System Modeling and Implementation of MPEG-4 Encoder under Fine-Granular-Scalability Framework

Literature Survey

Embedded Software Systems
Prof. B. L. Evans

by

Wei Li and Zhenxun Xiao

March 25, 2002

Abstract

Stream media over the Internet is now the mainstream of media communications due to the blooming of the Internet as well as the increasing demanding. A critical issue in providing quality service of media over the Internet is scalability.

In this literature survey, we will review the four basic scalable profiles proposed by Moving Picture Experts Group (MPEG), namely data-partitioning, SNR scalability, spatial scalability and temporal scalability. Furthermore, we will investigate the newly proposed amendment of MPEG-4, known as fine-granular-scalability (FGS) framework. We will discuss the relationship between FGS framework and MPEG scalable profiles. The objective of the project is to try to adopt the scalable profiles with the newly proposed FGS framework.

The work will be separated into two stages, system modeling and software implementation.

Introduction:

MPEG-1, MPEG-2 and MPEG-4 are a set of standards developed by moving pictures experts group (MPEG) for compression of motion video optimized at various rates. Standards developed by MPEG only define the syntax of encoded bit streams, and they can be applied to all kinds of applications.

Due to the blooming of the Internet and increasing demand, streaming videos over the Internet has become the mainstream of media communications [1, 6].

In this survey, we will first present an overview about the streaming media over the Internet, specifically the scalability issue. Then, we will investigate the scalability profiles in MPEG-2. After that, we will introduce the amendment of fine-granular-scalability framework for MPEG-4 as well as the related open issues. Finally, we will formulate our objective and the way that we are going to approach it.

Overview of Streaming Media over the Internet:

Providing streaming media services over the Internet poses higher requirements than traditional video coding and decoding techniques, which are designed to provide optimized services for certain video quality at given bit rates [4]. These special requirements are classified as scalable issue, and some constraints for providing scalable service include,

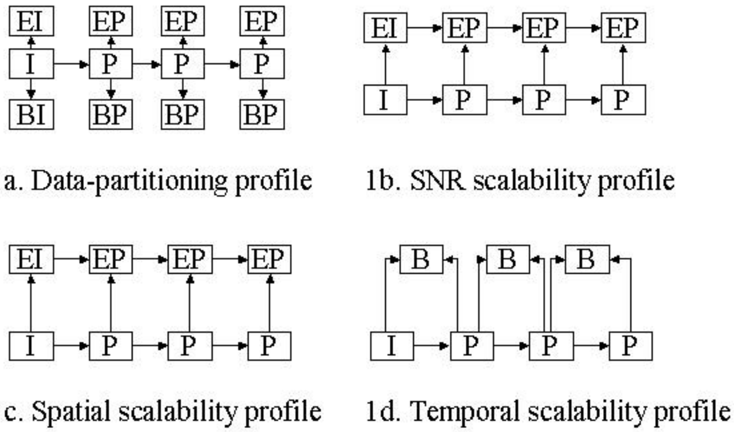
1. be thrifty in using network bandwidth;
2. be able to adapt to the dynamic distribution of bandwidths caused by network congestion;
3. be able to survive on top of best-effort IP service;
4. not pose requirement for extra disk space.

MPEG standards adopt a technique called the layered scalability to provide efficient and economical scalable service. The idea is to separate the media information into different layers, with a base-layer containing coarse quality video and enhancement layers containing fine information. By separating information into different layers, there are three obvious advantages,

1. If the media server knows in advance of the users' processing capabilities and the bandwidth of channels, it can select different layers of information to transmit accordingly;
2. If the media server does not know in advance about the network bandwidth, it can give different priorities to different layers. And the routers on the channel can select the various quality of services on-the-fly;
3. If the media server does not know in advance about the users' processing capabilities, it can just transmit the layered content to different users, and let the users to choose the encoded information according to their own processing capabilities accordingly.

Different Scalable Profiles in MPEG-2:

The scalability issue has been extensively investigated in MPEG-2, and there are four basic profiles proposed to provide the scalable service, known as data partitioning, SNR scalability, spatial scalability and temporal scalability [5].



1a. Data-partitioning profile 1b. SNR scalability profile

1c. Spatial scalability profile 1d. Temporal scalability profile

Figure 1. Illustration of different scalable Profiles, data-partitioning, SNR scalability, spatial scalability and temporal scalability, in MPEG-2, with only two layers, known as base-layer and enhancement-layer. I, P and B in figure refer to I-frame, P-frame and B-frame in MPEG standard. BI, BP, EI and EI denote base-layer I-frame, base-layer P-frame, enhancement-layer I-frame and enhancement-layer P-frame in scalable profiles in MPEG-2 respectively. The arrows connecting different frames show the reference relationship between them.

As illustrated in Fig. 1a, the profile of data partitioning partitions an encoded single bit stream into base partition and enhancement partition. Base partition contains the lower order DCT coefficients and enhancement partition contains the high-frequency DCT coefficients.

The profile of SNR scalability, as shown in Fig. 1b, generates separate bit streams for different layers. The enhancement layer carries the DCT refinement coefficients that can be used to increase the signal-to-noise ration when used together with the DCT coefficients carried in the base layer. And different SNR layers have the same spatial and temporal resolution.

The profile of spatial scalability, as shown in Fig. 1c, is to generate separate layers at different resolutions in spatial domain. Base layer and enhancement layer have the same resolution in time domain.

The profile of temporal scalability, as shown in Fig, 1d, is to generate separate layers at different resolutions in time domain. Base layer and enhancement layer have the same resolution in spatial domain.

In the four scalability profiles described above, the enhancement layer contains the fine information on different domains. When combined with the corresponding base layer for decoding, it can be used to supplement the fine information into the coarse information carried in the base layer.

FGS Framework under MPEG-4:

In the proposed scalability profiles of MPEG-2, the enhancement layer is encoded through run-level entropy encoding schema. As a consequence of this encoding schema, the decoder needs to decode all the information in the enhancement layer to be able to enhance the base layer or unable to use it at all. Video coding performance analysis of the quality of service provided by the scalability profiles proposed in MPEG-2 in respect of continuously increasing of available bit rate show a staircase effect, (Fig. 2).

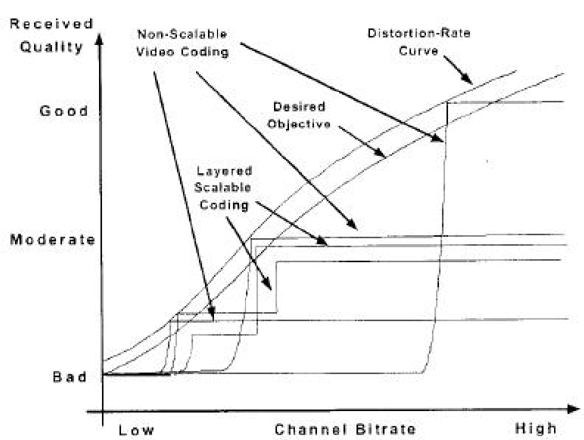


Figure 2. Illustration of video coding performance [2]

For the sake of proving scalable service, fine-granular-scalability (FGS) framework was proposed as an amendment in MPEG-4 [2, 3]. The key issue of FGS framework is the introducing of a new encoding schema, known as bit-plane (Fig. 3), in substitution of run-length entropy encoding schema which is used in MPEG-2 profiles for the encoding enhancement layer.

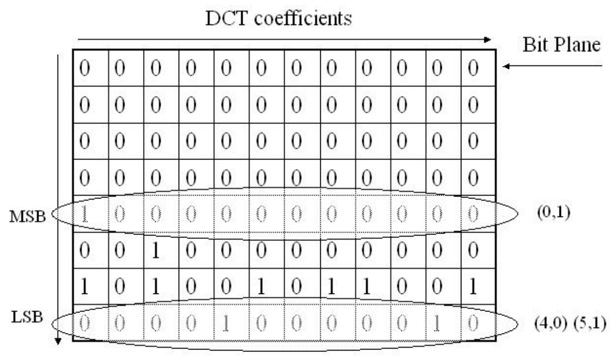


Figure 3. An example of bit-plane encoding schema.

In the figure, each column is the 8-bit representation of a DCT coefficient. Traditional run-level encoding schema is to encode each DCT coefficient individually. In bit-plane schema, all DCT coefficients are arranged as a matrix, each row with the bit from all DCT coefficients at the same significant position are encoded separately. The bit-plane with all zero bits are not coded.

As illustrated in Figure 3, the encoding order for bit-plane encoding schema is from the most significant bit (MSB) to the least significant bit (LSB). When enhancement layer is encoded in this way, decoder is able to utilize partial information of the enhancement layer together with the information from base layer for high quality service.

Open Issues:

Drifting and code efficiency are two critical factors to be considered in various bandwidth environment such as in current Internet. Since the prediction is always based on the base layer, the coding efficiency of the FGS might be worse than traditional SNR scalability scheme [7]. On the other hand, if the prediction were entirely based on the enhancement layer as in traditional SNR scalability scheme, packet losses from enhancement layer would propagate to the end of a group of picture even if there were sufficient bandwidth later.

Progressive Fine-Granular-Scalability (PFGS) was proposed to make the prediction more accurate and increase the coding efficiency by also using high quality references [7]. The drawbacks of PFGS are extra frame buffer and computational complexity. Furthermore, it is found that different reference layers contribute to the code efficiency differently. Therefore, how to select minimal reference layers to obtain significant code efficiency is an open problem [7]. Fluctuation and increase of DCT coefficients when switching from low quality reference to high quality reference need to be further investigated [7]. Meanwhile, PFGS also suffers from the drifting problem due to the reference to the enhancement layers.

An effective and flexible scheme to reduce drift is to control the PFGS coding at macro block level as presented in [8]. In this scheme, references are selected per macroblock by using one of the following three INTER modes: an enhancement macroblock is predicted and reconstructed both from the previous low quality reference in mode 1, both from the previous high quality reference in mode 2, or predicted from the previous high quality reference, while reconstructed from the previous low quality reference at both encoder and decoder in mode 3. In this way, a good trade off between drift reduction and code efficiency is achieved.

Objective and Approach:

The fine-granular-scalability framework proposed in [2, 3] only introduces minor modifications on top of codec structure, and the codec tools that implement SNR, spatial and temporal scalability profiles proposed in MPEG-2 can be easily modified and migrate to accommodate the FGS framework. The objective of this project is to model a MPEG-4 encoder in system level and implement it in software.

Video coding is a data-intensive application. The special feature of the video coding is that it has a constant data flow rate. From the perspective of system modeling, synchronous data flow (SDF) is an ideal modeling tool for this kind of application. Following figure is a high-level SDF model for the encoder with SNR scalability under FGS framework.

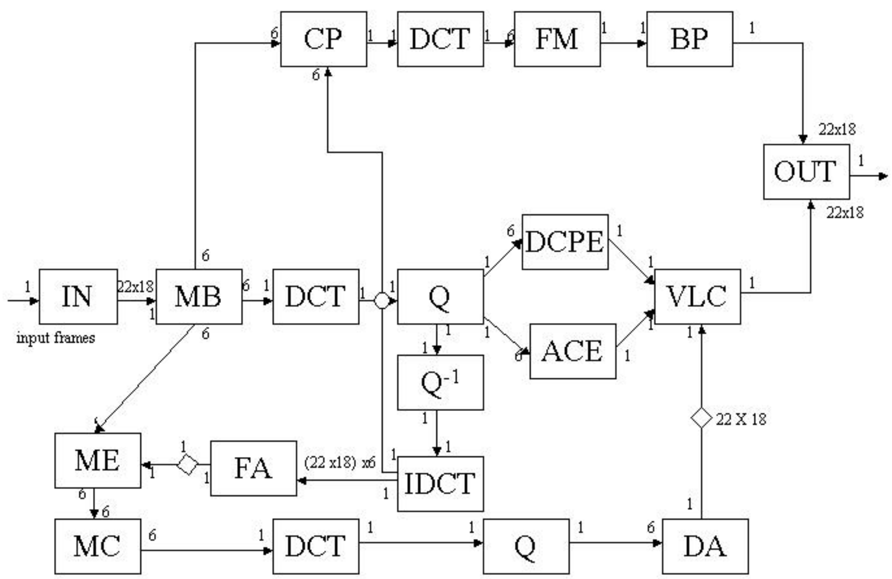


Figure 4. High-level system model for MPEG-4 encoder

The target video quality is QCIF format with resolution of 176 x 144. In the figure, IN refers to input, OUT refers to output, DCT refers to discrete cosine transform (8 x 8), IDCT refers to inverse discrete cosine transform, Q refers to quantization, Q^{-1} refers to inverse quantization, VLC refers to variable length coding, MB refers to separating frame into marcorblocks for DCT processing, CP refers to clipping, FM refers to finding maximum significant bit-plane, BP refers to bit-plane variable length encoding, DCPE refers to predication encoding DC coefficient, ACE refers to encoding AC coefficients, FA refers to frame accumulation for motion estimation, ME refers to motion estimation, MC refers to motion compensation and DA refers to coding of DC/AC coefficients.

References:

[1] D. P. Wu, Y. T. Hou, W. W. Zhu, Y. Q. Zhang, J. M. Peha, “Streaming Video over the Internet: Approaches and Directions”, *IEEE Transactions on Circuit and Systems for Video Technology,* Vol. 11, No. 1, p. 1-20, February 2001

[2]. W. P. Li, “Overview of Fine Granularity Scalability in MPEG-4 Video Standard:, *IEEE Transactions on Circuit and Systems for Video Technology*, Vol. 11, No. 3, p. 310-317, March 2001

[3] H. Radha, M. van der Schaar and Y. W. Chen, “The MPEG-4 Fine-Grained Scalable Video Coding Method for Multimedia Streaming Over IP”, *IEEE Transactions on Multimedia*, Vol. 3, No. 1, p. 53-68, March 2001

[4] H. Sun, W. Kwok, and J. W. Zdepski, “Architectures for MPEG compressed bitstream scaling”, *IEEE Transactions on Circuit and Systems for Video Technology*, Vol. 6, No. 2, p.191-199, April 1996

[5] B. G. Haskell, A. Puri, and A. N. Netravali, “Digital Video: An Introduction to MPEG-2”, *New York: Champion and Hall, ISBN 0-412-08411-2*, 1997

[6] G. J. Conklin, G. S. Greenbaum, K. O. Lillevold, A. F. Lippman, and Y. A. Renzik, “Video Coding for Streaming Media Delivery on the Internet”, *IEEE Transactions on Circuit and Systems for Video Technology*, Vol. 11, No. 3, p.269-281, March 2001

[7] F. Wu, S. Li, and Y.-Q. Zhang, “A framework for efficient progressive fine granularity scalable video coding”, *IEEE Transactions on Circuit and Systems for Video Technology*, Vol. 11, No.3, p. 332–344, Mar. 2001.

[8] F. Wu, S. Li, X. Sun, R. Yan and Y,-Q. Zhang, “ Macroblock-based progressive fine granularity scalable coding”, *ISO/IEC JTC1/SC29/WG11, MPEG2001/N6779, Pisa meeting*, January, 2001.