Video Standard Compliant Layered P2P Streaming

Man Yau Chiu[1], Kangheng Wu[1], Zhibin Lei[1] and Dah Ming Chiu[2]

Abstract. Peer-to-peer (P2P) streaming over Internet needs to consider the dynamic bandwidth fluctuations and various client device capabilities. To maximize the network utilization and to achieve stable P2P streaming, we propose a novel multi-layered P2P streaming system in which the scalable video bit-stream is fully decodable by legacy non-scalable video decoder. So, even the client device only has the legacy video decoding capability, e.g. H.264/AVC, once it joins the proposed layered P2P streaming, this client has the ability to decode the scalable video bit-stream fully to render video with high spatial resolution or quality and at the same time achieve stable and robust quality of experience with quick start up, fast channel zapping and reduced player freezing/buffering.

Keywords: P2P streaming · Spatial scalable coding · Layered coding · Dynamic bandwidth · Three-screen applications · Networking

1 Introduction

Single layer video streaming cannot satisfy heterogeneous customer requirements and heterogeneous download capacities [4]. There are existing systems that use multiple versions of video content (each encoded at different resolution or visual quality) to minimize the overall transmission costs. For example, lower resolution video can be sent to mobile devices while higher resolution or high quality video is sent to PC clients or set-top box (STB) receivers. However, in the P2P network, if there are too many independent video data transmissions, the users' inbound and outbound bandwidth will not be efficiently used. Peers in different versions will not help each other. The overall video quality received will not be optimal.

[1] Man Yau Chiu, Kangheng Wu, Zhibin Lei (✉)
Hong Kong Applied Science and Technology Research Institute, Hong Kong Science Park, Shatin, Hong Kong
e-mail: {edmondchiu, khwu, lei}@astri.org

[2] Dah Ming Chiu (✉)
The Chinese University of Hong Kong, Shatin, Hong Kong
e-mail: dmchiu@ie.cuhk.edu.hk

For layered video encoding, video sequence is encoded into a base layer bit-stream and one (or more) enhancement layer bit-stream(s). The base layer contains coded video of lower spatial resolution (for spatial scalability) or lower quality (for quality scalability). Additional enhancement layers contain data that can progressively refine the reconstructed video spatial resolution or quality. According to the available bandwidth, participating nodes subscribe a subset of the layers to reconstruct the video with certain spatial resolution or quality degradation [12].

Layered video has the advantage of bandwidth efficiency and at the same time meets the real-time streaming requirement of peer clients with wide range of variation in processing power, display capability and network conditions. In other words, although heterogeneity exists for peers in the P2P network, an optimal viewing experience can be achieved for each peer based on its own access bandwidth and capabilities.

Current layered P2P streaming literature has not addressed the requirement that the scalable video bit-stream (consists of both base layer and enhancement layers) should be fully decodable by legacy non-scalable video decoder, e.g. H.264/AVC [3, 10]. Most of these prior publications simply assume that such scalable bit-streams exist already and is generated by some international video coding standards, such as MPEG-2 Spatial Scalable Profile [2], Scalable Video Coding (SVC) extension of H.264/AVC [7, 8] etc. However, this requires all the peers should have corresponding scalable profile decoding capability, e.g. SVC, in order to fully decode the scalable bit-stream. This is not a practical requirement for the current deployed nodes or devices that have legacy non-scalable video decoding capability only, as many client devices (e.g. STB receivers or Internet Protocol Television, IPTV) may have its video decoder hard-wired inside the chip and cannot be changed at all. As a result, such devices can decode the base layer bit-stream only and hence only base layer video can be obtained.

Hence, it is desirable that the scalable video bit-stream, including both base and enhancement layer bit-streams, is fully decodable by legacy non-scalable video decoders that have been widely deployed today. In doing so, even the client device has non-scalable video decoding capability, e.g. with H.264/AVC compliant decoder, it can decode the scalable video bit-stream fully and take advantage of the layered P2P video streaming too.
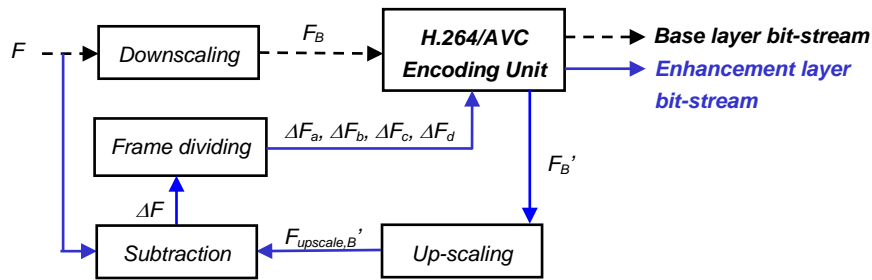
This paper is organized as follows. At first, our proposed layered video encoder/decoder (codec) is described in section 2. Here, our proposed layered encoder can generate the scalable video bit-stream which can be fully decoded by non-scalable video decoder. Furthermore, performance comparison of our proposed layered codec with some standard scenarios is given in this section. Next, the layered P2P streaming architecture and performance metrics are given in sections 3 and 4 respectively. Finally, a conclusion is given in section 5.

2 Video Standard Compliant Multi-Layer Codec

2.1 Proposed Layered Coding Method

With proper video pre- and post-processing techniques, our proposed scalable video coding method provides spatial scalability at the receiving clients even such clients have the non-scalable video decoding capability only, e.g. H.264/AVC.

Scalable Video Encoder

Scalable Video Decoder

- - -▶ Base layer encode / decode path

──▶ Enhancement layer encode / decode path

F: Source video

F_B: Downscaled source video

F_B': Decoded base video

$F_{upscale,B}'$: Up-scaled decoded base video

ΔF: Residual frame

$\Delta F_a, \Delta F_b, \Delta F_c, \Delta F_d$: Residual sub-frame

$\Delta F'$: Re-constructed residual frame

$\Delta F_a', \Delta F_b', \Delta F_c', \Delta F_d'$: Decoded residual sub-frames
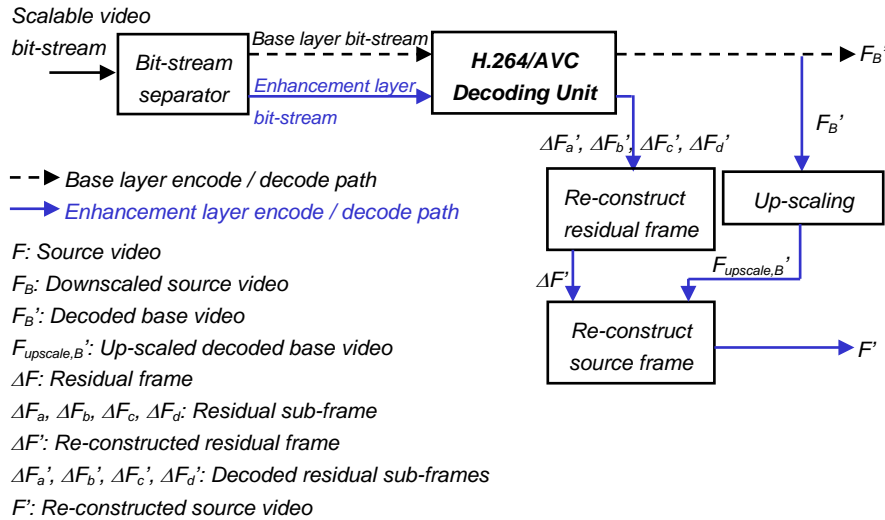
F': Re-constructed source video

Fig. 1 Proposed video standard compliant layered codec

To simplify our illustration, we assume that our layered encoder is going to code the source video into two-layer representation. Here, the base layer video represents coded video at lower resolution (2:1 spatially down-scaled, both horizontally and vertically, of the source video resolution) and the enhancement layer

video represents coded video at source video resolution. Without loss of generosity, this codec structure can be extended to code video using more than two layers as long as the resolution of the enhancement layer video, both horizontally and vertically, is integral multiples of the base layer video resolution.

Fig. 1 shows the structure of the proposed video standard compliant layered codec to code source video into 2-layer representation. At the encoding side, the encoder down-scales the original source video (F) into lower resolution video (F_B) and then encode the lower resolution video by legacy video encoding unit, e.g. H.264/AVC, to generate base layer bit-stream. Next, the encoder generates the residual frames (ΔF) as the difference between the original source video frame (F) and the up-scaled decoded base layer video frame ($F_{upscale,B}$'). The residual frames, ΔF, are then divided into non-overlapping residual sub-frames (ΔF_a, ΔF_b, ΔF_c, ΔF_d) and the resolution of these residual sub-frames is the same as that of base layer video frames. Finally, these residual sub-frames, ΔF_a, ΔF_b, ΔF_c, ΔF_d, are encoded by the same legacy video encoding unit to generate the enhancement layer bit-stream.

At the decoding side, the base layer bit-stream is decoded by legacy video decoding unit, e.g. H.264/AVC, to obtain base layer video, F_B'. The decoded base layer video is up-scaled to original source resolution. For the enhancement layer bit-stream, it can be decoded by the same legacy decoding unit to obtain residual sub-frames, ΔF_a', ΔF_b', ΔF_c', ΔF_d', and these residual sub-frames can be combined to re-construct residual frames, ΔF'. By combining the up-scaled base layer video with the residual frames, the original source video can be re-constructed.

2.2 Layered Codec Performance Evaluation

The performance of the proposed layered codec is evaluated by comparing its rate-distortion performance with single layer H.264/AVC encoding (single layer), SVC encoding as well as simulcast scenarios. Performance evaluation is conducted in the ways similar to the method as described in [8, 11]. Here, the video encoding and decoding units used in our proposed layered codec are compliant to H.264/AVC.

Totally, four 4CIF (704x576) video sequences are tested and they are City, Crew, Harbour and Soccer. These video sequences are encoded into two-layer representation in which the base layer represents the CIF (352x288) resolution while the enhancement layer represents the 4CIF resolution. Rate points for the test sequences are defined as shown in Table 1. All the video sequences are coded in IPPP..IPPP… structure (i.e. without B-frames for zero-coding delay) and interval between I-frames is 16. For single layer, SVC and simulcast simulations results, they are all generated by H.264/SVC reference software, JSVM 9.19.15 [1].

Fig. 2 shows the rate-distortion plots for the City and Harbour sequences. Distortion values in the plots are the PSNR values for enhancement layer video, while

the bit-rate represents the total bit-rates of the scalable bit-stream. For conducting performance evaluation, firstly, each video sequence is encoded by SVC encoder at each rate points to obtain SVC results in the plots. Next, the proposed layered encoder is used to encode the video sequences to achieve base and enhancement layer PSNR values close to that of SVC. This can be done by adjusting the layered encoder target base and enhancement layer bit rates. For single-layer encoding, it refers to the coding of video sequences at 4CIF resolution only by H.264/AVC while for simulcast, it refers to the coding of each sequence at CIF and 4CIF resolutions separately by H.264/AVC. For single-layer and simulcast encoding, the coded CIF and 4CIF video should have the PSNR values close to that of SVC.

Table 1 Tested bitrates for spatial scalability test

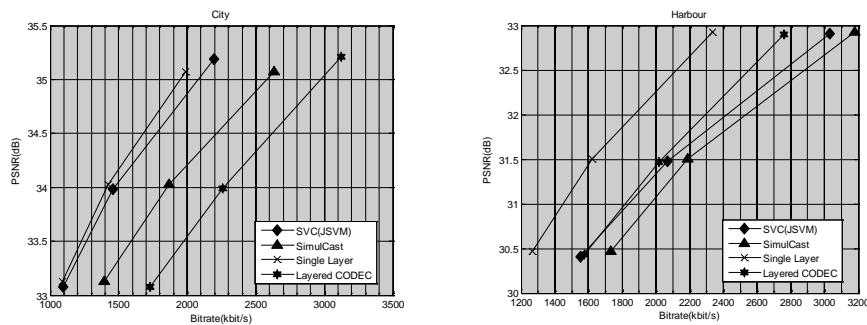| Sequence	Format	Bit rates (kbit/sec) Rate point index:		
		1	2	3
City	CIF 30Hz	256	384	512
	4CIF 30Hz	1024	1536	2048
Crew, Harbour, Soccer	CIF 30Hz	384	512	768
	4CIF 30Hz	1536	2048	3072

Fig. 2 Rate-distortion plots for City, and Harbour sequences. The SVC, simulcast and single layer results are all generated by the JSVM software

Table 2 Comparison of the proposed layered codec to simulcast and SVC scenarios

Sequence	Simulcast		SVC	
	ΔBit-rate	ΔPSNR	ΔBit-rate	ΔPSNR
City	21.1%	0.02	51.5%	0.01
Crew	-10.7%	0.04	-15.6%	0.01
Harbour	-10.0%	-0.03	-3.5%	0.02
Soccer	-10.2%	0.05	0.46%	0.05

Furthermore, the ΔBit-rate and the ΔPSNR in which the proposed layered codec as compared to simulcast and the proposed layered codec as compared to SVC for all sequences are given in Table 2. For ΔBit-rate, negative value indicates that the proposed layered codec requires lower bit-rate than the scenario under com-

parison. For ΔPSNR, its values are all close to zero indicates that the 4CIF resolution videos are all coded at similar quality under all scenarios.

From these figures and tables, we observe that in general our proposed layered codec can achieve better performance than simulcast encoding except for City sequence. As compared to SVC, the proposed layered codec also has better performance in encoding some sequences with moderate motion and edge details, e.g. Crew and Harbour sequences. If excluding City sequence results, our layered codec out-performs the simulcast encoding by an average of 10.3% and out-performs SVC by an average 6.2% in terms of bit-rate.

Here, the poor performance of the proposed layered codec in coding City sequence would be because there are many fine edge details within each frame and at the same time, there is a continuous panning motion between frames. As a result, the residual frames, ΔF, generated during layered coding contain many high frequency details and frames with such high frequency characteristic are not efficiently encoded by H.264/AVC encoding unit.

3 Layered P2P Architecture

Layered P2P architecture is an extension of the single layer Peer-to-Peer Streaming Protocol (PPSP) with modification of adding the layer information to meet the layered streaming requirement [5].

At *peer side*, different from the single layer P2P streaming, peer and tracker need to keep the layer information. The following components are considered:

- Transmission of PUT-LAYER messages, by which source tells trackers the layer information the streaming is using.
- Transmission of GET-LAYER messages, by which peers request what they want and get the layer information from trackers.
- Transmission of PUT-CHUNK messages, by which peers tell trackers what they have. The bitfield can represent chunk $_i$.
- Transmission of GET-PEERLIST messages, by which peers request what they want and get candidate peers list from trackers.
- Transmission of STATISTICS requests and responses, by which trackers can get peers status, network performance, layer $_i$, etc.

At *tracker side*, the peer-tracker protocol should allow a peer to indicate which layers are of interest. The tracker will then take that into consideration in returning peer-list. Optionally, the peers in peer-list may come with layer information. The management/status reports from peer to tracker should also indicate layer information about the local peer.

For *scheduling algorithm*, each peer doing layered P2P has a more complicated scheduling module, figuring out which chunks in which layer have higher priority to get. The peer-to-peer protocol for signalling (exchanging bitmap information) needs to have an extension to describe layered bitmaps and related data structures.

The peer-to-peer data plane protocol can be almost the same as the single-layer case; the only extension is to add indicator for which layer the requested chunk belongs to. In [6], a simple greedy data scheduling strategy is proposed. Each peer requests lower layers from lower bandwidth neighbors and higher layers from high bandwidth neighbors. This scheduling strategy is proven to be optimal for a given single peer within a certain time slot but may not be globally optimal. Different coding scheme has different optimal data scheduling strategy. Our approach is actually standard neutral, as long as it is compliant with media codec standard (e.g. H.264/AVC, MPEG-4, On2, wmv, real etc) at each and every layer to avoid a requirement for non-standard compliant decoder in the peer player. Our algorithm [9] consists of two parts: 1) layer adaptation, where peers adaptively adjust the number of subscribed layers to ensure a smooth playback of the highest possible video quality; and 2) piece selection, in which a peer selects a missing data piece to request based on its utility.

4 Performance Metrics

Although the layered encoding scheme brings more flexibility for participating peers to achieve adaptive video quality, it also causes challenges to the P2P protocol for layer streaming. In [12], the following four performance metrics are mentioned.

- Throughput and Delay: layered P2P shall maximize overall throughput and keep low packet delay.
- Layer Delivery Ratio: in multiple layered P2P streaming, subscribing many layers but with low delivery ratio for each layer can result in high throughput. But, the video quality is poor because of the layer dependency. Hence high delivery ratio for subscribed layers is important.
- Useless Packets Ratio: if lower layer packets are missed, related upper layer packets become useless. Useless packet ratio should be kept low.
- Jitter Prevention: since Internet is dynamic, if the node subscribes more layers after bandwidth increased, it may have to drop some layers if later bandwidth decreases. This short-term subscribe-drop pair, or jitter, brings fluctuation in quality of service and causes its buffer overflow or underflow.

In addition to individual peer level, layered P2P can help enhance overall system performance by utilizing the heterogeneous conditions of peers. For example, it can reduce the start-up delay by serving base layer first while enhancement layers may be delivered later. Several quality metrics may be considered for live streaming and progressing download.

- Start-up Delay: the duration between a peer makes a request for a stream and the stream actually plays.

- Playback Continuity: the percentage of the playing streaming successfully played at the correct time.
- Playback Delay: the delay between a video data which is generated by the source and it is being viewed by the peer.

5 Conclusion

In this paper, we have proposed a layered P2P streaming system in which the transmitted scalable video bit-stream is fully decodable by client with legacy non-scalable decoding capability, e.g. H.264/AVC. With this special video bit-stream property, spatial scalability and other advantages of layered P2P streaming can be easily achieved by those devices with legacy decoding capability in which these devices would have been widely deployed today.

References

1. H.264/SVC reference software JSVM.
   http://www.hhi.fraunhofer.de/de/kompetenzfelder/image-processing/research-groups/image-video-coding/svc-extension-of-h264avc/jsvm-reference-software.html. Accessed 9 July 2013;
2. ISO/IEC (1996). Information technology—Generic coding of moving pictures and associated audio information: Video. ISO/IEC Standard 13818-2.
3. ISO/IEC (2010). Information technology—Coding of audio-visual objects – Part 10: Advanced Video Coding. ISO/IEC Standard 14496-10.
4. Liu, Z., Shen, Y., Ross, K.W., Panwar, S., Wang, Y. (2009). LayerP2P: Using Layered Video Chunks in P2P Live Streaming. IEEE Transactions on Multimedia.
5. Peer to Peer Streaming Protocol (IETF). http://datatracker.ietf.org/wg/ppsp/charter/. Accessed 9 July 2013;
6. Rejaie, R., & Ortega, A. (2003). PALS: Peer-to-Peer Adaptive Layered Streaming. NOSSDAV.
7. Schwarz, H., Marpe, D., & Wiegand, T. (2007). Overview of the Scalable Video Coding Extension of the H.264/AVC Standard. IEEE Transactions on Circuits and Systems for Video Technology, vol. 17, no. 9, pp. 1103 - 1120.
8. Segall, C. A., & Sullivan, G. J. (2007). Spatial Scalability Within the H.264/AVC Scalable Video Coding Extension. IEEE Transactions on Circuits and Systems for Video Technology, vol. 17, no. 9, pp 1121 - 1135.
9. Wen, Z. (2013). Algorithm design of layered peer-to-peer video on demand streaming networks. http://hub.hku.hk/handle/10722/184256. Accessed 9 July 2013;
10. Wiegand, T., Sullivan, G. J., Bjontegaard, G., & Luthra, A. (2003). Overview of the H.264/AVC Video Coding Standard. IEEE Transactions on Circuits and Systems for Video Technology, vol. 13, no. 7, pp 560 - 576.
11. Wien, M., Schwarz, H., & Oelbaum, T. (2007). Performance Analysis of SVC. IEEE Transactions on Circuits and Systems for Video Technology, vol. 17, no. 9, pp. 1194 - 1203.
12. Xiao, X., Shi, Y., Gao, Y., & Zhang, Q. (2009). LayeredP2P: A New Data Scheduling Approach for Layered Streaming in Heterogeneous Networks. Infocom.