

Video Splicing for Tune-in Time Reduction in IP Datacasting over DVB-H

Miska M. Hannuksela

Nokia Research Center
Finland
Miska.Hannuksela@nokia.com

Mehdi Rezaei

Tampere University of Technology
Finland
Mehdi.Rezaei@tut.fi

Moncef Gabbouj

Tampere University of Technology
Finland
Moncef.Gabbouj@tut.fi

ABSTRACT

A novel video splicing method is proposed which minimizes the tune-in time of “channel zapping”, i.e. changing from one audiovisual service to another, in IP datacasting (IPDC) over Digital Video Broadcasting for Handheld terminals (DVB-H). DVB-H uses a time-sliced transmission scheme enabling a receiver to turn radio reception off for those time-slices that are not of interest to the user and thus reducing the power consumption used for radio reception. One of the significant factors in tune-in time is the time from the start of media decoding to the start of correct output from decoding, which is minimized when a time-slice starts with a random access point picture such as an independent decoding refresh (IDR) picture in H.264/AVC. In IPDC over DVB-H, encapsulation to time-slices is performed independently from encoding in a network element called IP encapsulator. At the time of encoding, time-slice boundaries are typically not known exactly, and it is therefore impossible to govern the location of IDR pictures relative to time-slices. It is proposed that an additional stream consisting of IDR pictures only is transmitted to the IP encapsulator, which replaces pictures in a normal bit stream with IDR pictures according to time-slice boundaries in order to achieve the minimum tune-in delay. Replacing pictures causes a mismatch in the pixel values of the reference pictures between the encoder and decoder and the mismatch error is propagated in the reconstructed video. It has to be ensured that the propagated error is subjectively negligible. Furthermore, the “spliced” stream resulting from the operation of the IP encapsulator should comply with the Hypothetical Reference Decoder (HRD) specification of H.264/AVC. Error propagation caused by the proposed splicing method is analyzed and a video rate control system is proposed to satisfy the HRD requirements for the spliced stream. Simulation results show that in addition to fulfilling H.264/AVC compliancy, good average quality of decoded video is achieved with minimum tune-in delay and complexity.

1. INTRODUCTION

DVB-H (Digital Video Broadcasting for Handheld terminals) is an ETSI standard specification for bringing broadcast services to battery-powered handheld receivers [1]. DVB-H is largely based on the successful DVB-T specification for digital terrestrial television, adding to it a

number of features designed to take into account the limited battery life of small handheld devices, and the particular environments in which such receivers must operate. The use of time-slicing leads to significant power savings. DVB-H also employs additional forward error correction to further improve mobile and indoor reception performance of DVB-T.

A simplified block diagram of a conventional IPDC system over DVB-H is depicted in Figure 1. As shown, a content encoder receives a source signal in analog format, uncompressed digital format, compressed digital format, or any combination of these formats. Content encoder encodes the source signal into a coded media bit stream. Content encoder may be capable of encoding more than one media type, such as audio and video. Alternatively, more than one content encoder may be required to code different media types of the source signal. Figure 1 illustrates the processing of one coded media bit stream of one media type.

The coded media bit stream is transferred to a server. Examples of the format used in transmission include an elementary self-contained bit stream format, a packet stream format, or one or more coded media bit streams encapsulated into a container file. Content encoder and server may reside on the same physical device or may be included in separate devices. Content encoder and server may operate with live real-time content, in which case the coded media bit stream may not be stored permanently, but rather buffered for small periods of time in content encoder and/or in server to smooth out variations in processing delay, transfer delay, and coded media bit rate. Content encoder may also operate considerably earlier than when the bit stream is transmitted from the server. In such a case, the system may include a content database, which may reside on a separate device or on the same device as content encoder and/or server.

The server may be an IP multicast server using real-time media transport over Real-time Transport Protocol (RTP). The server is configured to encapsulate the coded media bit stream into RTP packets according to an RTP payload format. Although not shown in this Figure, the system may contain more than one server. The server is connected to an IP encapsulator, also referred to as a multi-protocol

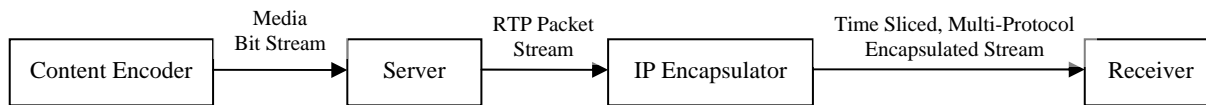


Figure 1: Simplified block diagram of a conventional IP datacasting system

encapsulator. The connection between the server and an IP network may be a fixed-line private network.

The IP encapsulator packetizes IP packets into Multi-Protocol Encapsulation (MPE) Sections which are further encapsulated into MPEG-2 Transport Stream (TS) packets. The IP encapsulator optionally uses MPE Forward Error Correction (MPE-FEC) based on Reed-Solomon (RS) codes. An IPDC system over DVB-H further includes at least one radio transmitter which is not essential for the operation of the proposed splicing system and it is not discussed further.

To reduce the power consumption in handheld terminals, the service data is time-sliced and then it is sent into the channel as bursts at a significantly higher bit rate compared to the bitrate of the audio-visual service. Time-slicing enables a receiver to stay active only a fraction of the time, while receiving bursts of a requested service. Finally, the system includes one or more recipients, capable of receiving, demodulating, decapsulating, decoding, and rendering the transmitted signal, resulting into uncompressed media stream.

Tune-in time or delay in DVB-H refers to the time between the start of the reception of a broadcast signal and the start of the media rendering. The tune-in delay for newly-joined recipients consists of several parts including: delay until the start of the desired time-slice, reception duration of a complete time-slice or MPE-FEC frame, delay to compensate the size variation of MPE-FEC frames, delay to compensate the synchronization between the associated streams (e.g. audio and video) of the streaming session and delay until a media decoder is refreshed by a random access point to produce correct output samples. One of the critical factors in tune-in delay is the time until a media decoder is refreshed to produce correct output frames, which can be minimized if MPE-FEC frame is started with a random access point such as an IDR picture in H.264/AVC. It should be remarked that if the decoder started decoding from an IDR picture that is not at the beginning of a time-slice immediately when the time-slice is received, the input buffer for decoding would drain before the arrival of the next time-slice and there would be a gap in video playback corresponding to the payout duration from the beginning of the time-slice to the first IDR picture.

In IPDC over DVB-H, the content encoding and the encapsulation to MPE-FEC frames are implemented

independently and it is hard to govern the exact location of IDR pictures relative to the boundaries of MPE-FEC frames. Moreover, very frequent IDR pictures in the bit stream drop the compression efficiency remarkably.

We propose a modification in the operation of IPDC system, which minimizes the tune-in delay required for decoder refresh by an IDR picture splicing method to desired locations. When a bit stream is modified, there is a mismatch in the pixel values of the reference pictures between the encoder and decoder and the mismatch propagates in reconstructed video due to predictive video coding known as inter prediction or motion compensation. We investigate the propagated error to verify that it is subjectively negligible. Furthermore, we introduce a method to obtain HRD compliancy of the spliced stream.

Typical intervals between time-slices containing content for a particular audio-visual service may range from one second to a couple of seconds. If IDR pictures are placed randomly and the average IDR picture are interval is about equal to the time-slice interval, the expected tune-in delay due to decoder refresh is approximately half of the time-slice interval, i.e. typically from half a second to few seconds. From the tune-in time reduction point of view, the proposed splicing method typically decreases the decoder refresh time from to zero or very close to zero.

The paper is organized as follows: Section 2 of this paper presents details of proposed splicing method. The propagated error is investigated in section 3 and the HRD compliancy of spliced stream is discussed in section 4. The paper is finalized with conclusions in section 5.

2. PROPOSED SPLICING METHOD

A simplified block diagram of the proposed IPDC system is depicted in Figure 2. At the content encoding level one or two video encoders encode the uncompressed input video to two encoded primary bit streams including a Spliceable Bit Stream (SBS) and a Decoder Refresh Bit Stream (DRBS) from the same source picture sequence. The SBS includes frequent spliceable pictures which are reference pictures constrained as follows: no picture prior to a spliceable picture, in decoding order, is referred to in the inter prediction process of any reference picture at or after the spliceable picture, in decoding order. Non-reference pictures after the spliceable picture may refer to pictures earlier to the spliceable picture in decoding order. These

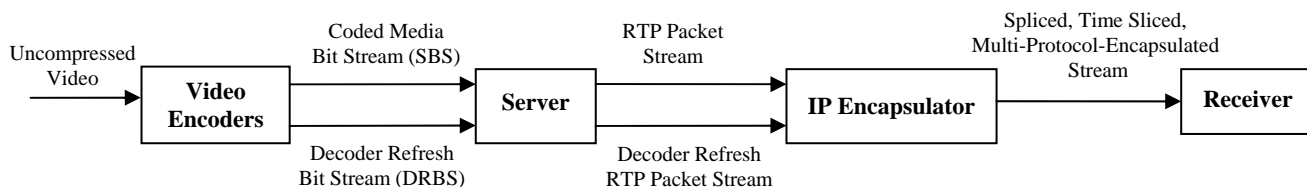


Figure 2: Block diagram of proposed IP datacasting system

non-reference pictures cannot be correctly decoded if the decoding process starts from the spliceable picture, but can be safely omitted as they are not used as reference for any other pictures. The DRBS contains only intra or IDR pictures corresponding to spliceable pictures and with a picture quality similar to corresponding spliceable pictures. The DRBS and the SBS are transmitted from the server to the IP encapsulator. The IP encapsulator composes MPE-FEC frames, in which the first picture in decoding order is an intra/IDR picture from the DRBS and the other pictures are from the SBS. The intra/IDR pictures at the beginning of MPE-FEC frames minimize the tune-in time for newly-joined recipients. No changes in the receiver operation are required in the proposed system.

Replacing an inter picture with an intra picture in the SBS causes a mismatch in the pixel values of the reference pictures between the encoder and decoder. The mismatch propagates until the next IDR picture in the spliced stream. A technically elegant solution would be to use SP and SI pictures of H.264/AVC, but they are only included in the Extended profile of H.264/AVC [2]. The Extended profile of H.264/AVC is not allowed in the current DVB-H standard [3]. Furthermore the HRD compliancy of the spliced stream is hard to verify in the IP encapsulator. Moreover, the initial delay for coded picture buffering in the HRD is hard to derive in the IP encapsulator. More details about these problems are discussed in the sequel.

3. ERROR PROPAGATION IN SPLICED STREAM

According to the proposed splicing method the spliceable pictures in SBS are replaced with corresponding intra/IDR pictures in DRBS. The replaced picture causes a mismatch in the pixel values of the reference pictures between the spliceable and spliced bit stream until the next IDR picture in spliced bit stream. To study the error propagation in spliced bit stream, a simulation was run on different video sequences with various encoding and splicing parameters such as picture size, quantization parameter, number of reference frames, number of skip frames and IDR frequency in spliced stream. In each case we measured the propagated error by several criteria such as PSNR, maximum sample-wise absolute difference, Bjontegaard Delta PSNR and Bjontegaard Delta Rate [5].

The graphs depicted in Figure 3 show an example of simulation results. In this example different bit streams have been compared in terms of PSNR on the luma component while all of them are encoded with a constant quantization parameter of 20. A bit stream including only IDR frames has been compared with spliceable bit stream, normal bit streams and spliced bit streams. Moreover, the PSNR of two spliced bit streams have been depicted in which replacing of spliceable picture with IDR picture started in different locations. As it is expected the PSNR of spliceable bit stream is between those of two normal bit streams with two different number of reference frames 1 and 4. Furthermore, the quality degradation on spliceable bit stream results of constrained referencing in inter prediction for spliceable pictures is insignificant even if the period of spliceable pictures be very short. As example, a period of 5 pictures has been used for the spliceable pictures of SBS depicted in Figure 3.

As an important result, the degradation in quality of spliced streams in comparison to the spliceable bit streams is saturated to a constant value after a small number of frames. This means that the degradation in quality is independent of frequency of IDR frames in spliced bit stream. Hence, using just one IDR picture in each

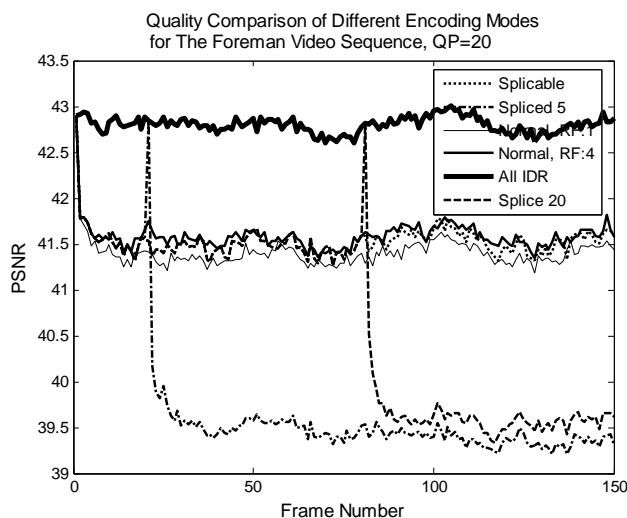


Figure 3: Quality comparison of different bit streams

MPE-FEC frame independently of MPE-FEC frame size can minimize the tune-in time and bitrate. In other words it is possible to have very large MPE-FEC frames each including only one intra/IDR picture with minimum tune-in time at the expense of a constant degradation in the quality of decoded video.

Moreover, as graphs show that the quality degradation of spliced video relative to the normal video is comparable with the quality degradation of inter frames relative to their previous intra/IDR frame while they are encoded with a constant quantization parameter.

The graphs depicted in Figure 4 show the Bjontegaard Delta PSNR between SBS and spliced bit streams including one spliced IDR picture for a number of video sequences each including 150 frames. In these graphs the distance of spliced IDR to the next IDR picture or end of sequence is considered as variable. Almost linear curves prove that the average quality degradation in each sequence is proportional to the number of video pictures which are affected by the propagated error. While there is no quality degradation before the spliced picture, the almost linear graphs means that the propagation error is saturated to a constant value. In other words the error level for all pictures after spliced picture is almost similar in each sequence and there is no accumulation error. The average quality degradation (on different video sequence) between SBS and spliced bit stream is about 1.58 dB.

In another simulation we encoded spliceable pictures and corresponding IDR pictures with a higher quality than other pictures. Moreover, a smaller quantization parameter was used for spliceable pictures than corresponding IDR pictures to get a similar quality. In this case, the average degradation in quality decreased to 1.06 dB.

While the error propagation degrades the quality of reconstructed video by decoding spliced stream in

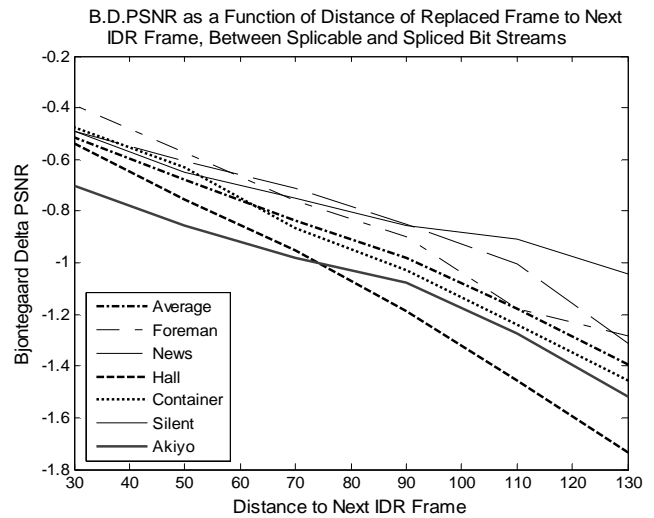


Figure 4: Bjontegaard Delta PSNR between spliceable and spliced bit streams with one spliced IDR picture

comparison to SBS, the quality of decoded SBS is lower than the quality of a fully rate-distortion optimized bit stream due to the proposed constraint on the reference frames for spliceable pictures. Depending on the frequency of spliceable picture in SBS, the degradation in quality of SBS in comparison to a normal unconstrained bit stream can be about 0.1 to 0.05 dB which is insignificant.

Furthermore, pixel wise error is relatively low. For example, the average simulation results on a number of video sequences with different encoding and splicing parameters show a value of 2.5 for mean absolute error.

Moreover, despite of drop in PSNR, the results of a small-scale visual quality test indicated that non-expert viewers do not typically perceive the error. An inter video frame from SBS has been compared with corresponding frame in spliced bit stream in Figure 5, while they have been encoded



Figure 5: Sample video splicing results using a quantization parameter of 28 for the both SBS and DRBS, from left to right: spliceable bit stream, spliced bit stream and splicing error after propagated over 14 inter pictures

with a constant quantization parameter of 28 and the splicing error has propagated over 14 video frames.

It is remarkable that in the conventional IPDC over DVB-H, if we use more than one IDR picture in each MPE-FEC frame to decrease the tune-in time, we should decrease the quality to get the same bit rate. In this case depending on the frequency of IDR frames the degradation in quality can be much higher than the spliced stream while the tune-in time can not be minimized. As example, simulation results on a number of 10 video sequences show that if we decrease the average tune-in time from 1.5 to 0.25 seconds just by increasing the frequency of IDR pictures, for a bit stream with the bitrate 64 kb/s and frame rate of 15 f/s, the quality of encoded video will degrade more than 4 dB in PSNR. Moreover, for shorter tune-in times the drop in quality is much higher than degradation above.

Considering the simplicity of proposed splicing method, and the penalty above, the proposed splicing method can be utilized when the use of SP and SI pictures of H.264/AVC is disallowed.

4. HRD COMPLIANCY OF SPLICED STREAM

According to the proposed splicing method the spliceable pictures and corresponding intra/IDR pictures in two primary streams should be encoded with similar qualities. In a similar quality an IDR frame can consume a bit budget from 5 to 10 times more than corresponding inter picture. Furthermore, similar quality for corresponding frames in two primary streams means that only the bit rate of one primary stream can be controlled. Consequently, there is no real short-term control on the bit rate of spliced streams and therefore it is hard to verify the HRD compliancy of spliced streams. Moreover the encoding parameters cannot be controlled according to the results of splicing, because the encoding and splicing are performed independently and without any feedback link. Results of HRD simulation on a number of spliced bit streams confirm that the problem of HRD compliancy of spliced bit stream is serious and it is impossible to control the bitrate of spliced bit stream just by dropping frames in IP encapsulator.

To solve the problem above, a comprehensive rate control system is proposed which is implemented in both the content encoder and the IP encapsulator. The content encoding level rate control system (CLRCS) controls the bitrate of two primary bit streams considering an average value for the frequency of IDR pictures in a desired spliced stream. However the frequency of IDR picture in the spliced stream can vary around the average value since the number of video pictures in MPE-FEC frames is not fixed. Moreover, in offline encoding the IDR frequency which has been used for the rate control of primary streams at the content encoder level may be very different from the average IDR frequency of spliced stream. The encapsulating

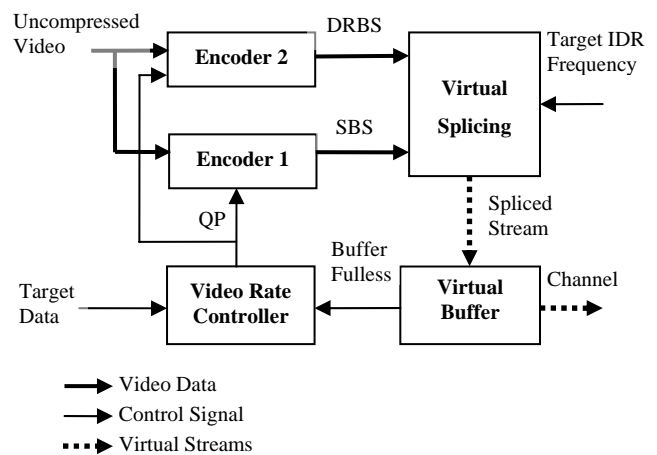


Figure 6: Block diagram of proposed rate control system at content encoder level (CLRCS).

level rate control system (ELRCS) implements another control to compensate the variations above and to provide HRD compliancy for the spliced stream. Moreover, the H.264/AVC buffering period supplemental enhancement information (SEI) message parameters for buffering of spliced stream are provided by ELRCS.

The CLRCS controls the bit rate of primary streams according to encoding target data which are set by the user. The encoding target data include target bit rate of spliced stream and average frequency of IDR pictures in the desired spliced stream. Figure 6 depicts a general block diagram for the proposed CLRCS. As shown two encoders encode the DRBS and SBS primary streams from a common input uncompressed video. They use similar quantization parameters provided by the video rate controller. Various variable video rate controllers with buffer constraint such as our presented controller in [4] can be used in this structure. The rate controller utilizes a virtual buffer for the rate control. The virtual buffer is charged by a virtual spliced stream with a constant IDR period and it is discharged with the target bitrate of spliced stream. The encoded primary bit streams and some meta data including encoding target data and additional statistics about the encoding results are sent to the server and then to IP encapsulator.

The ELRCS at IP encapsulator controls the rate of spliced stream according to the encoding meta data, encapsulating target data defined by the server. The encapsulating target data including target bit rate of spliced stream and IDR frequency of spliced stream are homogeneous with the encoding target data while they may have different values in offline applications.

A HRD-CPB (HRD Coded Picture Buffer) is running at the ELRCS which simulates the decoder buffer. ELRCS utilizes the HRD-CPB to control the bit rate and also to compute the

buffering period SEI message parameters related to buffering of the spliced stream. The ELRCS controls the bitrate of spliced streams by control of frame rate, i.e. it can drop a number of pictures to decrease the bitrate.

To evaluate the standard HRD compliancy of the proposed spliced bit stream, we encoded 45 minutes of video with 10 different contents to provide the primary streams for spliced streams with different target bit rates, IDR frequencies and frame rates. Simulation results show that the proposed rate control system provides standard compliant bit stream with a small percentage of dropped frames and extra IDR frames while the location of IDR pictures is not known for encoders.

5. CONCLUSIONS

In this paper we proposed a video splicing method which minimizes the tune-in time of channel zapping in IP datacasting (IPDC) over DVB-H (Digital Video Broadcasting for Handheld terminals). In order to ensure that the spliced bit stream is compliant with the Hypothetical Reference Decoder specification of the video coding standard, a comprehensive rate control system was proposed. Simulation results show that the proposed splicing method and rate control system can minimize the tune-in time of IPDC over DVB-H at the expense of a relatively small degradation in quality of spliced video.

REFERENCES

- [1] ETSI, "Digital Video Broadcasting (DVB): Transmission systems for handheld terminals," ETSI standard, EN 302 304 V1.1.1, 2004.
- [2] M. Karczewicz, R. Kurceren, "The SP- and SI-frames design for H.264/AVC", *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 13, No. 7, July 2003.
- [3] ETSI, "Specification for the use of Video and Audio Coding in DVB services delivered directly over IP Protocols DVB," ETSI Standard, TS 102 005, 1 November 2005.
- [4] M. Rezaei, S. Wenger, M. Gabbouj, "Video Rate Control for Streaming and Local Recording Optimized for Mobile Devices," *IEEE International Symposium on Personal Indoor and Mobile Radio Communications (PIMRC)*, Berlin, September 2005.
- [5] G. Bjontegaard, "Calculation of average PSNR differences between RD-curves," *Video Coding Experts Group (VCEG)*, Document VCEG-M33 Austin, Texas, USA, 2-4 April, 2001.