

Scalable Video Adaptation for IPTV

Huifeng Shen*

University of Science & Technology of China
Hefei, P. R. China 230027

shenhf@mail.ustc.edu.cn

Xiaoyan Sun, Feng Wu, Shipeng Li

Microsoft Research Asia
Beijing, P. R. China 100080

{xysun, fengwu, spli}@microsoft.com

ABSTRACT

IPTV based on full-duplex networks potentially provides much richer media experiences than traditional TV. However, video contents compressed with non-scalable coding schemes greatly constrain the potential. This paper proposes a new IPTV architecture, which fully takes the advantages of scalable video for adaptation to both networks and devices. Firstly, network adaptation is mainly performed in scalable IPTV servers. An algorithm is proposed to optimally allocate available bandwidth between different frames and different channels, considering the dependency, weighting factors and rate-distortion data of scalable layered streams. Secondly, home gateway provides device adaptation, which is accomplished by reshaping the received scalable video. Considering the existing set-top boxes equipped with hardware decoder, we also propose a fast and effective algorithm to convert a scalable video to a non-scalable one in the home gateway.

Categories and Subject Descriptors

C.3 [Special-purpose and application-based systems] Real-time systems

General Terms

Algorithms, Design, Experimentation

Keywords

IPTV, scalable video coding, rate allocation, scalable to non-scalable transcoder

1. INTRODUCTION

IPTV (Internet Protocol Television) has gotten rapid developments in the last couple of years [1]–[3]. It is greatly different from traditional TV, in which a full-duplex IP broadband network, no matter that it is cable, telco or a wireless one, is used as the medium. In general, IPTV upon the full-duplex network can provide much richer media experiences than traditional TV.

Unfortunately, video content delivered in current IPTV systems is still compressed with traditional non-scalable technologies (e.g. MPEG-2, MPEG-4 AVC/H.264 and WMV). The generated non-

scalable stream is difficult to be reshaped for adaptation to networks and devices without computation-intensive transcoding. It considerably constrains the potential of IPTV on utilizing network bandwidth in an effective way and providing more novel services. For example, in traditional TV systems, a fixed bandwidth is pre-allocated to each channel to facilitate radio transmission, which is also the target bit rate for video compression. However, the programs delivered in channels are quite different. Some of them are very simple and can be compressed very effectively at medium bit rates, while others are very complicated and need higher bit rates to achieve acceptable display quality. Due to the equal bandwidth allocation, overmuch bandwidth is always spent on simple video channels, while insufficient bandwidth is assigned to complicated video channels in case the total available bandwidth is not so adequate. It implies inefficient utilization of network resources. However, in IPTV systems, if compressed video offers enough adaptation and flexibility, the bandwidth of each channel and even the sending rate of a channel at different periods can be regulated dynamically according to the requirements of both video content and IP network.

So far, IPTV is usually served with set-top boxes and viewed on television sets. The compressed program delivered by IPTV is of high quality and of high resolution. However, it is a tendency that more and more devices that can playback media content, such as cell phones, pocket PCs and portable media centers, etc., are going to be involved in IPTV applications. These devices are quite different in display resolutions, computation capacities and connection methods. The IPTV systems using non-scalable video media have to employ multiple versions of transcoding to serve these devices.

Recently, scalable video coding technologies have been proposed to compress video contents in a scalable fashion. The resulting streams can be partially decoded at a desired frame rate, quality and resolution. It facilitates compressed video to flow freely from one device to another through wired or wireless networks like water without the hassle of transcoding. In the past, scalable video suffered from severe loss on coding efficiency as the price of scalability. It prevented scalable video coding from being deployed extensively in industries. However, the coding efficiency of scalable video coding has been significantly improved in recent years. For example, MPEG-4 FGS is a promising scheme on providing fine-granularity temporal and quality scalabilities [4]. Advanced FGS coding schemes proposed in [5]–[9] can achieve more than 2dB gain over MPEG-4 FGS. In comparison with the corresponding non-scalable coding schemes, the performance losses of the scalable approaches in the target rate range are only about 1dB in terms of PSNR. Scalable video coding methods adopting motion compensated temporal filter (MCTF) can achieve comparable and even better performance than non-scalable video coding schemes [10][11].

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

IPTV Workshop, International World Wide Web Conference, May 23, 2006, Edinburgh, Scotland, United Kingdom

*This work has been done while the author is with Microsoft Research Asia.

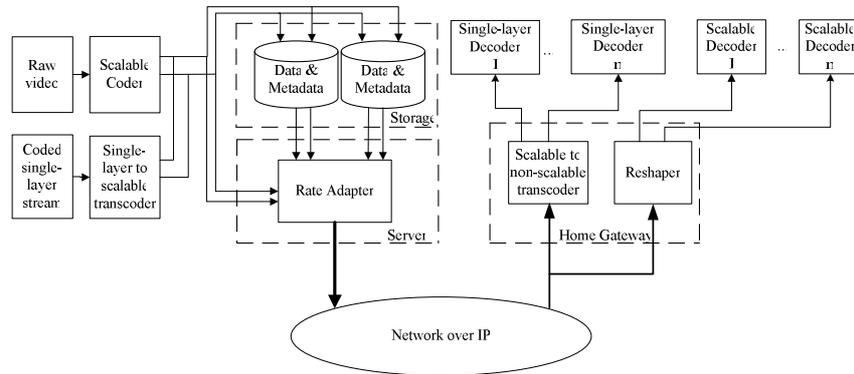


Fig. 1. The proposed architecture of IPTV with scalable video.

In this paper, a new IPTV architecture is proposed to provide full adaptation to networks and devices. First of all, video contents stored, transmitted and played back in the proposed IPTV system are in the scalable fashion. Besides compressed video data, each scalable stream owns its metadata to describe the scalable-related information, such as rate-distortion pairs and dependencies between sub-streams. Thus, a two-level adaptation is supported in the proposed IPTV system. Firstly, the scalable IPTV server provides the network adaptation. An algorithm is presented to dynamically allocate network bandwidth within a channel and between channels according to users' request and stream metadata. It not only increases network utilization in IPTV systems but also enables consistent quality between channels. Secondly, we introduce a home gateway for device adaptation in the architecture, which can easily reshape the received IPTV data for different devices. At the same time, since existing set-top boxes can not process received scalable streams, we also propose a fast and effective transcoder in the home gateway to convert scalable stream to non-scalable high-quality standard-definition (SD) stream. Totally, the new IPTV architecture with scalable video enables many new features that can't be provided by current IPTV systems.

The rest of this paper will be organized as follows. Section 2 describes the proposed new architecture for IPTV. Section 3 discusses the rate allocation within a channel and between different channels at the server side. Section 4 proposes a fast transcoder to convert scalable stream to non-scalable one in the home gateway. Finally, conclusions are presented in section 5.

2. THE PROPOSED ARCHITECTURE FOR IPTV

As illustrated in Fig.1, a new architecture is proposed for transmitting scalable video in IPTV systems. It consists of three main parts: scalable video storage, scalable IPTV server and home gateway.

2.1 Scalable Video Storage

Scalable video usually consists of a series of layered streams. There are three basic scalabilities on video content: SNR scalability for picture quality, spatial scalability for frame size, and temporal scalability for frame rate. In many practical applications, one scalable stream may contain more than one kind of basic scalabil-

ity. Each layer can be partially and even completely dropped during transmission and decoding.

Unlike non-scalable streams, there is some additional information which is very important for IPTV systems to benefit more from scalable video. They are unnecessary to be delivered to clients but are very helpful for servers to select data for transmission. Here we divide them into three categories.

- Relationship between layered streams.

The relationship between layered streams is either dependent or independent. When the server selects a layered stream according to user request, network bandwidth, and device capability, all layered streams on which it is dependent also need to be delivered to clients.

- Weighting factors

It is used to describe the dependency degree of dependent layered streams in terms of distortion effect. It indicates how much the distortion of a layered stream propagates to its dependent streams. It should be used together with the following rate and distortion data.

- Rate-distortion data

When a layered stream is selected, it does not mean that all bits in this layer shall be delivered to clients. Each layered stream contains a set of rate and distortion pairs. How many bits that are delivered in each layer is decided based on rate-distortion, weighting factor and available network bandwidth.

All these additional information can be easily obtained when video sequences are compressed into streams by scalable encoders [4]~[9]. The additional information will be stored as metadata into scalable streams. Now MPEG is extending ISO base file format to contain them in media files [12]. Though a number of approaches have been presented to achieve and storage the additional information, all of them are out of the scope of this paper.

2.2 Scalable IPTV Server

Commonly, scalable video streams stored in IPTV systems are of high resolution and of high frame rate at high bit rates to satisfy a broad range of applications. Upon the clients' request, Rate Adapter in Fig.1 retrieves the streams either from storage or directly gets the streams from encoder for real-time applications. According to network conditions and different requests of receivers, the high-quality streams are firstly truncated by Rate Adapter

and then the bit streams after truncation are delivered over networks. Allocating the current bandwidth between different frames and different channels is a key functionality of Rate Adapter. The simplest rate allocation is uniform bit-rate allocation. But it often results in quality fluctuation between frames and channels. It has been proved that human visual perception is very sensitive to the quality fluctuation between consecutive frames. In addition, in TV-providing systems, when consumers switch channels on end-receivers, large quality variation between channels could also annoy human's eyes. Thus, it is very important for the service provider to control quality of each channel by allocating the bandwidth for different demands.

To better take the advantages of scalable streams and fully utilize available network bandwidths, an optimal rate allocation is needed at the server side. In the literatures [14]~[17], many approaches use look-ahead sliding window in rate allocation, which results in delay. Therefore, such methods are unsuitable for real-time applications. In this paper, a simple but effective way for scalable streams rate allocation is proposed to achieve constant quality between different frames and controllable quality between different channels in no-delay mode. It is discussed in detail in Section 3.

2.3 Home Gateway

The home gateway in the proposed IPTV system provides device adaptation. It can further reshape the received scalable streams to fit for each device capability (e.g., display resolution, computation power, and memory). Fig.2 depicts the advantages provided by scalable video at home gateway. On one hand, when a MPEG SD non-scalable stream is served, it can be directly sent to existing set-top boxes with MPEG hardware decoder. But since the SD resolution and high computation are not suitable for portable devices, the HD stream has to be transcoded to a lower resolution version requiring low-complexity decoding. As the resolution and computation power in different portable devices can be diverse in a wide range, multiple transcoders have to be equipped on the home gateway. On the other hand, when scalable video is adopted in the IPTV system, only one transcoder is needed to convert scalable video to MPEG SD stream for the old set-top boxes without scalable video decoder. The adaptation to all other devices can be readily supported by stream truncation for a desired resolution, frame rate and bit rate.

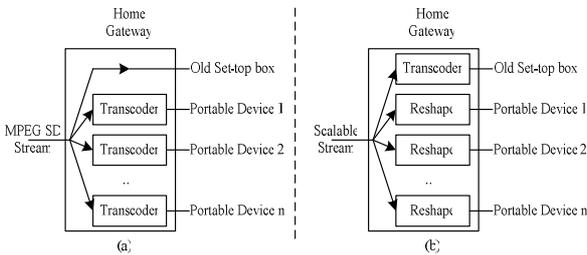


Fig. 2. Comparison between the home gateways, (a) the received data is MPEG SD stream and (b) the received data is scalable stream.

In general, the transcoding from scalable video to MPEG non-scalable video performed in the home gateway should be fast and easy in terms of computational complexity. Therefore, we propose a fast transcoder to satisfy this demand. Unlike traditional

transcoders [18] which reduce bit rates to adapt to the bandwidth of the network, the proposed transcoder aims to convert the scalable stream to non-scalable stream and to achieve better performance.

3. RATE ALLOCATION

In this section, an optimal rate allocation method is presented. It considers the rate-distortion relationships between channels as well as within one channel to control the playback quality of channels at server side.

3.1 Problem Description

Rate allocation for minimizing the quality variation in a single channel is always formulated as minimizing the average of absolute differences between adjacent frames [13][15]. Let N denote the number of channels. $M(i)$ presents the number of frames within the channel i in a certain time slot, here $0 \leq i < N$. $D_{i,j}(R_{i,j})$ denotes the distortion of frame j in channel i , where $R_{i,j}$ denotes the number of coded bits. Minimizing the quality variation between frames in a single channel can be formulated as

$$\min_{R_{i,j}} J_i = \frac{1}{M(i)-1} \sum_{j=0}^{M(i)-2} |D_{i,j}(R_{i,j}) - D_{i,j+1}(R_{i,j+1})|, \quad (1)$$

which is subject to the rate constraint described in formula (2).

$$\sum_{i=0}^{N-1} \sum_{j=0}^{M(i)-1} R_{i,j} \leq R_{budget}, \quad (2)$$

where R_{budget} denotes the current total available bits for all channels in this time slot. Ideally, the problem can be solved when the distortion of each frame is identical with the neighboring frames within the channel (i.e. $D_{i,j} = D_{i,j+1}$, $0 \leq i < N$ and $0 \leq j \leq (M(i)-2)$). In other words, there is no distortion variation within a channel.

Since the enhancement layer of scalable stream can be arbitrarily truncated, the R-D function $R_{i,j}(D)$ of enhancement layer is continuous and monotonous, and the total curve can be linear-interpolated by discrete R-D sample points which are pre-stored during encoding[13]. Therefore, it is possible to achieve a desired distortion distribution between channels under an overall rate constraint, while minimizing the total distortion optimally.

3.2 Optimal Rate Allocation

According to the formulas given in the previous subsection, a 2-D look-backward sliding window-based rate allocation method is presented in this subsection to meet real-time demand. Since the rate allocation is performed based on the previous transmitted frames, it is suitable for IPTV applications with less buffer requirement and lower delay. Moreover, the R-D information of both within one channel and between channels is taken into account in the proposed approach.

In the case of IPTV scenario, we can construct a frame-based look-backward sliding window with length L in terms of frame number. Thus the bits' number WR_t assigned to the 2-D window can be computed as

$$WR_t = \begin{cases} R_{budget} & t = L-1 \\ WR_{t-1} - \sum_{i=0}^{N-1} R_{i,t-1} + R_{budget}/L & t > L-1 \end{cases}, \quad (3)$$

here t is the last frame number in the window. Under the constant distortion constraint within a channel and the distortion relationship constraint between channels, we use bisection search algorithm to find the optimal distortion value for each channel and then compute the rates for the last frame in the window according to $R_{i,j}(D)$.

We present the rate allocation method in detail as follows.

Step 1: initializing the start point of the distortion search. Let $D_b(i,j)$ denote the distortion of the base layer of frame i in channel j . The minimal distortion value D_{min} of all $D_b(i,j)$ is selected to be the high bound of current search. The low bound of current search is set to zero.

Step 2: using bisection search algorithm to find the optimal distortion D that meets the criteria

$$\left| \sum_{j=t-L+1}^t \sum_{i=0}^{N-1} R_{i,j}(wD) - WR_t \right| < \delta, \quad (4)$$

where δ is an adjustable factor that makes a trade-off between rate accuracy and computation complexity of the algorithm.

Step 3: computing the rate that should be allocated to the frame t for each channel by $R_{i,j}(D)$.

Step 4: sliding the 2-D window by one frame step and update window's information using the equation (3). Then go to step 1 for the successive frames.

3.3 Simulation Results

As an example of SNR scalable coding schemes, PFGS scheme proposed in [7] is used in this subsection to illustrate the performance of the optimal rate allocation method, in which the enhancement layer is coded by the bit-plane method and with LPLR mode. The test sequences, City, Crew and Harbour at 4CIF and 30fps are used in the experiment as different channels. There are 150 frames for each sequence coded with intra-period of 15. The base layers' quantization parameters (QPs) are set to 32 for I frames and 44 for P frames in City and Crew sequences and 34 for I frames and 46 for P frames in Harbour sequence. The total bandwidth for all the enhancement layer bits is 18432 Kbps.

Fig. 3 compares the distortion variation frame by frame between uniform bit allocation and our proposed method with the sliding window length of 10. It can be seen that the optimal bit allocation significantly reduces the distortion variation between frames and between channels.

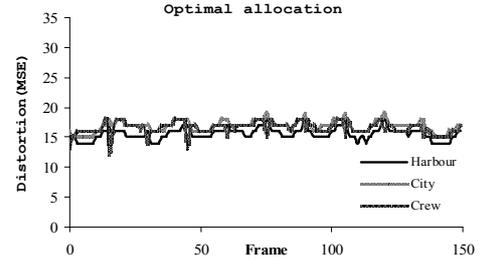
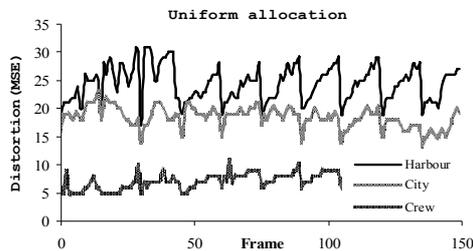


Fig. 3. Simulation results comparing the distortion per frame using different rate allocation methods.

4. FAST TRANSCODER IN HOME GATEWAY

An efficient transcoding method is presented in this section to enable scalable to non-scalable device adaptation. Among current advance FGS coding schemes, the PFGS system proposed in [6][7] achieves good coding efficiency as well as scalability. Compared with traditional non-scalable coding methods, three INTER coding modes and one more motion compensation loop are required for PFGS enhancement layer coding, which brings more difficulties in transcoding. In this chapter, we choose PFGS scalable coding scheme to clarify the proposed transcoding approach. H.264 coding method is selected as the exemplified non-scalable scheme.

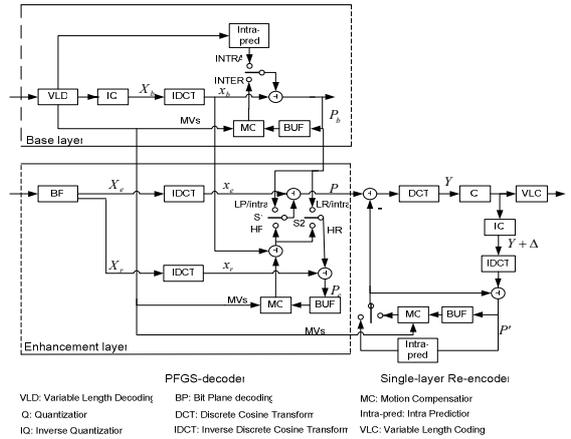


Fig. 4. The block diagram of the cascaded pixel-domain transcoder

4.1 The Cascaded Pixel Domain Transcoder

The straightforward solution for PFGS scalable video to H.264 non-scalable video conversion is cascaded transcoder in which the scalable video content is firstly fully decoded by PFGS decoder and then the resulted signal is re-encoded by H.264 encoder, as illustrated in Fig.4. Obviously, the cascaded transcoder is undesirable to real applications especially for IPTV scenario due to the high complexity.

4.2 The Fast Transcoder

Let P , P' denote the reconstructed pictures of input stream and output stream, respectively. X_b , X_e , and X_r are corresponding to the base layer bits, the enhancement layer bits, and the enhancement layer reference bits. Δ means the re-quantization error. Based on

- [3] Microsoft, "Delivering IPTV with the Windows media platform", <http://go.microsoft.com>.
- [4] W. Li, "Streaming video profile in MPEG-4", IEEE Trans. on Circuits and Systems for Video Technology, vol. 11, no 3, pp 301-317, March 2001.
- [5] F. Wu, S. Li and Y.-Q. Zhang, "A framework for efficient progressive fine granularity scalable video coding", IEEE Trans. on Circuits and Systems for Video Technology, vol. 11, no 3, pp 332-334, Mar 2001.
- [6] X. Sun, F. Wu, S. Li, W. Gao, Y.-Q. Zhang, "Macroblock-based progressive fine granularity scalable video coding", International Conference on Multimedia and Expo, Tokyo, Japan, Aug 2001
- [7] Y. He, F. Wu, S. Li, Y. Zhong and S. Yang, "H.26L-based fine granularity scalable video coding", IEEE International Symposium on Circuits and Systems, vol. 4, pp 26-29 May 2002.
- [8] M. Schaar and H. Radha, "A hybrid temporal-SNR fine-granular scalability for Internet video", IEEE Trans. on Circuits and Systems for Video Technology, vol. 11, no 3, pp 318-331, 2001.
- [9] J.-R. Ohm, "Advances in scalable video coding", Proceedings of the IEEE, vol. 93, no 1, pp 42-56, Jan.2005.
- [10] R. Xiong, F. Wu, J. Xu, S. Li and Y.-Q. Zhang, "Barbell lifting wavelet transform for highly scalable video coding", Picture Coding Symposium, special issue on advances in MCTF/Wavelet, Oct. 2004.
- [11] ITU-T and ISO/IEC JTC1, "Scalable Video Coding - Working Draft 1", JVT-N020, Jan 2005
- [12] M. Z. Visharam, A. Tabatabai, D. Singer, "Supporting the Storage of MPEG-21: Part 13 Scalable Video by an extension of the ISO Base Media File Format", ISO/IEC, JTC1/SC29/WG11, m11422, Palma, October 2004.
- [13] X. Zhang, A. Vetro, Y. Shi, H. Sun, "Constant quality constrained rate allocation for FGS-coded video", IEEE Trans. on Circuits and Systems for Video Technology, vol.13, no.2, February 2003.
- [14] H. Cheng, X. Zhang, Y. Shi, A.Vetro, H. Sun, "Rate allocation for FGS coded video using composite R-D analysis", Proc. IEEE Int'l Conference on Multimedia and Expo, July 2003.
- [15] Y. Liu, G. Er, Q. Dai, "Improved rate allocation method based on sliding window for FGS video bit-stream", IEEE Int'l Conferences on Acoustics, Speech, and Signal Processing, 2004 (ICASSP'04).
- [16] X. Zhao, Y. He, S. Yang, Y. Zhong, "Rate allocation of equal image quality for MPEG-4 FGS video streaming", IEEE Packet Video, 2002.
- [17] L. Zhao, J. Kim and C. -C. Kuo, "MPEG-4 FGS video streaming with constant-quality rate control and differentiated forwarding", SPIE VCIP, 2002
- [18] A. Vetro, C. Christopoulos, and H.Sun, "Video transcoding architectures and techniques: an overview", Signal Processing Magazine, IEEE, Volume 20, Issue 2, March 2003 Page(s):18 – 29
- [19] Y.-C. Lin, C.-N. Wang, T. Chiang, A. Vetro, H. Sun, "Efficient FGS to single layer transcoding", International Conference on Consumer Electronics, 2002. ICCE. 2002 Digest of Technical Papers. 18-20 June 2002.
- [20] M. Horowitz, A. Joch, F. Kossentini, A. Hallapuro, "H.264/AVC baseline profile decoder complexity analysis", IEEE Trans. on Circuits and Systems for Video Technology, vol. 13, no. 7, July 2003

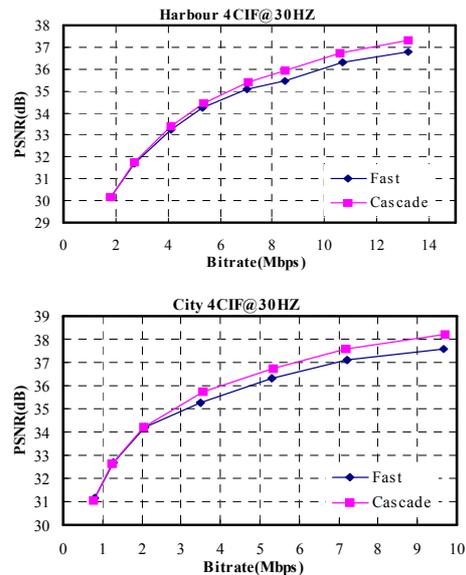


Fig. 6. Comparison on performance of the transcoding schemes.

TABLE I PERFORMANCE COMPARISON BETWEEN THE CASCADED TRANSCODER AND THE FAST TRANSCODER.

SEQUENCE	TRANSCODERS	QP = 36		QP = 30		QP = 24	
		SNR(dB)	Rate(kbps)	SNR(dB)	Rate(kbps)	SNR(dB)	Rate(kbps)
FOREMAN	CASCADE	31.03	117.95	33.72	257.81	35.67	517.66
	FAST	31.14	119.20	33.83	261.87	35.10	523.54
NEWS	CASCADE	32.46	72.94	35.70	136.03	38.96	270.75
	FAST	32.47	74.01	35.65	138.14	38.35	275.72
MOBILE	CASCADE	27.62	238.73	30.09	516.55	30.96	928.67
	FAST	27.58	249.61	29.73	537.38	30.33	941.04