Scalable and Adaptable Media Coding Techniques for Future Internet

Naeem Ramzan and Ebroul Izquierdo

School of Electronic Engineering and Computer Science, Queen Mary University of London, Mile end, London E1 4NS, United Kingdom
{Naeem.Ramzan, Ebroul.Izquierdo}@elec.qmul.ac.uk

Abstract. High quality multimedia contents can distribute in a flexible, efficient and personalized way through dynamic and heterogeneous environments in Future Internet. Scalable Video Coding (SVC) and Multiple Description Coding (MDC) fulfill these objective thorough P2P distribution techniques. This chapter discusses the SVC and MDC techniques along with the real experience of the authors of SVC/MDC over P2P networks and emphasizes their pertinence in Future Media Internet initiatives in order to decipher potential challenges.

Keywords: Scalable video coding, multiple description coding, P2P distribution.

1 Introduction

Future Media Internet will entail to distribute and dispense high quality multimedia contents in an efficient, supple and personalized way through dynamic and heterogeneous environments. Multimedia content over internet are becoming a well-liked application due to users' growing demand of multimedia content and extraordinary growth of network technologies. A broad assortment of such applications can be found in these days, e.g. as video streaming, video conferencing, surveillance, broadcast, e-learning and storage. In particular for video streaming, over the Internet are becoming popular due to the widespread deployment of broadband access. In customary video streaming techniques the client-server model and the usage of Content Distribution Networks (CDN) along with IP multicast were the most desirable solutions to support media streaming over internet. However, the conventional client/server architecture severely limits the number of simultaneous users for bandwidth intensive video streaming, due to a bandwidth bottleneck at the server side from which all users request the content. In contrast, Peer-to-Peer (P2P) media streaming protocols, motivated by the great success of file sharing applications, have attracted a lot of interest in academic and industrial environments. With respect to conventional approaches, a major advantage in using P2P is that each peer involved in a content delivery contributes with its own resources to the streaming session. However, to provide high quality of service, the video coding/transmission technology needs to be able to cope with varying bandwidth capacities inherent to P2P systems and end-user characteristics

J. Domingue et al. (Eds.): Future Internet Assembly, LNCS 6656, pp. 381–389, 2011.

such as decoding and display capabilities usually tend to be non-homogeneous and dynamic. This means that the content needs to be delivered in different formats simultaneously to different users according to their capabilities and limitations.

In order to handle such obscurity, scalability emerged in the field of video coding in the form of Scalable Video Coding (SVC) [1–4] and Multiple Description Coding (MDC) [5-6]. Both SVC and MDC offers an efficient encoding for applications where content needs to be transmitted to many non-homogeneous clients with different decoding and display capabilities.

Moreover, the bit-rate adaptability inherent in the scalable codec designs provides a natural and efficient way of adaptive content distribution according to changes in network conditions.

In general, a SVC sequence can be adapted in three dimensions, namely, temporal, spatial and quality dimensions, by leaving out parts of the encoded bit-stream, thus reducing the bit-rate and video quality during transmission. By adjusting one or more of the scalability options, the SVC scheme allows flexibility and adaptability of video transmission over resource-constrained networks.

MDC can be considered as an additional way of increasing error resilience and end user adaptation without using intricate error protection methods. The objective of MDC is to generate numerous independent descriptions that can bestow to one or more characteristics of video: spatial or temporal resolution, signal-to-noise ratio. These descriptions can be used to decode the original stream, network congestion or packet loss, which is common in best-effort networks such as the Internet, will not interrupt the reproduction of the stream but will only cause a loss of quality. Descriptions can have the same importance namely “balanced MDC schemes” or they can have different importance “unbalanced MDC schemes”. The more descriptions received, the higher the quality of decoded video. MDC combined with path/server diversity offers robust video delivery over unreliable networks and/or in peer-to-peer streaming over multiple multicast trees.

The eventual objective of employing SVC/MDC in Future Internet is to maximize the end-users' quality of experience (QoE) for the delivered multimedia content by selecting an appropriate combination of the temporal, spatial and quality parameters for each client according to the limitation of network and end user devices .

This chapter starts with an overview of SVC and MDC source coding techniques in section 2 and 3. Section 4 describes how to adapt SVC for P2P distribution for Future Internet. MDC over P2P is explained in section 5. Finally, this chapter concludes in section 6.

2 Scalable Video Coding

During the last decade a noteworthy amount of research has been devoted to scalable video coding with the aspire of developing the technology that would offer a low-complexity video adaptation, but preserve the analogous compression efficiency and decoding complexity to those of conventional (non-scalable) video coding systems. This research evolved from two main branches of conventional video coding: 3D wavelet [1] and hybrid video coding [2] techniques. Although some of the earlier video standards, such as H.262 / MPEG-2 [3], H.263+ and MPEG-4 Part 2 included limited support for scalability, the use of scalability in these solutions came at the significant increase in the decoder complexity and / or loss in coding efficiency. The latest video coding standard, H.264 / MPEG-4 AVC [2] provides a fully scalable extension, SVC, which achieves significant compression gain and complexity reduction when scalability is sought, compared to the previous video coding standards.

The scalability is usually required in three different directions (and their combinations). We define these directions of scalability as follows:

- Temporal scalability refers to the possibility of reducing the temporal resolution of encoded video directly from the compressed bit-stream, i.e. number of frames contained in one second of the video.
- Spatial scalability refers to the possibility of reducing the spatial resolution of the encoded video directly from the compressed bit-stream, i.e. number of pixels per spatial region in a video frame.
- Quality scalability, or commonly called SNR (Signal-to-Noise-Ratio) scalability, or fidelity scalability, refers to the possibility of reducing the quality of the encoded video. This is achieved by extraction and decoding of coarsely quantised pixels from the compressed bit-stream.

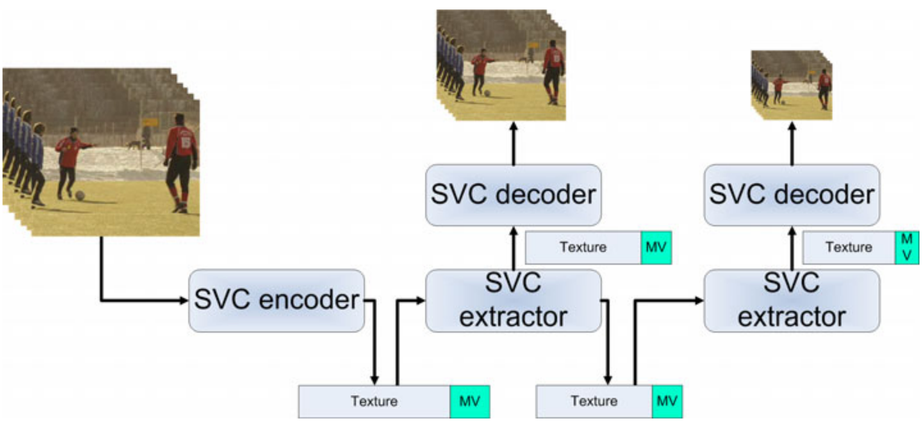

Fig. 1. A typical scalable video coding chain and types of scalabilities by going to lower-rate decoding

An example of basic scalabilities is illustrated in Figure 1, which shows a typical SVC encoding, extraction and decoding chain. The video is encoded at the highest spatio-temporal resolution and quality. After encoding, the video is organised into a scalable bit-stream and the associated bit-stream description is created. This description indicates positions of bit-stream portions that represent various spatio-temporal resolutions and qualities. The encoder is the most complex between the three modules. The compressed video is adapted to a lower spatio-temporal resolution and / or quality by the extractor. The extractor simply parses the bit-stream and decides which portions of the bit-stream to keep and which to discard, according to the input adaptation parameters. An adapted bit-stream is also scalable and thus it can be fed into the extractor again, if further adaptation is required. The extractor represents the least complex part of the chain, as its only role is to provide low-complexity content adaptation without transcoding. Finally, an adapted bit-stream is sent to the decoder, which is capable of decoding any adapted scalable video bit-stream.

2.1 H.264/MPEG-4 SVC

The latest H.264/MPEG-4 AVC standard provides a fully scalable extension, H.264/MPEG-4 SVC, which achieves significant compression gain and complexity reduction when scalability is sought, compared to the previous video coding standards [4]. According to evaluations done by MPEG, SVC based on H.264/MPEG-4 AVC provided significantly better subjective quality than alternative scalable technologies at the time of standardisation. H.264/MPEG-4 SVC reuses the key features of H.264/AVC and also employs some other new techniques to provide scalability and to improve coding efficiency. It provides temporal, spatial and quality scalability with a low increase of bit-rate relative to the single layer H.264/MPEG-4 AVC.

The scalable bit-stream is structured into a base layer and one or several enhancement layers. Temporal scalability can be activated by using hierarchical prediction structures. Spatial scalability is obtained using the multi-layer coding approach. Within each spatial layer, single-layer coding techniques are employed. Moreover, inter-layer prediction mechanisms are utilized to further improve the coding efficiency. Quality scalability is provided using the coarse-grain quality scalability (CGS) and medium-grain quality scalability (MGS). CGS is achieved by requantization of the residual signal in the enhancement layer, while MGS is enabled by distributing the transform coefficients of a slice into different network abstraction layer (NAL) units. All these three scalabilities can be combined into one scalable bit-stream that allows for extraction of different operation points of the video.

2.2 Wavelet-Based SVC (W-SVC)

A enormous amount of research activities are being continued on W-SVC although H.264/MPEG-4 SVC was chosen for standardisation. In the W-SVC, the input video is subjected to a spatio-temporal (ST) decomposition based on a wavelet transform. The rationale of the decomposition is to decorrelate the input video content and offer the basis for spatial and temporal scalability. The ST decomposition results in two distinctive types of data: wavelet coefficients representing of the texture information remaining after the wavelet transform and motion information obtained from motion estimation (ME), which describes spatial displacements between blocks in neighbouring frames. Although the wavelet transform generally performs very well in the task of video content decorrelation, some amount of redundancies still remains between the wavelet coefficients after the decomposition. Moreover, a strong correlation also exists between motion vectors. For these reasons, further compression of the texture and motion vectors is performed. Texture coding is performed in conjunction with so-called embedded quantisation (bit-plane coding) in order to provide the basis for quality scalability. Finally, the resulting data are mapped into the scalable stream in the bit-stream organisation module, which creates a layered representation of the compressed data. This layered representation provides the basis for low-complexity adaptation of the compressed bit-steam.

3 Scalable Multiple Description Coding (SMDC)

SMDC is a source coding technique, which encodes a video into a N (where N≥2) independent decodable sub-bitstreams by exploiting the scalability features of SVC. Each sub-bitstream is called "description". The descriptions can be transmitted through different or independent network paths to reach a destination. The receiver can create a reproduction of the video when at least one of the descriptions is received. The quality of the received video is proportional to the number of descriptions received on time. Thus, the more descriptions are received, the better quality of reconstruction is.

General principles and different approaches for MDC are reviewed in [5]. Approaches for generating multiple descriptions include data partitioning (e.g., even/odd sample or DCT coefficient splitting) [5], multiple description (MD) quantization (e.g.,MD scalar or vector quantization) [6], and multiple description transform coding (e.g., pairwise correlating transform). Another MD encoding approach is based on combinations of video segments (codeblocks, frames, orgroup of pictures) encoded at high and low rates. Different combinations of codeblocks coded at high and low rates have been introduced in [5] for wavelet-based flexible MD encoders.

4 Scalable Video over P2P Network

The proposed system is based on two main modules: scalable video coding and the P2P architecture. In this system, we assume that each peer contains the scalable video coder and the proposed policy of receiving chunk is to make sure that each peer at least receives the base layer of the scalable bit-stream for each group of picture (GOP). Under these circumstances, peers could download different layers from different users, as shown in Figure 2.

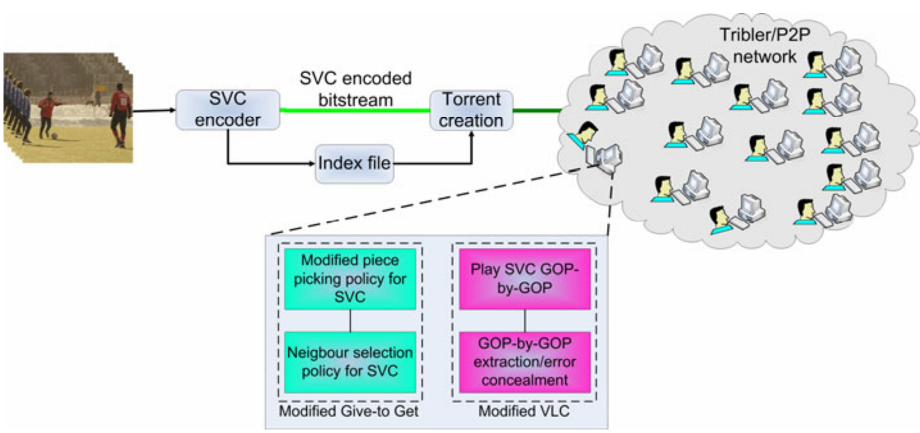

Fig. 2. An example of the proposed system for scalable video coding in P2P network.

In this section, we formulate how the scalable layers are prioritized in our proposed system. First we explain how the video segments or chunks are arranged and prioritized in our proposed system

4.1 Piece Picking Policy

The proposed solution is a variation of the "Give-To-Get" algorithm [8], already implemented in Tribler. Modifications concern the piece picking and neighbour selection policies. Scalable video sequences can be split into GOPs and layers [7] while BitTorrent splits files into pieces. Since there is no correlation between these two divisions, some information is required to map GOPs and layers into pieces and vice versa. This information can be stored inside an index file, which should be transmitted together with the video sequence. Therefore, the first step consists of creating a new torrent that contains both files. It is clear that the index file should have the highest priority and therefore should be downloaded first.

Once the index file is completed, it is opened and information about offsets of different GOPs and layers in the video sequence is extracted. At this point, it is possible to define a sliding window, made of W GOPs and the pre-buffering phase starts. Pieces can only be picked among those inside the window, unless all of them have

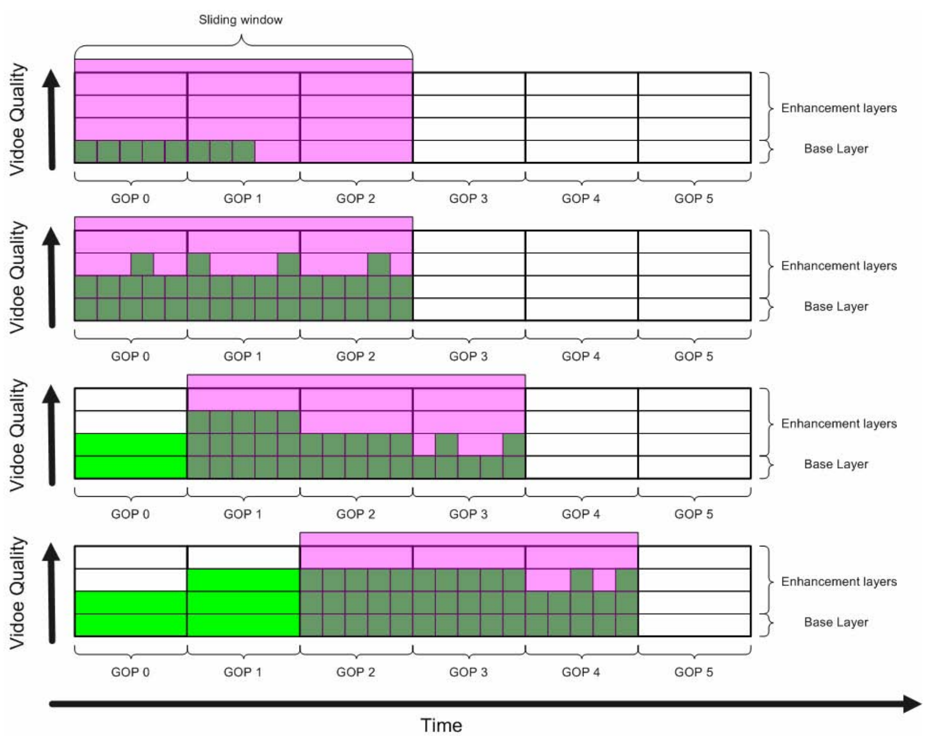

Fig. 3. Sliding window for scalable video

already been downloaded. In the latter case, the piece picking policy will be the same as the original BitTorrent, which is rarest [piece] first.

Inside the window, pieces have different priorities as shown in Figure 3. First of all, a peer will try to download the base layer, then the first enhancement layer and so on. Pieces from the base layer are downloaded in a sequential order, while all the other pieces are downloaded rarest-first (within the same layer).
The window shifts every t(GOP) seconds, where t(GOP) represents the duration of a GOP. The only exception is given by the first shift, which is performed after the pre-buffering, which lasts W * t(GOP) seconds.

Every time the window shifts, two operations are made. First, downloaded pieces are checked, in order to evaluate which layers have been completely downloaded. Second, all pending requests that concern pieces belonging to a GOP that lies before the window are dropped. An important remark is that the window only shifts if at least the base layer has been received, otherwise the system will auto-pause. The detail of modified piece picking policy can be found in [7]. Another issue is the wise choice of the neighbours.

4.2 Neighbour Selection Policy

It is extremely important that at least the base layer of each GOP is received before the window shifts. Occasionally, slow peers in the swarm (or slow neighbours) might delay the receiving of a BT piece, even if the overall download bandwidth is high. This problem is critical if the requested piece belongs to the base layer, as it might force the playback to pause. Therefore, these pieces should be requested from good neighbours. Good neighbours are those peers that own the piece with the highest download rates, which alone could provide the current peer with a transfer rate that is above a certain threshold. During the pre-buffering phase, any piece can be requested from any peer. However, every time the window shifts, the current download rates of all the neighbours are evaluated and the peers are sorted in descending order.

The performance of the proposed framework [7] has been evaluated to transmit wavelet-based SVC encoded video over P2P network as shown in Figure 4.

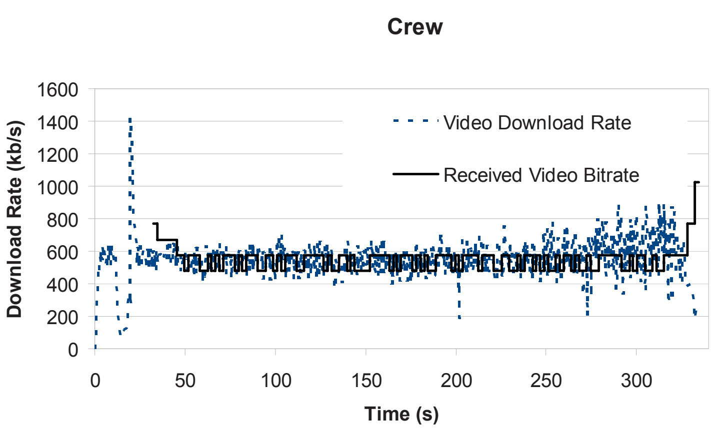

Fig. 4. Received download rate and received video bitrate for Crew CIF sequence

5 Multiple Description Coding over P2P Network

Most of the work on MDC is proposed for wireless applications in which there are issues such as hand-over of a client to another wireless source is present. However, in IP networks, it may be more complicated to have autonomous links among peers. Thus, additional redundancy introduced by using MDC over internet need to be carefully evaluated.

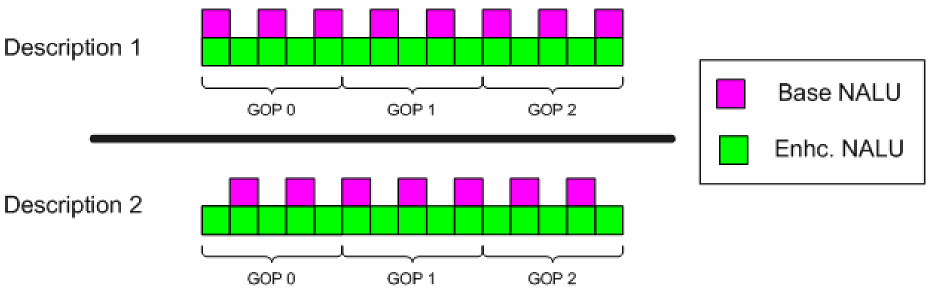

Fig. 5. An example of multiple description using scalable video coding

A simple way to generate multiple descriptions using scalable video coding is to distribute the enhancement layer NAL units to separate descriptions. Thus when both descriptions is received it is possible to have all the enhancement NAL units. Moreover, it is possible to control the redundancy by changing the quality of the base layer as shown in Figure 5.

5.1 Piece Picking Policy

Similar to SVC P2P technique described above, the sliding window is defined for Scalable MDC over P2P. The highest priority is given to the pieces of each description inside the sliding window however the further classification can be enabled with respect to playing. The chunks belong to nearer to the playing time gives higher priority and declared as critical in the sliding window and vice versa as shown in Figure 6. This enhances the overall performance of the system. The main advantage of Scalable

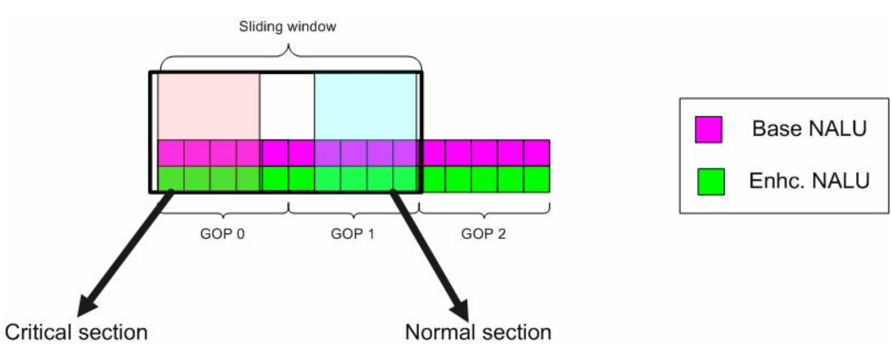

Fig. 6. Sliding window for multiple description of scalable video.

MDC over SVC is that the receiver/client can make a reproduction of the video when any of the description is received at the cost of additional redundancy due to the presence of base layer in each description.

6 Conclusions

This chapter has presented an overview of SVC and MDC with the perspective of content distribution over Future Internet. These coding schemes provide natural robustness and scalability to media streaming over heterogeneous networks. The amalgamation of SVC/MDC and P2P are likely to accomplish some of the Future Media Internet challenges. Tangibly, SVC/MDC over P2P presumes an excellent approach to facilitate future media applications and services, functioning under assorted and vibrant environments while maximizing not only Quality of Service (QoS) but also Quality of Experience (QoE) of the users. At last, we persuade Future Internet initiatives to take into contemplation these techniques when defining new protocols for ground-breaking services and applications.

Acknowledgement. This research has been partially funded by the European Commission under contract FP7-248474 SARACEN.

References

1. Mrak, M., Sprljan, N., Zgaljic, T., Ramzan, N., Wan, S., Izquierdo, E.: Performance evidence of software proposal for Wavelet Video Coding Exploration group, ISO/IEC JTC1/SC29/WG11/ MPEG2006/M13146, 76th MPEG Meeting, Montreux, Switzerland (April 2006)
2. ITU-T and ISO/IEC JTC 1: Advanced video coding for generic audiovisual services, ITU-T Recommendation H.264 and ISO/IEC 14496-10 (MPEG-4 AVC)
3. ITU-T and ISO/IEC JTC 1: Generic coding of moving pictures and associated audio information – Part 2: Video, ITU-T Recommendation H.262 and ISO/IEC 13818-2 (MPEG-2 Video)
4. Schwarz, H., Marpe, D., Wiegand, T.: Overview of the scalable video coding extension of the H.264 / AVC standard. IEEE Transactions on Circuits and Systems for Video Technology 17(9), 1103–1120 (2007)
5. Berkin Abanoz, T., Murat Tekalp, A.: SVC-based scalable multiple description video coding and optimization of encoding configuration. Signal Processing: Image Communication 24(9), 691–701 (2009)
6. Tillo, T., Grangetto, M., Olmo, G.: Redundant slice optimal allocation for H.264 multiple description coding. IEEE Trans. Circuits Syst. Video Technol. 18(1), 59–70 (2008)
7. Asioli, S., Ramzan, N., Izquierdo, E.: A Novel Technique for Efficient Peer-To-Peer Scalable Video Transmission. In: Proc. of European Signal Processing Conference (EUSIPCO-2010), Aalborg, Denmark, August 23-27 (2010)
8. Pouwelse, J.A., Garbacki, P., Wang, J., Bakker, A., Yang, J., Iosup, A., Epema, D.H.J., Reinders, M., van Steen, M.R., Sips, H.J.: Tribler: A social-based based peer to peer system. In: 5th Int'l Workshop on Peer-to-Peer Systems, IPTPS (Feb. 2006)