains all possible truncation points is built internally during encoding process and some heap manipulations, such as "shiftdown" and "update root," are conducted according to R-D property of each truncation point. The heap manipulation operations stop when the heap tree is balanced and the root of the tree meets the target bitrate constraint. At this point, the final bitstream is composed. Another approach that uses quadtree merge-based algorithm is proposed in [14]. Similar to the heap-based proposal, this method tries to achieve fast R-D optimization by applying simple operations to manipulate the data structure during the bit allocation process. One major disadvantage of fast search algorithm with well-designed data structure is that the memory required may be extremely large in order to build the complete data structure to store all coding unit information, especially for video coding.

3. Parameterized R-D Models for Wavelet Video

The concept of rate distortion function is first published by Shannon [15]. Based upon the idea of Shannon, several literatures [16][17][18] pointed out that rate and distortion have the relationship shown in (1), where R(·) is the source rate, E(b) represents the entropy of the signal source b, and D is the distortion measured by square error.

$$R(D) \geq E(b) - \frac{1}{2}\log_2(2\pi eD) \qquad (1)$$

From (1), let $R_L(D) = E(b) - \log_2(2\pi eD)/2$, one can infer the general form of the Rate vs. Square-Error-Distortion function as shown by (2):

$$\begin{aligned} R_L(D) &= \frac{1}{2}\log_2 2^{2E(b)} - \frac{1}{2}\log_2(2\pi eD), \\ &= \frac{1}{2}\log_2 \frac{\omega}{D} \\ \omega &= \frac{2^{2E(b)}}{2\pi e}. \end{aligned} \qquad (2)$$

The parameter, ω, in (2) is related to the probability density function of the source signal. Take Gaussian distribution for example, assume the probability density function of the source is $p(x)$ with mean μ and variance σ^2, the entropy of the source signal is:

$$E(p) = \frac{1}{2}\log_2 2\pi e\sigma^2. \qquad (3)$$

Therefore,

$$\omega = \sigma^2, \text{ and } R_L(D) = \frac{1}{2}\log_2 \frac{\sigma^2}{D}. \qquad (4)$$

Note that in (4), when $^2 < D$, $R_L(D)$ becomes negative. If this is not desirable, (4) can be rewritten as follows:

$$\begin{aligned} R_L(D) &= \max(\frac{1}{2}\log_2 \frac{\sigma^2}{D}, 0) \\ &= \begin{cases} \frac{1}{2}\log_2 \frac{\sigma^2}{D}, & 0 \leq D \leq \sigma^2 \\ 0, & D \geq \sigma^2 \end{cases} \end{aligned}. \qquad (5)$$

In this section, the general rate distortion relationship is established. The rate distortion model can be extended by using different distortion measures or content probability density functions. Some literatures apply the function to embedded wavelet coder [8][14] and make a little empirical adjustment on the parameters. The revised relationship with an additional parameter, χ, is in (6):

$$R(D) = \frac{1}{2}\chi \log_2 \frac{\omega}{D}, \quad \omega = \frac{2^{2E(b)}}{2\pi e}. \qquad (6)$$

The parameter χ characterizes the exponentially decaying rate. Base on the analysis of the experimental results in [8][14], the parameter is shown to be related to the distribution of the source. The general R-D function for embedded wavelet coder with square-error measure is shown in (7):

$$R(D) = \gamma \ln \frac{\omega}{D}, \qquad (7)$$

where $\gamma = (\chi \log 2e)/2$ is the scaling factor for changing the base of the log function.

We conducted an experiment using a wavelet video codec [5] to examine the

precision of the rate distortion relationship in (7). The test sequence is STEFAN in CIF resolution. The partial results for two coding blocks are shown in Fig. 4. 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 parameter γ and ω in (7) using least-squares-error curve fitting method. Due to different local source distributions, these two coding blocks have different values of the parameters. $D(R) = 3739.1\ e^{-0.0120R}$ for coding block one and $D(R) = 19794\ e^{-0.0137R}$ for coding block 2. The experiment shows the precision and the reliability of the rate distortion function when applying to coding blocks with different characteristics.

So far, we have introduced the theoretical background of scalable rate control algorithms. However, there are still some gap between the theory and actual implementations. For example, the determination of the Lagrange multiplier value is difficult in practice, and the overall bit allocation procedure should be restructured in order to achieve computational efficiency. Solutions to these issues will be developed in the proposed scheme in the next section.

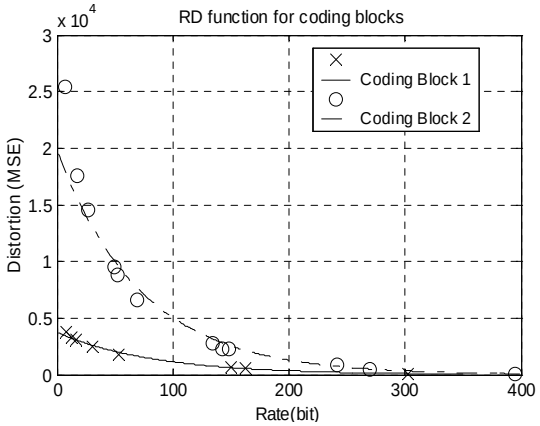


Fig. 4. R-D models for coding blocks in a wavelet video codec

4. The Proposed Multiple Adaptations and Context-Adaptive FEC schemes

In this section, we presents 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 Microsoft Research Asia (MSRA) wavelet codec [5]. A bitstream encoded using the MSRA codec is organized in the format shown in Fig. 5. A bitstream parser extracts the information for the truncated candidates from the headers. After all the required data are collected, the bitstream truncation procedure begins without entropy decoding involved. The truncation module decides the truncation point in order to meet the resolution, frame rate, and bit rate criterions. The bitstream is then composed again with new header information and truncated body bits. The new bitstream should conform to the usage scenario and can be transmitted over the network to the target receiver.

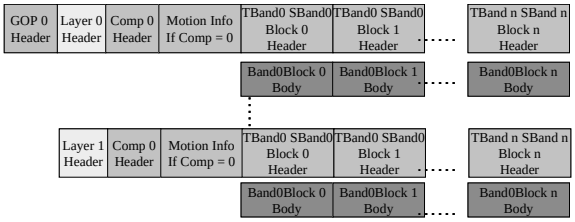


Fig. 5. MSRA Wavelet Bitstream Format

4.1. Proposed Rate Allocation Algorithm for Multiple Adaptation Applications

On a PC platform, the truncation process account for 72% of the bitstream adaptation time. The proposed framework (Fig. 6) tries to build a closed-form *R-* 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 value. Therefore, the truncated point for each coding block can be selected given the target bit rate.

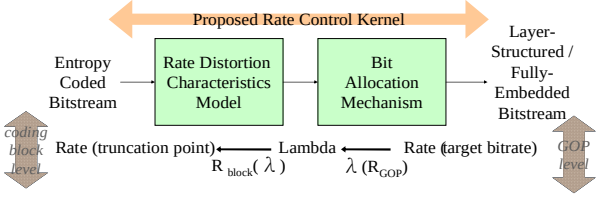


Fig. 6. The framework proposed rate control extractor

4.1.1. R-Lambda Model Analysis

The proposed *R*- model for each coding block is established by combining the parameterized R-D model mentioned in section 3 and the Lagrange multiplier optimization technique. Recall that in (7) 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 minimization of the Lagrange cost function $J(R) = D +$ R can be obtained when $dJ(R)/dR = 0$, that is,

$$\lambda = -\frac{dD(R)}{dR}. \qquad (8)$$

Taking the inverse of (7), we have $D(R) = \omega\, e^{-\frac{R}{\gamma}}$ $D(R) =$ $e^{-R/}$. Substitute $D(R)$ into (8), we obtain the relationship between the Lagrange multiplier and the bitrate:

$$\lambda = \left(\frac{1}{\gamma}\right)\omega e^{-R/\gamma}. \qquad (9)$$

As a result, the R– model in coding block level can be written as in (10), where the parameters α and β are source dependent:

$$\lambda = \alpha\, e^{\beta R}. \qquad (10)$$

For each coding block, a parameter pair of (α, β) will be estimated by curve-fitting to real *R*- data points. Fig. 7 shows an example of a coding block of the FOOTBALL sequence using the MSRA codec. Each point in the figure is the R-D point of a possible truncation point. Therefore, the R-D information of the whole coding block can be represented using simply two parameters, instead of 11 data points as in in Fig. 7.

The GOP level *R*- model can be extended from the coding block model. First, by adopting the R-D form of 錯誤! 找不到參照來源。, we have $R = \max(\frac{1}{\beta}\ln\frac{\lambda}{\alpha}, 0)$ be a nonnegative R-D model. For > 0 and < 0, the *R*- model at GOP level is derived as follows:

$$\begin{aligned} R_{GOP} &= \sum_i R_{block\ i} = \sum_i \max(\frac{1}{\beta_i}\ln\frac{\lambda}{\alpha_i},\ 0) \\ &= \sum_j \frac{1}{\beta_j}\ln\frac{\lambda}{\alpha_j}, \quad where\{j \in S \,|\, \alpha_j > \lambda\ in\ S\} \\ &= (\sum_j \frac{1}{\beta_j})\ln\lambda - (\sum_j \frac{1}{\beta_j}\ln\alpha_j). \end{aligned} \qquad (11)$$

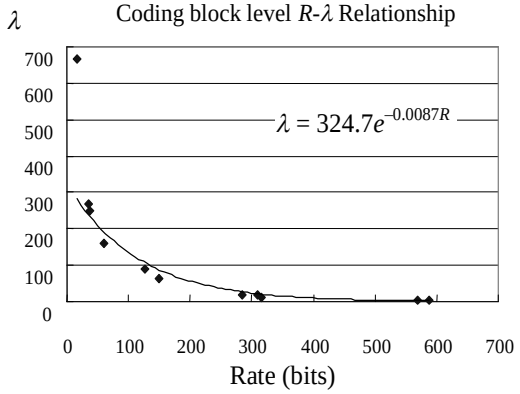


Fig. 7. Example of coding block level *R*– relationship

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 value. We define the two summation terms in (11) as follows:

$$p_{GOP} = \sum_j \frac{1}{\beta_j} \text{ and } q_{GOP} = \sum_j \frac{1}{\beta_j}\ln\alpha_j. \qquad (12)$$

In order to keep the model simple, we assume that these two summations can be modeled by polynomials as shown in (13):

$$P_{GOP} = a_1(\ln(\lambda))^{n-1} + a_2(\ln(\lambda))^{n-2} + \cdots + a_n$$

and (13)

$$q_{GOP} = b_1(\ln(\lambda))^{n-1} + b_2(\ln(\lambda))^{n-2} + \cdots + b_n.$$

Finally, the relationship of the GOP level *R*- model is established in (14):

$$\begin{aligned} R_{GOP} &= p_{GOP}\ln\lambda - q_{GOP} \\ &= \gamma_1(\ln\lambda)^n + \gamma_2(\ln\lambda)^{n-1} + \cdots + \gamma_{n+1} \end{aligned}. \qquad (14)$$

The graph in Fig. 8 illustrates the accuracy of the proposed *R*- model in the GOP level. The order of the function is determined empirically. In general, a cubic

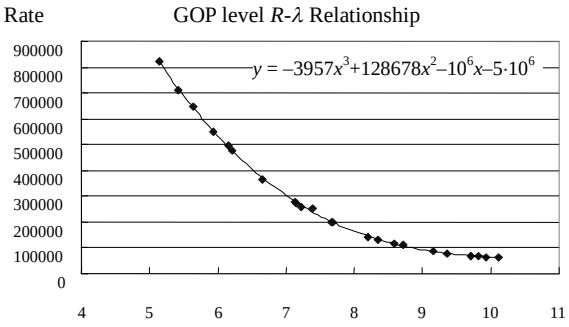


Fig. 8. Example of GOP level *R*- model

function can be used to fit the data points well for a wide range of rate.

4.1.2. Embedded Rate-Distortion Information Generation

In order to allow for multiple adaptation applications, we must embedded the R-D information into the bitstream so that a terminal receive 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- 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 pre-optimized 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 through the proposed algorithm.

The coding block level model (10) 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 parameter α and β can be calculated as follows. First, (10) can be rewritten as $\ln\lambda = \ln\alpha + \beta \cdot R$. Therefore, for $n > 2$ we have an over-determined system of equations 錯誤! 找不到參照來源。,

$$\begin{pmatrix} \ln\lambda_0 \\ \ln\lambda_1 \\ \vdots \\ \ln\lambda_{n-1} \end{pmatrix} = \begin{pmatrix} 1 & R_0 \\ 1 & R_1 \\ 1 & \vdots \\ 1 & R_{n-1} \end{pmatrix} \begin{pmatrix} \ln\alpha \\ \beta \end{pmatrix}. \quad (15)$$

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*- model (see (14)) is adaptively built by the least-squares curve fitting method. For certain GOP, assume that

$$A = \begin{pmatrix} (\ln\lambda_1)^n & (\ln\lambda_1)^{n-1} \cdots & 1 \\ (\ln\lambda_2)^n & (\ln\lambda_2)^{n-1} \cdots & 1 \\ \vdots & \vdots & \vdots \end{pmatrix},$$
$$Y = \begin{pmatrix} R_{GOP1} \\ R_{GOP2} \\ \vdots \end{pmatrix}, \text{ and } X = \begin{pmatrix} \gamma_1 \\ \gamma_2 \\ \vdots \\ \gamma_{n+1} \end{pmatrix}. \quad (16)$$

The parameters $\gamma_1, \gamma_2, \ldots, \gamma_3$ can be solved by computing the pseudo inverse $X = (A^TA)^{-1}A^TY$. As the whole GOP level *R*- model is established, the value can be solved using closed form solutions for $n < 5$ (typical n is 3).

The overall proposed algorithm which adopts the *R*- model is illustrated in Fig. 9. In the bit allocation mechanism, the *R*- model is used to search the lambda value in the GOP level, and in the rate distortion optimization procedure, the *R*- function is used to represent the rate distortion properties in the coding block level.

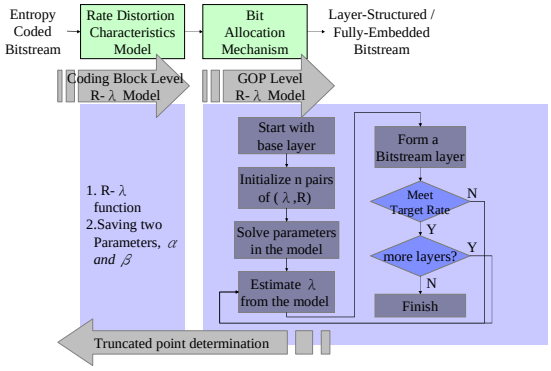


Fig. 9. Overall framework of the proposed rate adaptation mechanism

The algorithm used to embed R-D information into a MSRA encoded bitstream are 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 (γ_1, γ_2, …, γ_3).
   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 another quality layers.
2. Represent R-D property of a coding block:
   In procedure d), a bitstream layer is formed given a Lagrange multiplier value. The truncation point of each coding block is determined at the fractional bitplane pass with the nearest Lagrange multiplier value. To achieve the typical coding block level rate allocation, the Lagrange multiplier value of each fractional bitplane pass in all coding blocks should be stored during tier 1 of entropy coding. In order to reduce the memory usage of the information and distribute the rate among all coding blocks based on information theory, the coding block level R-lambda model is applied to describe the property of each coding block. Therefore, only the parameters α and β should be stored for a single coding block, and the coding block level rate allocation can be easily done by adopting the inverse R-lambda model with a given Lagrange multiplier. In the proposed method, the truncation point would be the fractional bitplane pass with the nearest rate.

4.1.3. Rate Adaptation Procedure

Once the bitstream is formed, run-time adaptation to a target bitrate becomes a question of searching for a value that marks all the truncation points to form a target bitstream that following the rate constraint. For discrete R-D information used by the original MSRA codec, bisection search is used for determining the value. The search process starts from the initial maximum and minimum value estimates. By half-eliminating the search range at each iterative step, the search results converge and the value which meets the target bitrate is obtained at the end. Fig. 10 shows an example of such process. In Fig. 10, the MSRA bitstream extractor is used to adapt the full FOOTBALL bitstream to two quality layers. The first layer is QCIF resolution, 7.5 frames per second, bitrate 128 kbps. The second layer is QCIF resolution, 15 frames per second, bitrate 192 kbps.

For the proposed algorithm, the 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. These guesses are marked with arrows in Fig. 11. Then the model is fitted to these data points. The closed-form model is then solved to determine the value. If this value results in a bitstream that meets the target rate, the process stops, otherwise, the process will be repeated with the new (*R*,) pair replacing the first data point. Usually, the estimation process can meet the target bitrate in two steps.

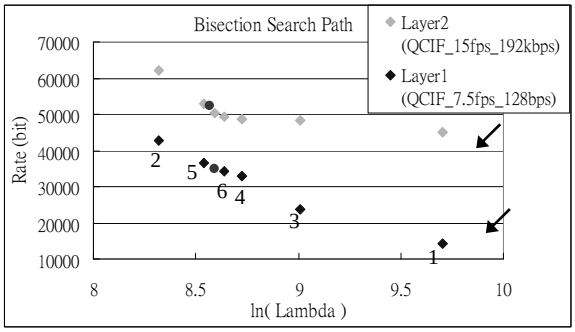


Fig. 10. Example search path for the MSRA method (numbers are the iteration number)

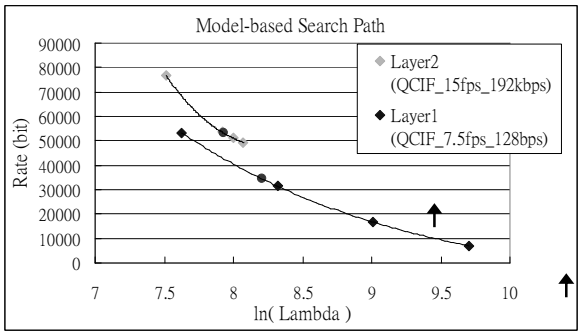


Fig. 11. Example search path for the proposed method (both searches stop in one step)

4.2. Proposed Context-Adaptive Streaming Framework

Another proposed application for the *R*-model is for content-adaptive FEC protection of the embedded bitstream. For video streaming FEC protection video packets is sometimes preferred over retransmission since the round trip time of a congested network may be too long for effective retransmission of missing packets. However, FEC protection of video data imposes noticeable overhead that unequal error protection should be used to achieve the best rate-distortion tradeoff.

For example, 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 (for example, a symbol is composed of 8 bits of bitstream data). The number of parity symbols is 2*s*, 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 2*s* errors per codeword. Therefore, in order to perform content-adaptive FEC protection, the degree of protection level *s* should be based on the importance of the video data.

In the generic wavelet coding framework (Fig. 2), video frames are first temporally decomposed into several levels of low- and high-frequency subbands, and then 2D spatial decomposition is applied to each temporal subband for spatial subband decomposition of the signals. After temporal and spatial subband decomposition, all the spatio-temporal subband data are encoded by the 3D-ESCOT entropy coder. Typically, low frequency band data are more important than high frequency band data. In addition, the importance of the coefficients within a coding block in a particular subband can be ranked based on the R-D model 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 proposed content-adaptive protection mechanism determines the level of FEC protection for different subband coefficients based on their subband index and the R-D model of their coding block. To compute the level of FEC protection, we first compute the subband factor W_i based on the subband indices as in (17):

$$W_i = \exp\left[(-1)\cdot\left(\frac{(T-\omega_i)\cdot C_1}{\mathrm{T}} + \frac{1}{(S-\tau_i)}\right)\right], \quad (17)$$

where T is the maximum temporal level index, S is the maximum spatial subband index, and C1 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 *i* is divided into *m* codeword. The protection level *s* of different portions of coding block *i* is computed by

$$s_{i,x} = \left\lfloor \left(\frac{\alpha_i \cdot \exp\left(C \cdot \beta_i \cdot \sum_{j=0}^{x} R_{i,j} \right)}{\omega} + W_i \right) \cdot n_{pl} \right\rfloor + o, \quad (18)$$

$$o = \begin{cases} 1 & \text{if } s_{i,x} \text{ is even} \\ 0 & \text{if } s_{i,x} \text{ is odd} \end{cases},$$

where x = 0, 1,…, m-1, the parameters αi and βi are the close-form R- model (10) parameters for the coding block i, Ri,x is the length of the xth RS codeword in coding block i, C is a weighting factor, npl denotes the estimated number of packet losses per second, and ω is a scale factor determined empirically. Equation (18) is designed so that si,0 ≥ si,1 ≥…≥ si,m-1, that is, the level of protection decreases following fractional bitplane coding pass order. Note that the operation ⌊·⌋ stands for "taking the largest integer less then."

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. The coding parameters used in the experiments are as follows. The GOP size is 64 frames, and the frame rate is 30 fps. The parameter n in the GOP level model is set to 3, and the bitrate error threshold TR is set to 3% of the target bitrate.

5.1. Computational Cost Reduction for Bitstream Adaptation

The number of iterations required before the solution converges for the proposed method and the bisection search used in the MSRA codec is shown in TABLE I. The average computation cost saving is about 47% when the resolution and the frame rate settings for each layer are all different. When the number of layers for each resolution and frame rate setting increases, the search procedure can converges even faster by taking advantage of the *R-* model from the previous layer. According to our experiments, the saving ratio is about 60% when the layer number is 5, and up to 80% when the layer number is 12 (Fig. 12).

TABLE I. Number of iterations comparison for search

Sequence	*MSRA Bisection*	*R- λ Model*	*Saving Ratio*
Mobile	9.67	5.30	45.17 %
Foreman	10,68	4.55	57.41 %
Football	7.84	4.70	40.05 %
Average	**9.40**	**4.85**	**47.54 %**

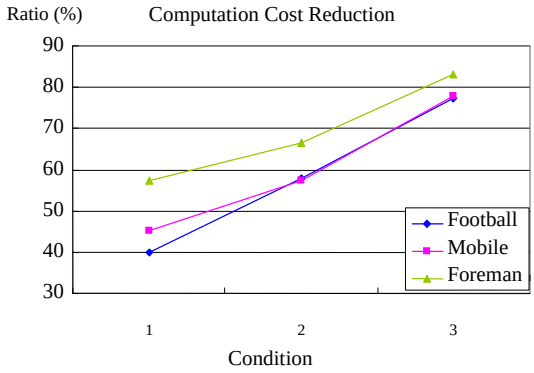
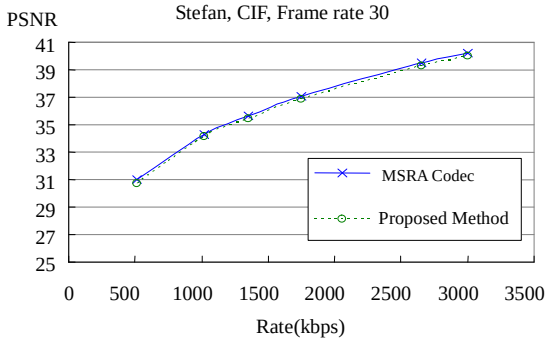


Fig. 12. Computation reduction ratio of the proposed method

Since the proposed mechanism allocates rate for each coding block differently from that of the MSRA codec's, the rate distribution (and quality) in a GOP is different from that of the MSRA codec's. The coding efficiency is shown in Fig. 13, Fig. 14, and Fig. 15. The test sequences are STEFAN, FOOTBALL, and FOREMAN in CIF resolution and are truncated at frame rate 30 and 15. The figures show that the proposed rate adaptation mechanism achieves similar PSNR performance in comparison with that of the MSRA codec's at any rates. The average PSNR degradation is only 0.25dB.



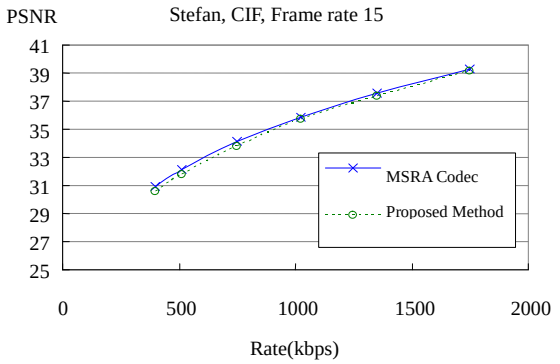


Fig. 13. PSNR performance comparison of STEFAN

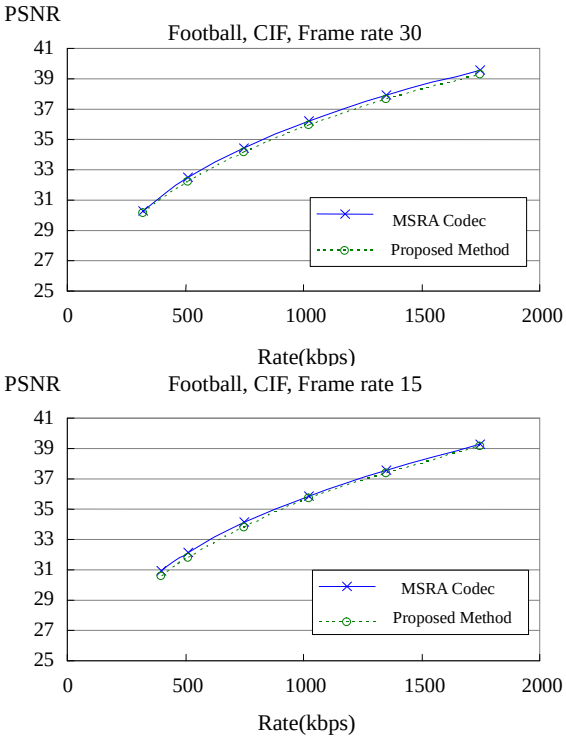
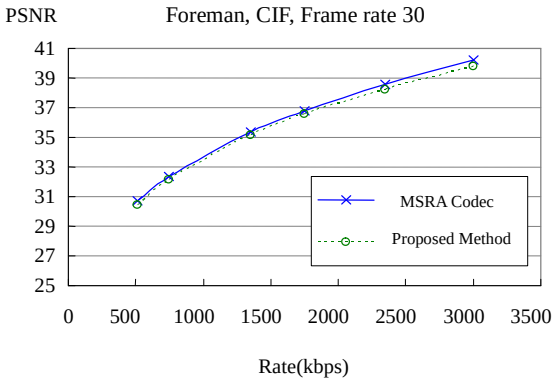


Fig. 14. PSNR performance comparison of FOOTBALL



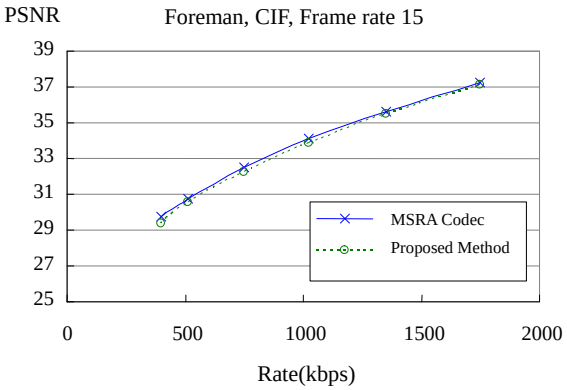


Fig. 15. PSNR performance comparison of FOREMAN

5.2. Side Information Saving for Multiple Adaptation Scheme

The experimental result in the Table 2 shows the saving ratio in different resolutions and framerates for different sequences in a multiple-adaptation scenario. The average saving ratio of the side information is about 50%, and the side information percentage in the bitstream is reduced from 3.5% to 1.7%.

TABLE II. Side information saving ratio

Sequence Name	*Resolution / frame rate / bit rate (kbps)*	*Side information bits (% in bitstream)*		*Saving Ratio*
		MSRA	*Proposed Method*	
Mobile	CIF / 30 / 768	266,864 (3.39%)	128,475 (1.63%)	51.86%
	CIF / 30 / 512	165,691 (3.16%)	78,030 (1.49%)	52.91%
	CIF / 30 / 384	112,793 (2.87%)	52,924 (1.35%)	53.08%
	CIF / 15 / 256	73,981 (2.82%)	32,718 (1.26%)	55.78%
Foreman	CIF / 30 / 768	332,417 (4.23%)	166,249 (2,12%)	49.99%
	CIF / 30 / 512	234,343 (4.47 %)	107,864 (2.06%)	53.97%
	CIF / 30 / 256	109,067 (4.17%)	49,737 (1.90%)	54.40%
	CIF / 15 / 192	86,234 (4.40 %)	41,488 (2.12%)	51.89%
Football	CIF / 30 / 1380	347,746 (3.06%)	186,822 (1.64%)	46.28%
	CIF / 30 / 1024	282,381 (3.11%)	149,396 (1.64%)	47.09%
	CIF / 30 / 768	213,047 (3.13%)	110,072 (1.61%)	48.33%
	CIF / 15 / 512	141,205 (3.11%)	76,603 (1.69%)	45.75%
Average		3.5%	1.7%	50.94%

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 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 [19][20][21], 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. The close-form R- model (10) is compared against a coarse discrete R- model where the first in the coding block is used to determine the importance of all the coefficients in the block [22]. The PSNR of the luma channel of the reconstructed video sequences are shown in Fig. 15 and Fig. 16. In either case, the maximal packet loss protection level can only recover up to 4% packet losses on average.

As one can see from the figures, the close-form R-D model has higher performance than the coarse R-D model, especially in the low bitrate cases. At low bitrate, accurate R-D models are crucial for both rate control and FEC protection decisions for video servers.

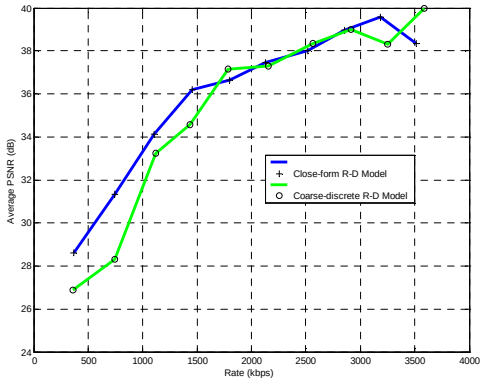


Fig. 16. Content-adaptive FEC test for the STEFAN sequence (5% losses)

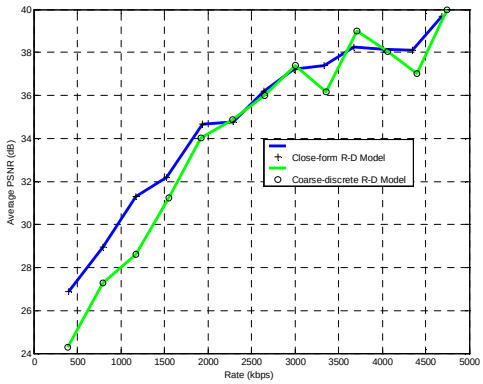


Fig. 17. Content-adaptive FEC test for the MOBILE sequence (5% losses)

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 use a closed-form *R-* model to reduce the size of R-D side information embedded in the coded bitstream while maintaining the accuracy of the rate-distortion information of the video data. In addition, the proposed technique can reduces the computational complexity of wavelet video adaptation by 50%. Although the existing model achieves good performance, there are till 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 efforts on scalable motion vector coding. Similar ideas can be applied to the construction of an R-D model for motion vector bits to increase the performance further.

7. 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, Mar. 2003.

[2] W. Li, "Overview of Fine Granularity Scalability in MPEG-4 Video Standard," *IEEE Trans. on Circuits and Systems for Video Technology*, Vol. 11, No. 3, Mar. 2001, pp.301-317.

[3] S.J. Choi and J.W. Woods, "Motion-Compensated 3-D Subband Coding of Video," *IEEE Transactions on Image Processing*, vol. 8, Feb. 1999, pp. 155-167.

[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, Special Issue on Wavelet App.*, vol. 10, 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. Pearlman, "A New, Fast and Efficient Image Codec Based on Set Partitioning in Hierarchical Trees," *IEEE Trans. on Circuit and System Video Technology*, vol. 6, June 1996, pp. 243-250.

[7] D. Taubman, "High performance scalable image compression with EBCOT," *IEEE Trans. on Image Processing*, vol. 9, July 2000, pp. 1158-1170.

[8] Po-Yuen Cheng, Jin Li, and C.-C. Jay Kuo, "Rate Control for Embedded Wavelet Video Coder," *IEEE Trans. on Circuit and System Video Technology*, vol. 7, NO. 4, Aug. 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: Special Issue on Multimedia Systems*, pp. 66-80, Jan. 2003.

[10] J. Dong and Y. F. Zheng, "Content-based Retransmission for 3-D Wavelet Video Streaming on the Internet," in *Proceedings IEEE Int. Conf. on Information Technology, Coding and Computing*, pp. 452-457, April 2002.

[11] W. Tan and A. Zakhor, "Real-time Internet Video Using Error Resilient Scalable Compression and TCP-friendly Transport Protocol," *IEEE Transactions on Multimedia*, 1(2):172-186, May 1999.

[12] A. Aminlou and O. Fatemi, "Very Fast Bit Allocation Algorithm, Based on Simplified R-D Curve Modeling," *Proc. of 10th IEEE International Conferences on Electronics, Circuits, and Systems*, Dec. 2003, pp. 112-115.

[13] Wei Yu, "Integrated Rate Control And Entropy Coding For JPEG 2000," *IEEE Proc. of the Data Compression Conference*, 2004.

[14] Jin Li and C.-C. Jay Kuo, "Embedded Wavelet Packet Image Coder With Fast Rate- Distortion Optimized Decomposition," *Proc. SPIE: Visual Communications and Image Processing'97*, Vol. 3024, 1997, pp. 1077-1088.

[15] C. E. Shannon, "A Mathematical Theory of Communication," *Bell Syst. Tech. J.*, vol.27, 1948, pp.379-423 and 623-656.

[16] T. Berger, *Rate Distortion Theory*, Englewood Cliffs, NJ: Prentice Hall, 1971.

[17] A. N. Netravali and B. G. Haskell, *Digital Pictures: Representation and Compression*, New York, NY: Plenum, 1988.

[18] Hsueh-Ming Hang and Jiann-Jone Chen, "Source Model for Transform Video Coder And Its Application – Part I: Fundamental Theory," *IEEE Trans. on Circuit and System for Video Technology*, vol. 7, No. 2, April. 1997.

[19] J. M. Boyce and R. D. Gaglianello, "Packet loss effects on MPEG video sent over the public Internet," *Proc. of ACM Multimedia*, pp. 181-190, September 1998.

[20] K. Lai, M. Roussopoulos, D. Tang, X. Zhao, and M. Baker, "Experiences with a Mobile Testbed," *Proc. of the Second Intern. Conf. on Worldwide Computing and its Applications (WWCA'98)*, March 1998.

[21] R. Rosa, C. -P. Angel, D. -F. Manuel, O. Luis, and G. Antonio, "On the Traffic Disruption Time and Packet Lost Rate during the Handover Mechanisms in Wireless Networks," *Proceedings AINA*, 2004, pp.351-354.

[22] C.-P. Ho and C.-J. Tsai, "Content-Adaptive Packetization and Streaming of Wavelet Video over IP Networks," Submitted to Journal of Image and Video Processing, Aug. 2006.

研究成果二：

Content-Adaptive Packetization and Streaming of Wavelet Video over IP Networks

1. Introduction

There is a growing demand for video transmission over heterogeneous networks for communication and entertainment applications. Scalable video coding (SVC) techniques are often proposed for such systems since, ideally, a video sequence can be encoded once and adapted on-the-fly to different frame rate, bitrate, and resolution for different applications. Although scalable video is an interesting concept, it takes complete end-to-end system design to show the advantage of SVC over single-layer coding techniques. With single-layer coding, techniques like bitstream switching and simulcasting can be used to achieve video adaptations. However, it is easier to achieve good rate versus source-and-channel distortion tradeoff with scalable coding techniques.

The mainstream video compression techniques are based on hybrid motion compensated transform coding approach, where the transform algorithms are typically either Discrete Cosine Transform (DCT) or 3-D wavelet transform [1]. So far, DCT-based SVC approaches have demonstrated better coding efficiency than wavelet-based SVC techniques [2], especially for low bitrate applications. However, a wavelet-based SVC framework can provide fine-granularity bitrate (i.e. SNR) scalability with less system complexity than that of an FGS-based DCT framework. In addition, many on-going efforts show that wavelet-based SVC approaches still have room for improvement [3]. Therefore, in this paper, wavelet-based SVC is used as the core codec for the development of a scalable video streaming framework.

The most challenging problem for scalable video streaming over IP networks is about how to optimally adapt source data rate and degree of packet loss protection to real-time network conditions. Video packet packetization and scheduling algorithms are mostly responsible for mitigating the effects of bandwidth variation and packet losses in the network. The packetization and scheduling algorithms are mainly based on resource-versus-distortion optimization [5][6][7][8], where resource can be available computation power, rate, delay, etc. A general resource allocation treatment for streaming systems is presented in [6]. Some researches tries to apply the rate-distortion optimization (RDO) principle [4] of source coding theories to video streaming over lossy networks [5]. For a streaming system, the distortion is a result from both source coding and channel losses. A key issue in an RDO-based streaming system is that the distortion due to packet losses is much more difficult to quantify than the distortion due to lossy source coding.

Several frameworks for 3-D wavelet based video streaming system have been proposed in the literature recently. Chu and Xiong [9] introduced a combined packetized wavelet video coding and FEC approach for video streaming and multicast. The packetized wavelet video coder marks the truncation points of the bitstream at the nearest packet boundaries (instead of the end of each fractional bit-plane). In the FEC-based error protection scheme, it applies Reed-Solomon (RS) coding to produce parity packets. And then the scheme broadcast all source packets to one multicast group and parity packets to different multicast groups. Hence, for each client, the optimal number of layers and error protection to subscribe to can be determined by the packet loss ratio and the available channel bandwidth. However, data-interleaving is not used in this work, which makes the system less robust to burst errors. Dong and Zheng [10] proposed a content-based retransmission framework for wavelet video streaming. The compression module adopt dynamical grouping and bounded coding scheme for improving

compression efficiency and removing unnecessary dependency to each coefficient subband. In the transmission module, a video packet includes one or more subbands, and a content-based retransmission is used to provide robustness against transmission errors. The content-based retransmission scheme is based on the importance of packet content which is computed by the square-sum of coefficients for each wavelet subband. Later, Zhao et al. [11] incorporated an error concealment scheme into this content-based retransmission framework to increase its error resilience capability. Nevertheless, retransmission-based error control requires longer jitter buffer and may consume too much extra bandwidth in high error rate channels [12][13].

Chou and Miao [5] developed a framework for RDO streaming of packetized media. The RDO framework is flexible to extend the optimizing packet transmission scheduling to a wide range of receiver/sender/proxy driven streaming systems [14]. However, the scheme maps (probability of) packet losses into rate increment of redundant packet forward transmission (ARQ can be avoided in this approach). However, although redundant packet transmission makes the RDO system simpler for analysis, it is not cost-effective for practical systems. R-D performance can be greatly improved if FEC is used instead. Zhu et al. [7] proposed a congestion-distortion optimized scheme. Zhai et al. [8] presented an integrated joint source-channel coding framework for video streaming. Wang et al. [15] proposed a cost-distortion optimization framework. Chang et al. also proposed a sender-based [16] and a receiver-based [17] RDO frameworks for 3-D wavelet video streaming, which basically follow the framework introduced by Chou and Miao. The proposed system uses source rate-distortion profiles to optimize for playout latency and bandwidth allocation among a group of data packets in a way that minimizes distortion in the reconstructed frames.

There are many error control schemes for video streaming, including Forward Error Correction (FEC) [18][19][20][21], Unequal Error Protection (UEP) [22][23][24], and Automatic Retransmission reQuest (ARQ) [25]. Until recently, error control schemes for streaming systems are designed independently to rate control schemes. Joint design of error and rate control is important to a variable bandwidth lossy network. For example, when the channel bandwidth increases during runtime, should more bits be allocated to sending extra (enhancement) source data, or to increase the level of protection of crucial (a.k.a. base-layer) source data? Based on the RDO principle, one should pick whichever approach that reduces more distortion. However, this is not trivial since distortions from channel losses are non-deterministic. Another issue is that not all source data bits carries equal amount of information (i.e. entropy). Although some of the error control techniques tries to put different degree of protection based on the degree of importance of the content, unequal error protection is done coarsely since the error-control scheme are based on either single-layer video coding model or coarse-granularity layered scalable video coding mode.

In this paper, a content-adaptive packetization scheme for wavelet-based streaming video is proposed. The mechanism is based on detail analysis of the mainstream wavelet-based video codec [4]. Due to its fine-granularity SNR scalability feature, the proposed packetization scheme can apply various degrees of Reed-Solomon (RS) codes on interleaved video subband data so that the streaming video is very robust over IP networks. In addition, the paper proposes to map the distortion caused by packet loss to distortion caused by source data rate reduction due to extra FEC protection (for error-free transmission). Since measuring operational video distortion from packet loss is very difficult while measuring source coding distortion is much simpler, the proposed mechanism can be applied to practical systems. In summary, the main features of the proposed system are

highlighted as follows:

- The streaming algorithm searches along the R-D curve for an optimal operating point between the scalable source coding rate and the FEC protection level.
- The FEC protection level is also influenced by run-time packet loss rate feedback from the client. Therefore, it is adaptive to both the video content entropy and the runtime packet loss rate.
- The rate-distortion tradeoff of the system takes into account both distortion due to source data rate reduction and distortion due to packet losses (predicted by FEC protection bits required for error-free transmission).

The rest of this paper is organized as follows. Section 2 presents a detail analysis on the wavelet compressed video bitstreams and its characteristics for content-adaptive protection. The detail of the proposed packetization scheme and streaming framework is described in section 3. Some experimental results of the proposed system are shown in section 4. Finally, some conclusions and discussions are given in section 5.

2. Wavelet Video Bitstream Analysis

For streaming applications, the quality of video is affected by packet loss distortion. In addition, one of the most difficult problems for RDO streaming is about how to measure packet loss distortion. In practice, distortion due to packet loss depends heavily on the source coding method. In this section, the wavelet video coding schemes presented in [4][4] are investigated in detail. In particular, some experiments are conducted to exhibit the effect of packet losses of different wavelet subband data on the reconstructed video quality.

The block-diagram of a wavelet-based video coding system is shown in Fig. 1. In a T+2D wavelet coder, an input video sequence will be temporally decomposed first using motion compensated temporal filtering (MCTF) [1]. The output of MCTF is then further decomposed by a 2-D spatial wavelet transform on a frame-by-frame basis. For example, a two-level temporal decomposition has three temporal subbands, namely, $P(H_t, YUV)$, $P(LH_t, YUV)$, and $P(LL_t, YUV)$. When a group of pictures (GOP) size is 8, a typical structure of the T+2D wavelet coder has 4 $P(H_t, YUV)$ frames, two $P(LH_t, YUV)$ frames, and two $P(LL_t, YUV)$ frames. In each frame, it consists of one luminance component (Y) and two chrominance components (U and V). After temporal and spatial subband transforms, the coefficients of different subbands are logically segmented into coding blocks, based on the structure of Fig. 19, and each coding block is independently coded by an entropy coder. For instance, a coding block size in Fig. 19 has block depth 2 (i.e. two frames), block height 36 (= $288/2^3$), and block width 44 (= $352/2^3$). Common entropy coding techniques for wavelet video are 3D Embedded Subband Coding with Optimized Truncation (3D-ESCOT) [4] and 3D Set Partitioning in Hierarchical Trees (3D-SPIHT) [28]. The 3D-ESCOT algorithm has higher compression efficiency and better scalability than the 3D-SPIHT algorithm. Therefore, the proposed scheme is based on 3D-ESCOT coding technique.

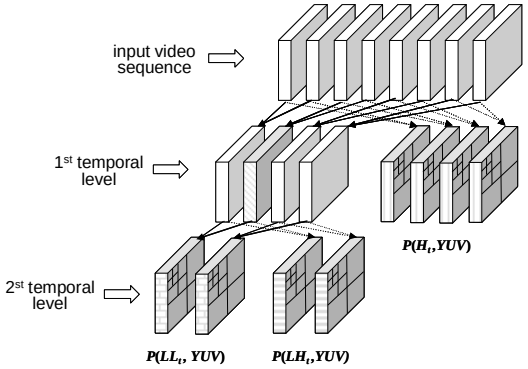


Fig. 18. Wavelet Video Coding Block Diagram.

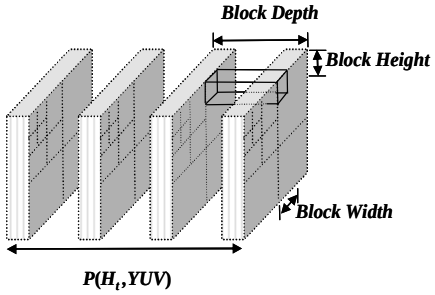


Fig. 19. Examples of Coding Block in Wavelet Video Coding.

During the 3D-ESCOT entropy coding process, the entropy coder (fractional bit-plane coding and context-based arithmetic coding) operates one coding block at a time, and each coding block consists of N total bitplanes, where N is the number of bits in the most significant coefficients. Three encoding operations of the context-based arithmetic coding (zero coding, sign coding, and magnitude refinement) are used to characterize the significance of coefficients in a bit-plane. Following the 3D context modeling, fractional bit-plane coding ensures that the bitstream is arranged with fine granularity of SNR scalability for each coding block. The fractional bit-plane coding procedure consists of three distinct passes which is significant propagation pass, magnitude Refinement pass, and normalization pass. Since the first bitplane of coding block can only process with a normalization pass, a coding block contains 3N–2 coding passes. After the entropy coding, candidate truncation point of a coding block is associate with rate distortion slopes (R-D slope). For truncating the bitstream to an optimal truncation point, those points not on the convex hull are eliminated, and the R-D slopes are λ0, λ1, …, λ(3N-2), where | λ0 | > | λ1 | >…> | λ(3N-2) |. All coding blocks have a similar R-D curve as the example shown in Fig. 20, and the top coding passes contain the most important video data. Therefore, the higher level of protection is required as in the top bitplane coding passes.

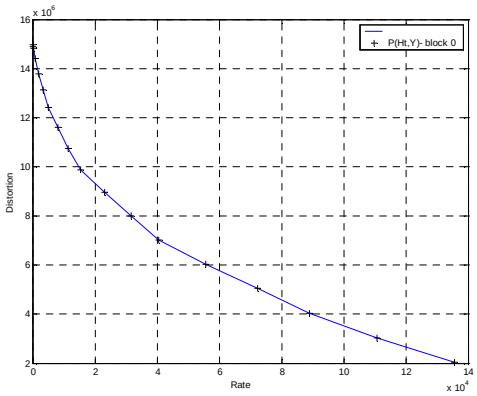
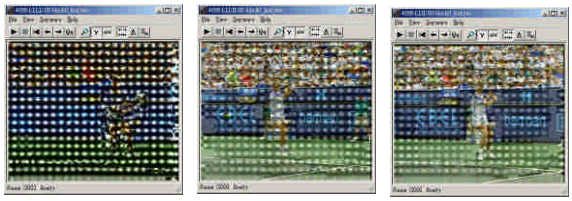


Fig. 20. The R-D curve of coding block 0 of subband $P(H_t, Y)$ of STEFAN.

In order to obtain better perception into the significance of different bitstream segments across different temporal subbands, some experiments are conducted. For example, using a four-level MCTF temporal decomposition, a group of frames are temporally decomposed into the LLLL, LLLH, LLH, LH, and H subbands. Besides, each temporal subband may comprise many spatio-temporal subbands generated by spatial decomposition. As shown in Fig. 21, for an encoded video with four-level temporal transform and three-level spatial decomposition, each temporal subband (TSB) is split into nineteen spatial subbands (SSB) from 0 to 18. The distortion impact of the first coding block within a higher spatio-temporal subband (e.g. (b), (c), (d)) is indeed more sensitive than that of the last coding block within a lower spatio-temporal subband (e.g. (e)).

In practice, given an estimated packet lost rate, we want to apply different amount of error protection for different portions of coding block based on their importance. Therefore, further analyses of wavelet subband ‘rate’ versus ‘channel-distortion’ analysis are conducted as follows. Since the size range of coding blocks is various (see Fig. 22), it is not suitable to be a data interleaving unit. A coding block should be split into several smaller units for performing interleaving. Within each coding block, the first coding pass is usually a small size (see Fig. 23) and has the highest importance value (see Fig. 24 and Fig. 25). For evaluation the

performance degradation, 10% injured bits are placed in the different portion of coding blocks. When the injured bits locate at the beginning of coding blocks, it may cause a big perceived degradation of video quality. Hence, the error protection for different portions of coding block should be setting up a different strategy.



(a) P($LLLL_t$,Y) (b) P($LLLH_t$,Y) (c) P(LLH_t,Y)

(d) P(LH_t,Y) (e) P($LLLL_t$,Y) (f) P(H_t,Y)

Fig. 21. Reconstructed video when a chunk of TSB data is lost. The loss occurs in coding block 0 of SSB 0 for the TSB in (a)-(d), and coding block 0 of SSB 18 for the TSB in (e)-(f).

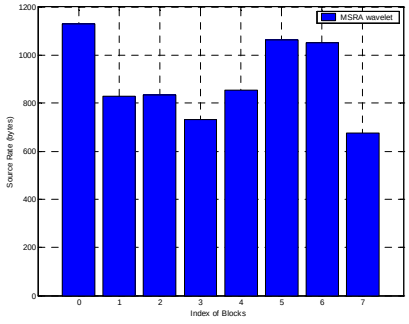


Fig. 22. Source data rate in SSB 0 of subband $P(H_t, Y)$ of STEFAN.

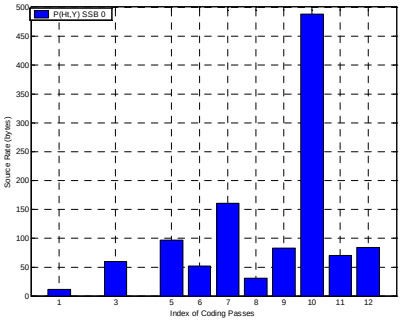


Fig. 23. Source rate of coding passes on the convex hull in block 0 of STEFAN.

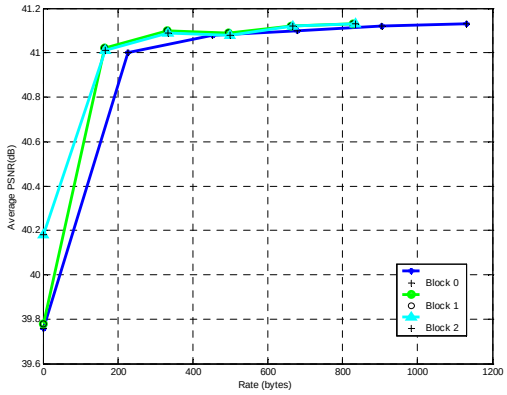


Fig. 24. PSNR with coding block loss in SSB 0 of the TSB $P(H_t, Y)$ of STEFAN.

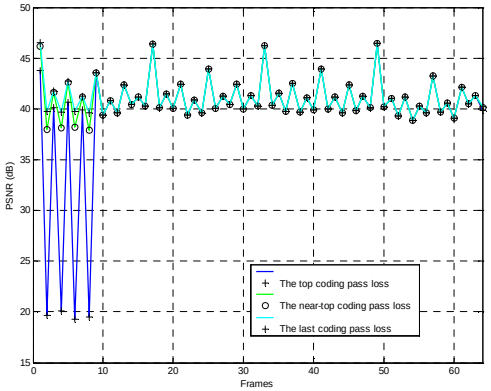


Fig. 25. PSNR with coding pass loss in block 0 of SSB 0 of the TSB $P(H_t, Y)$ of STEFAN.

Packet loss is a major cause of non-deterministic distortion for video streaming applications. For example, over fiber networks, bit errors rarely occur. The bit error rate of fiber networks is only 10^{-9} only [29]. The main reasons for packet losses are mostly because of network congestion, which causes packet losses in the ATM switch queue buffer [30]. As Fang et al. [29] and Biersack [30] pointed out, FEC protection scheme is effective to recover packet loss with minimum overhead for multimedia streaming. Hence, the proposed framework applies previous analysis on wavelet video to the design of a content-dependent interleaved FEC coding scheme for scalable streaming systems.

The basic concept of our context-adaptive FEC streaming scheme is to

add different FEC protection level (subject to predicted packet loss rate) to different wavelet subband data based on the data set's R-D slope (or, equivalently, the distortion-reduction rate). Fig. 26 illustrates this concept using some examples from the proposed algorithm. The content-adaptive FEC protection is applied to the coding block 0 of temporal subband P(Ht,Y) and spatial subband 0 of the STEFAN sequence. In this plot, the *y*-axis is the distortion reduction rate (i.e. the slopes of the conventional R-D curve as in Fig. 20) and the *x*-axis is the bitrate (including source data bits and FEC protection bits). The dash line is the original subband data without any protection, while the solid line with circle markers is the FEC protected data given 3% estimated packet loss rate and the solid lined with "plus" markers is the protected data given 8% estimated packet loss rate. The lower the rate point, the higher the protection level. The exact equation to compute the protection level will be described in a moment. Note that the function in Fig. 26 can be used for operational RDO streaming decision since it is exhibits rate versus source-and-channel distortion tradeoff.

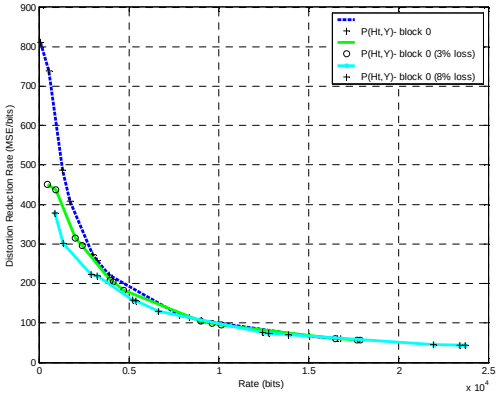


Fig. 26. Content-adaptive FEC protection examples.

In the proposed framework, for each group of video bitstream, an (*n*, *k*) Reed-Solomon (RS) code-based FEC is applied to add resiliency to the data. In Fig. 27, *n* is the codeword length of the RS encoder, *k* is the number of video data symbols (8 bits of bitstream data in this case), and *s* is the number of correctable symbols. The number of parity symbols is 2*s*, where 2*s* = *n*–*k*. If burst errors occur during transmission, then the RS decoder can correct up to *s* errors and detect up to 2*s* errors per codeword.

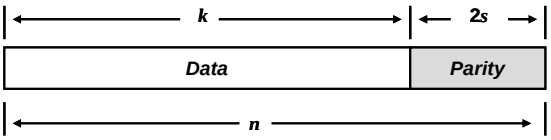


Fig. 27. An (n, k) RS codeword with *k* symbols of video data and 2s symbols of parity.

For 3D-ESCOT, each coding block *j* has temporal level index ω_j, component index v_j, spatial subband index τ_j, and block index ψ_j. Assume a coding block bitstream is generally divided into *l* codeword. Then, the importance of the various portions in one coding block can be expressed as in Eq (1).

$$c_{j,i}(x) = \alpha \cdot \sum_{n=0}^{i} \left(\frac{(T-\omega_j)\cdot U_1}{T} + \frac{U_2}{(Y-\upsilon_j)} + \frac{U_3}{(B-\tau_j)} + \frac{1}{(G-\psi_j)} \right), \quad (1)$$

where *i*=0, 1,…,*l*-1 , α is a scale-down factor, *T* is the maximum temporal level index, *Y* is the maximum component index, *B* is the maximum spatial subband index, *G* is the m[illegible] U_1, U_2, U_3 are w[illegible]mization of the co[illegible]tion problem can [illegible]e coding block $c_{j,i}(x)$ given by (1) subject to the network conditions. In addition, the bitstream of a coding block contains different number of coding passes. Larger value of *s* (number of correctable symbols) is required for earlier coding passes since the importance of data is arranged in coding pass order. Therefore, the allocation of expected protection to different coding passes stream is proposed to be Eq (2) and (3):

$$s_{j,i} = \left\lfloor \exp\left(\frac{|\lambda_0|}{\beta} \right) \cdot n_{pl} \right\rfloor - \left\lfloor \exp\left(c_{j,i}(x)\right) \right\rfloor + o$$

$$o = \begin{cases} 1 & \text{if } s_{j,i} \text{ is even} \\ 0 & \text{if } s_{j,i} \text{ is odd} \end{cases}$$

where λ0 is the R-D slope of the first

coding pass in block j and npl denotes the estimated packet losses given current bandwidth RBW, average packet size Ps and packet loss rate εpl. β is a scale factor determined from empirical analysis. Eq. (2) is designed so that sj,0 ≥ sj,1 ≥…≥ sj,l-1, that is, the level of protection decreases following coding passes order. Note that npl = ⌊εpl× RBW / Ps ⌋, where ⌊·⌋ stands for "the largest integer less then".

3. The proposed Packetization Scheme and Streaming Framework

In the following discussions, we use the terminology "block-bitstream segment" to describe a portion of bitstream bytes of a coding block across spatio-temporal subbands (see Fig. 19). A block-bitstream segment is composed of one or more coding passes. The packaging of the scalable bitstreams into UDP packets is accomplished following both rate control and error control constraints. These constraints try to fulfill the following goals:

1. Error protection level of a block-bitstream segment should depend on its entropy. The higher the entropy, the higher the protection level should be. Note that since a block-bitstream segment is only a small chunk of data in a coding block, the granularity of content adaptation of the FEC protection is at a very fine scale.
2. The streaming packet rate of the system should stay as low as possible. UDP packet size should be smaller than the MTU (Maximum Transmission Unit) allowed by the network links (typical size is around 1500 bytes for wired networks, and less than 1000 bytes for mobile networks). On the other hand, processing a lot of small packets causes very high overhead to the streaming system, especially on the client side. Therefore, a reasonable packet size is slightly smaller than the MTU.
3. Although interleaving with FEC works well for handling packet losses, it does introduce extra delay to the transmission of video data. Therefore, the selection of interleaving group size must take into account the end-to-end delay of the whole systems. In general, for video streaming, overall delay should be less than 10 seconds.

Packetization of FEC-protected data

As mentioned in the previous section, a systematic Reed-Solomon (RS) codeword comprising of data symbols and parity symbols is used for content-adaptive FEC protection. RS coding used for the protection of the block-bitstream segment is depicted in Fig. 28. Assume that the total number of coding block is L, i =0,..., L-1, for each coding block i, bitstream can be divided into m-data symbol unit, it begins with the first block-bitstream segment $C_{i,0}$ and continues through $C_{i,1}$, $C_{i,2}$, ... to $C_{i,m}$. An (n, k_x), x = 0,…,m, RS code is then applied to add resiliency to the m-data symbol unit. Since the block-bitstream segments have large variations in size, one must pack variable number of block-bitstream segments into a data unit to reduce packet overhead. In addition, different levels of protection are allocated to different portions of the coding block, $k_m \geq k_{m-1} \geq ... \geq k_0$. Furthermore, the data symbols gathered at the front end of the data unit, and the parity symbols are located at the back end of the data unit. For each data unit, there is a header that describes the protection level of the data unit. The header is also protected by RS coding. Also note that if data unit is not a multiple of k, zero-padding will be applied at the end of the data. These padding bytes do not have to be transmitted though.

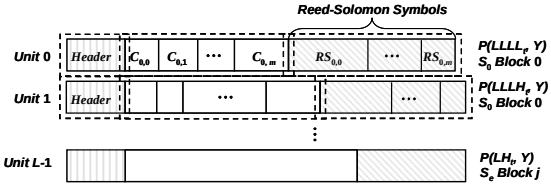


Fig. 28. Packetization for one group of video data.

Since we are dealing with a packet-loss

channel, not a bit-error channel, a byte-wise data-interleaving scheme is used to shuffle the RS coded data among several data packets before transmission. As illustrated in Fig. 29, a block-bitstream segment is spread across many packets (each packet is composed of the group of data in dashed lines in Fig. 29). For each packet, in addition to video data payload, we also have to transmit the highest protection level, temporal subband index, component index, spatial subband index, and block index in order to properly de-interleave the data. When interleaving is used, the interleaving depth must match the worst-case of channel conditions against burst errors. In addition, a large interleaving depth will have impact on the packet buffer size of the client and the end-to-end delay of packet transmissions. The interleaving depth should be appropriately chosen to handle the worst case error bursts of the networks. As mentioned in section 2, the number of parity symbols is 2*s*, where *s* means the number of correctable errors by an RS decoder. A data unit can be split into several *r* equal-length sub-units and each interleaved packet is composed of *q* data symbols from each sub-units. Hence, *q* is limited by the number of parity symbols *s*, and *p* is limited by the maximum end-to-end delay.

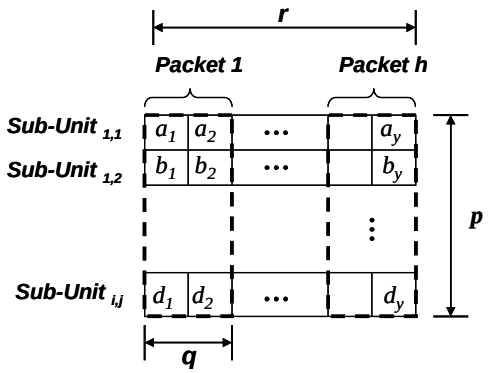


Fig. 29. Data-interleaving scheme for one group of video data.

Streaming policy

The proposed framework will adapt to the fast varying channel conditions by using the real-time network statistics feedbacks from the client side. Through standard RTCP receiver reports, the server can obtain the statistics such as round-trip time (RTT), jitter, short-term packet losses, and accumulative packet losses. The packet loss rate is used to compute the content-adaptive FEC-protected data rate-distortion tradeoff information as described in section 2. In addition, the server can compute the effective channel bandwidth through the last packet sequence number received by the client and loss rate. Based on the estimated channel bandwidth and the rate-distortion information, the system performs a dynamic rate allocation at discrete transmission time to enhance the perceived quality whenever the network bandwidth is good enough for perceptible quality improvement.

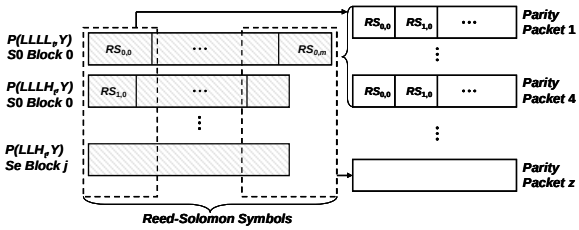


Fig. 30. Redundancy packets for protect source packets one group of video data.

For the correction of errors, parity packets are employed to recover from lost data packets. But some of parity packets may be lost or corrupted when transmitting packets over the networks based on the UDP protocol. For enhancing the system performance, error recovery mechanisms such as retransmission or error-correction can be applied to handle uncorrectable errors. Instead of using retransmission scheme to all parity packets, the proposed system delivers more redundancy parity packets to those packets carrying important portion of blocks and fewer to other packets. As seen in Fig. 30, all of the blocks are arranged according to the degree of importance of each spatial-temporal subband. In addition, the higher protection-level parity symbols are gathered together into one packet for maximum the efficiency of the error recovery scheme.

4. Experiments

This section presents the experimental results of the proposed video streaming system. The block diagram of the proposed streaming system is shown in Fig. 31. The system is based on the MPEG-21 Test Bed for Resource Delivery [31] (the source code of the original test bed can be downloaded from http://clabprj.ee.nctu.edu.tw/~mpeg21tb/). The test bed includes an IP transmission link emulator (based on the NIST Net [32]) that allows real-time emulation of various network conditions. We have added Reed-Solomon coding modules, a data-interleaving module, and a data de-interleaving module to the original test bed.

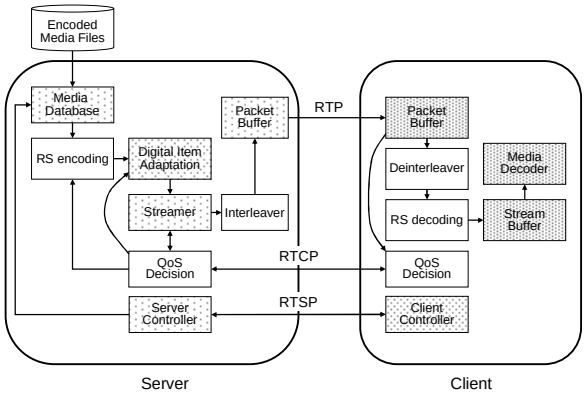


Fig. 31. Architecture of the proposed system.

The CIF version of the standard MPEG test sequences STEFAN and MOBILE are used for the experiments. Those sequences are encoded using MSRA 3-D wavelet video coding software [33] at 15 frames per second and a GOP is composed of 64 frames. Four levels of 5/3 MCTF temporal decomposition and three levels of 9/7 wavelet spatial decomposition are used for subband coding. The number of luminance (Y) blocks is around 1024 block-bitstream segments, and the number of chrominance (U and V) blocks is around 608 block-bitstream segments.

To evaluate the performance of the proposed system, reasonable range of packet loss rates should be used. Over wired links, studies showed that based on MPEG compressed video using the RTP and UDP transport protocols reported the average packet loss rates, ranging from 3.0 to 13.5 percent [34]. Over wireless links, Lai et al. [20] reported the characteristics of the MosquitoNet wireless network. The packet loss rates were 25.6% when packets were sent from a mobile host to a router, and 3.6% when packets are sent from a router to a mobile host. Rosa et al. [21] did a comprehensive study of the handover mechanisms during the disruption time in the wireless network. They reported that the packet loss caused by the handover mechanism was below 0.3%. Based on these published studies, we have set the packet loss rates of our experiments to 5%.

The proposed content-adaptive FEC protection framework is compared against a fixed FEC protection streaming system. The PSNR of the luma channel of the reconstructed video sequences are shown in 0 and Fig. 33. The level of protection for the content-adaptive FEC system is determined by Eq (2), while the level of protection of the fixed FEC is determined by the (predicted) average number of packet loss for each second. In either case, the maximal packet loss protection level can only recover up to 4% packet losses on average. As one can see from the figures, the adaptive FEC protection scheme works much better than the fixed level protection scheme.

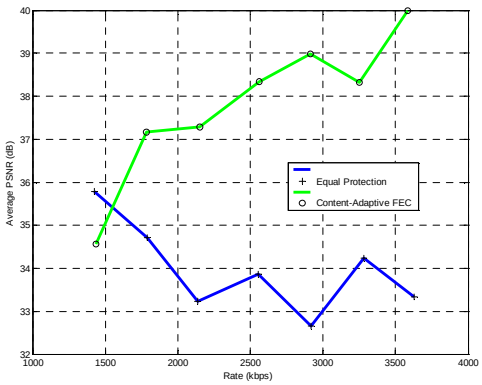


Fig. 32. Comparison between fixed and content-adaptive FEC protection for the STEFAN sequence. (frame rate:15, GOP size: 64, packet loss rate:5%)

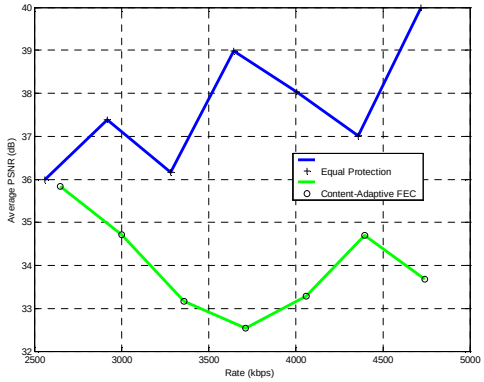


Fig. 33. Comparison between fixed and content-adaptive FEC protection for the MOBILE sequence. (frame rate:15, GOP size: 64, packet loss rate:5%)

5. Conclusions and Future Work

In this paper, a content-adaptive FEC protection and packetization framework for wavelet video streaming is proposed. The adaptive packet loss protection scheme using Reed-Solomon coding and data-interleaving is based on detail analysis of rate-distortion tradeoff of wavelet subband data. The experimental results show that with an adaptive fine-granularity FEC protection level packetization scheme, one can achieve much better quality than a fixed-level FEC protection scheme.

For future work, a runtime operational rate-distortion optimized streaming policy with joint optimization for minimal source coding distortion and packet-loss distortion will be investigated. Furthermore, the equation used for the determination of FEC protection level given estimated packet loss rate are designed based on empirical analysis. More rigorous derivation of the FEC protection level function is under investigation.

6. References

[1] S. J. Choi and J. W. Woods, "Motion-Compensated 3-D Subband Coding of Video," *IEEE Trans. on Image Processing*, vol. 8, pp. 155-167, 1999.

[2] ISO/IEC MPEG Test Group, "Subjective Test Results for the CfP on Scalable Video Coding Technology," *MPEG Documents N6383*, Mar., 2004.

[3] S. Brangoulo, R. Leonardi, M. Mrak, B. Pesquet Popescu, Jizheng Xu, "Draft Status Report on Wavelet Video Coding Exploration," *MPEG Documents N7571*, Oct., 2005.

[4] T. Berger. *Rate Distortion Theory: A Mathematical Basis for Data Compression*, Prentice-Hall, Englewood Cliffs, NJ, 1971.

[5] P. A. Chou and Z. Miao, "Rate-distortion optimized streaming of packetized media," *IEEE Transactions on Multimedia*, Vol. 8, No. 2, pp. 390-404, April 2006.

[6] A. K. Katsaggelos, Y. Eisenberg, F. Zhai, R. Berry, and T. N. Pappas, "Advances in Efficient Resource Allocation for Packet-Based Real-Time Video Transmission," *Proceedings of the IEEE, Special Issue on Advances in Video Coding and Delivery*, vol. 93, no. 1, pp. 135-147, Jan. 2005.

[7] X. Zhu, E. Setton, and B. Girod, "Congestion-Distortion Optimized Video Transmission Over Ad Hoc Networks," *EURASIP Signal Processing: Image Communication*, Vol. 20, no. 8, pp. 773-783, Sep. 2005.

[8] F. Zhai, C. E. Luna, Y. Eisenberg, T. N. Pappas, R. Berry, and A. K. Katsaggelos, "Joint Source Coding and Packet Classification for Real-Time Video Transmission over Differentiated Services Networks," *IEEE Transactions on Multimedia*, Vol. 7, No. 4, pp. 716-726, August 2005.

[9] T. Chu and Z. Xiong, "Combined Wavelet Video Coding and Error Control for Internet Streaming and Multicast," *EURASIP Journal on Applied Signal Processing: Special Issue on Multimedia Systems*, pp. 66-80, Jan. 2003.

[10] J. Dong and Y. F. Zheng, "Content-based Retransmission for 3-D Wavelet Video Streaming on the Internet," in *Proceedings IEEE Int. Conf. on Information Technology, Coding and Computing*, pp. 452-457,

April 2002.

[11] Y. Zhao, S. Ahalt, and J. Dong, "Content-based Retransmission for Video Streaming System with Error Concealment," *SPIE Defense and Security Symposium*, Orlando, April, 2004.

[12] W. Tan and A. Zakhor, "Real-time Internet Video Using Error Resilient Scalable Compression and TCP-friendly Transport Protocol," *IEEE Transactions on Multimedia*, 1(2):172-186, May 1999.

[13] J.C Bolot and T. Turletti, "Experience with Control Mechanisms for Packet Video in the Internet," *SIGCOMM Computer Communication Review*, 28(1), Jan. 1998.

[14] M. Kalman and B. Girod, "Techniques for Improved Rate-Distortion Optimized Video Streaming," *ST Journal of System Research*, Vol. 1, No. 3, Q1 2005.

[15] H. Wang, F. Zhai, Y. Eisenberg, and A. K. Katsaggelos, "Cost-Distortion Optimized Unequal Error Protection for Object-based Video Communications," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 15, no. 12, pp. 1505-1516, Dec. 2006.

[16] C.-L. Chang, S. Han, and B. Girod, "Sender-Based Rate-Distortion Optimized Streaming of 3-D Wavelet Video with Low Latency," *Proc. IEEE Intern. Workshop on Multimedia Signal Processing*, MMSP'04, Siena, Italy, September 2004.

[17] C.-L. Chang, S. Han, and B. Girod, "Rate-Distortion Optimized Streaming for 3-D Wavelet Video," *Proc. IEEE Intern. Conf. on Image Processing*, ICIP-04, Singapore, Oct. 2004.

[18] F. Zhai, Y. Eisenberg, C. E. Luna, T. N. Pappas, R. Berry, and A. K. Katsaggelos, "Packetization Schemes for Forward Error Correction in Internet Video Streaming," *Proc. 41st Allerton Conf. Communication, Control and Computing*, Oct. 2003.

[19] E. Martinian and C.-E. W. Sundberg, "Decreasing Distortion Using Low Delay Codes for Bursty Packet Loss Channels," *IEEE Transactions on Multimedia*, Vol. 5, No. 3, Sep. 2003.

[20] K. Shimizu, N. Togawa, T. Ikenaga, and S. Goto, "Reconfigurable Adaptive FEC System Based on Reed-Solomon Code with Interleaving," *IEICE Transactions on Information and Systems*, Volume E88-D, Number 7, pp. 1526-1537, 2005.

[21] V. Stankovic, R. Hamzaoui, and Z. Xiong, "Efficient Channel Code Rate Selection Algorithms for Forward Error Correction of Packetized Multimedia Bitstreams in Varying Channels," *IEEE Trans. on Multimedia*, Vol. 6, No. 2, pp. 240- 248, April 2004.

[22] M. Gallant and F. Kossentini, "Rate-Distortion Optimized Layered Coding with Unequal Error Protection for Robust Internet Video," *IEEE Trans. on Circuits and Systems for Video Technology*, Vol. 11, No. 3, pp. 357-372, March 2001.

[23] J. Goshi, A. E. Mohr, R. E. Ladner, E. A. Riskin, and A. Lippman, "Unequal Loss Protection for H.263 Compressed Video," *IEEE Trans. on Circuits and Systems for Video Technology*, Vol. 15, No 3. pp. 412- 419, March 2005.

[24] S. Dumitrescu, X. Wu, and Z. Wang, "Globally Optimal Uneven Error-Protected Packetization of Scalable Code Streams," *IEEE Trans. on Multimedia*, Vol. 6, No. 2, April 2004.

[25] M. Zink, J. Schmitt, and R. Steinmetz, "Layer-Encoded Video in Scalable Adaptive Streaming," *IEEE Trans. on Multimedia*, Vol. 7, No. 1, February 2005.

[26] ISO/IEC MPEG Video Group, "Wavelet Codec Reference Document and Software Manual V1.0," *MPEG Document N7573*, July, 2005.

[27] 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, Special Issue on*

Wavelet Applications, vol. 10, pp. 290-315, 2001.

[28] B.-J. Kim, Z. Xiong, W. Pearlman, "Low bit-rate scalable video coding with 3-D set partitioning in hierarchical trees (3-D SPIHT)," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 10, no. 8, pp. 1374-1387, December 2000.

[29] R. Fang, D. Schonfeld, R. Ansari, and J. Leigh, "Forward Error Correction for Multimedia and Tele-immersion Streams," *EVL Technical Report*, 2000.

[30] E. W. Biersack, "Performance Evaluation of Forward Error Correction in an ATM Environment," *IEEE Journal on Selected Areas in Communications*, 11(2):631--640, May 1993.

[31] ISO/IEC JTC 1/SC 29/WG11, *ISO/IEC TR21000-12: MPEG-21 Test Bed for Resource Delivery*, ISO, Jan. 2005.

[32] Mark Carson and Darrin Santay, "NIST Net: a Linux-based networkemulation tool," *Computer Comm. Review (ACM SIGCOMM)*, 33(3), pp. 111-126, 2003.

[33] R. Xiong, X. Ji, J. Xu, and F. Wu, "MSRA scheme for SVC CE1," *MPEG Input Document M11320*, Palma de Mallorca, ES, Oct. 2004.

[34] J. M. Boyce and R. D. Gaglianello, "Packet loss effects on MPEG video sent over the public Internet," *In Proceedings of ACM Multimedia*, pp. 181-190, September 1998.

[35] K. Lai, M. Roussopoulos, D. Tang, X. Zhao, and M. Baker, "Experiences with a Mobile Testbed," *Proc. of the Second Intern. Conf. on Worldwide Computing and its Applications (WWCA'98)*, March 1998.

[36] R. Rosa, C. -P. Angel, D. -F. Manuel, O. Luis, and G. Antonio, "On the Traffic Disruption Time and Packet Lost Rate during the Handover Mechanisms in Wireless Networks," *Proceedings AINA*, pp.351-354, 2004.