VIDEO ENCODING AND SPLICING FOR TUNE-IN TIME REDUCTION IN IP DATACASTING (IPDC) OVER DVB-H

[1]Mehdi Rezaei, [2]Miska M. Hannuksela, [3]Moncef Gabbouj

[1,3]Tampere University of Technology, [2]Nokia Research Center

ABSTRACT

A novel video encoding and splicing method is proposed which minimizes the tune-in time of "channel zapping", i.e. changing from one audiovisual service to another, in IPDC over Digital Video Broadcasting for Handheld terminals (DVB-H). DVB-H uses a time-sliced transmission scheme to reduce the power consumption used for radio reception. Tune-in time in DVB-H refers to the time between the start of the reception of a broadcast signal and the start of the media rendering. 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 is started 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 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 bitstream with IDR pictures according to time-slice boundaries in order to achieve the minimum tune-in time. It has to be ensured that the "spliced" stream resulting from the operation of the IP encapsulator complies with the Hypothetical Reference Decoder (HRD) specification of H.264/AVC. A video encoding and rate control system is proposed to satisfy the HRD requirements for the spliced stream. Simulation results show that in addition to fulfilling HRD compliancy, good average quality of decoded video is achieved with minimum tune-in time.

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.

In a conventional IPDC system over DVB-H, a content encoder receives source signal and encodes the source signal into a coded media bit stream. The coded media bit stream is transferred to a server. The server is typically a normal IP multicast server using real-time media transport over RTP. The server encapsulates the coded media bit stream into RTP packets. The server is connected to an IP Multi-Protocol Encapsulator. 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 a radio transmitter which is not essential for the operation of the proposed encoding and 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, typically capable of receiving, de-modulating, 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.

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

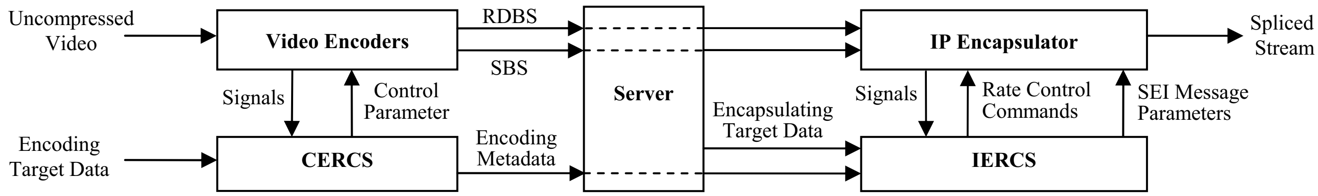

Figure 1: Block diagram of proposed Splicing and rate control method

IDR pictures relative to the boundaries of MPE-FEC frames. Moreover, very frequent IDR pictures in the bitstream drops the compression efficiency remarkably. A method for fast channel zapping in Set-Top Box applications has been presented in [2] in which an auxiliary bit stream including frequent low quality intra pictures is sent to the decoder in parallel to the main bit stream. When channel change the decoder replaces an Inter picture from the main bit stream with a low quality Intra picture from the auxiliary bit stream. Although this method can decrease the tune-in time in IPCD, it can not minimize the tune-in time. Furthermore, the auxiliary bit stream consumes the transmission bandwidth. Moreover, the receiver needs some modifications to switch between two bit streams.

We propose a modification in the operation of IPDC system, which minimizes the tune-in time required for decoder refresh by an IDR picture splicing method to desired locations. When a bitstream is modified, it has to be ensured that HRD (Hypothetical Reference Decoder) compliancy of the modified bitstream is maintained. We present a video encoding and rate control system to ensure the HRD compliancy of the resulting spliced bitstream. Sections 2 and 3 of this paper present details of proposed splicing method and encoding system respectively. Simulation results are provided 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 1. At the content encoding level two video encoders encode the uncompressed input video to two primary bit streams including a Splicable Bit Stream (SBS) and a Decoder Refresh Bit Stream (DRBS) from the same video sequence. The SBS includes very 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 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 temporally an error 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 [3]. The Extended profile of H.264/AVC is not allowed in the current DVB-H standard [4]. Experimental results show that the propagated error is saturated to a constant relative small value after several frames. Moreover visual quality tests show that the error is perceivable hardly visible.

The error propagation degrades the quality of reconstructed video by spliced stream in comparison to SBS. Moreover the quality of decoded SBS is degraded due to the proposed constraint on the reference frames. However, the resulting degradation in quality is small compared to a conventional system in which very frequent IDR frames in a normal bit stream degrade the average quality by consuming more bit budget and still it can not minimize the tune-in time.

3. PROPOSED RATE CONTROL SYSTEM

According to the proposed splicing method the spliceable pictures and corresponding IDR pictures in 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 qualities 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.

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 rate control system (CERCS) controls the bit rate of two primary streams considering a fixed value for the frequency of IDR pictures in a desired spliced stream. However the frequency of IDR picture in the spliced stream can have variation around an 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 may be different from the average IDR frequency of spliced stream. The IP encapsulating level rate control system (IERCS) implements another control to compensate the above variations and to provide HRD compliancy for the spliced bit stream. Furthermore the SEI message parameters related to buffering of the spliced bit stream can be provided by IERCS.

The CERCS controls the bit rate of primary streams according to encoding target data which are set by the user and also according to several signals which are extracted from the uncompressed and compressed video. The encoding target data include target bit rate of spliced stream and average frequency of IDR pictures in the desired spliced stream. Furthermore, some encoding metadata as complementary information are provided by CERCS which are sent to the server and then IP encapsulator.

The IERCS controls the bit rate of spliced stream according to the encoding metadata, encapsulating target data defined by the server. The encapsulating target data includes target bit rate of spliced stream and IDR frequency of spliced stream. Since CERCS can solve the HRD compliancy problem effectively in online applications in which encoding target data and encapsulating target data are similar, the proposed CERCS is presented in details in this paper while further discussion on IERCS is omitted.

Figure 2 illustrates the block diagram of proposed CERCS. The CERCS is configured to provide control on the bit rate of the spliced stream by controlling the bit rate of SBS taking into consideration the bit rate of DRBS. Moreover, the CERCS minimizes the changes of encoding parameters to provide high visual quality for encoded video. It is shown in [5] that minimizing the overall distortion is roughly equivalent to minimizing the variation in quality or in encoding parameters. In the system depicted in Figure 2, two separate video encoders, encoder1 and encoder2, encode the uncompressed video to provide SBS and DRBS, respectively. Two virtual buffers, virtual buffer 1 and virtual buffer 2, are utilized in this system. The virtual buffer 1 provides buffering constraints for the target spliced stream and the virtual buffer 2 moves the extra bit rate resulting from replacing spliceable pictures by IDR pictures to the virtual buffer 1 gradually to minimize short-term fluctuations in encoding parameters and to maximize the quality of encoded video. The VBR (Variable Bit Rate) video rate controller block in the diagram can employ any bitrate control algorithm with buffer constraint, such as our presented algorithm in [6]. The main advantage of proposed CERCS is that it can utilize different available video rate controllers designed for normal video streams without any modification. The quantization parameter provided by the rate controller is used by the two encoders. The target bit rate for the rate controller which is fixed in normal applications, is utilized as a control signal in this system. The target bit rate is provided by the TBRE (target bit rate estimator) block. More details about the virtual buffers and TBRE are presented in the sequel.

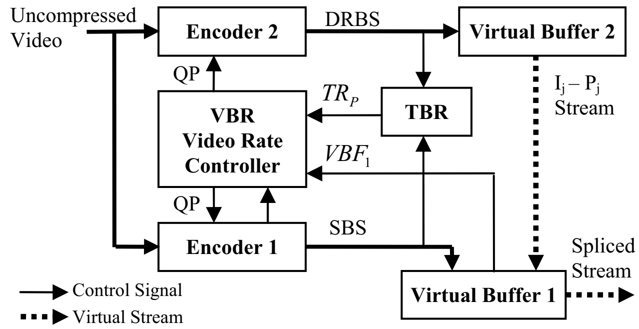

Figure 2: Block diagram of proposed rate control system at content encoder level (CLRCS).

3.1. Virtual Buffers

A receiver buffer model is proposed for the virtual buffer 1. The fullness of the buffer 1 is updated after encoding each picture in SBS as:

$$VBF_1(i+1) = VBF_1(i) - P_i - (I_j - P_j)/F + TR_S/FR, \quad (1)$$

where $VBF_1(i)$ denotes the fullness of virtual buffer 1 updated after encoding ith picture i.e. P_i. TR_S and FR refer to the target bit rate and frame rate of desired spliced stream, respectively. F denotes the average expected frequency of spliced IDR pictures in the desired spliced stream. I_j and P_j stand for the bit budgets consumed by the jth replaced IDR picture and corresponding spliceable picture respectively.

At the replacing locations, difference of an encoded IDR picture and corresponding spliceable picture is inserted at once to the virtual buffer 2 and then the difference is moved gradually to the virtual buffer 1 during the whole IDR period. The occupancy of the virtual buffer 2 is updated after encoding each frame in SBS. At the replacing locations it is updated as:

$$VBF_2(i+1) = VBF_2(i) + I_j - P_j - (I_j - P_j)/F, \quad (2)$$

otherwise it is updated as:

$$VBF_2(i+1) = VBF_2(i) - (I_j - P_j)/F \quad (3)$$

where $VBF_2(i)$ denotes the fullness of virtual buffer 2.

3.2. Target Bit Rate Estimator

The TBRE adjusts the target bit rate of SBS, as a control signal smoothly, according to target bit rate of spliced stream and according to results of encoding DRBS as:

$$TR_P = TR_S - \widetilde{R}_{I-P}, \quad (4)$$

$$\widetilde{R}_{I-P}(z) = H(z) \times R_{I-P}(z), \quad (5)$$

$$R_{I-P} = FR \times (I_j - P_j) / F, \quad (6)$$

where TR_P denotes the target bit rate of SBS. R_{I-P} stands for extra bitrate result of replacing *j*th spliceable picture with corresponding IDR picture. $\widetilde{R}_{I-P}$ refers to low pass filtered version of R_{I-P} provided by a low pass filter such as $H(z)$ presented in our previous paper [6]. The low pass filtering minimizes the fluctuation of target bit rate and thereafter encoding parameter to provide high average quality for encoded video.

4. SIMULATION RESULTS

To study the error propagation in spliced stream, several simulations were run on different video sequences with various encoding and splicing parameters. In each case we measured the propagated error by several criteria such as PSNR, maximum sample-wise absolute difference. The results of simulations show that the propagation error is saturated to a constant relative small value after several frames. The average (on different video sequence) degradation in quality of spliced bit stream relative to SBS is about 1.6 dB which is almost independent of the frequency of IDR pictures. Despite of PSNR drop, the subjective impact is hardly noticeable. Furthermore, the degradation in quality of SBS in comparison to a normal unconstrained bit stream is less than 0.1 dB which is insignificant. If we try to minimize the tune-in time in a normal bit stream just by frequent IDR pictures, a penalty much higher than above degradation in quality should be paid. Therefore, the proposed splicing method can be utilized when the use of SP and SI pictures of H.264/AVC is disallowed.

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 in SBS 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 can decrease the decoder refresh time to very close to zero or even to zero.

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 CERCS can provide standard compliant bit stream when the IDR pictures in the spliced stream in encapsulator have an average frequency about the frequency which is used for the rate control of primary streams by CERCS and while the location of IDR pictures is not known for encoders.

To evaluate CERCS from the video quality point of view, we encoded similar video contents by the CERCS and by the normal rate controller used in CERCS. The average qualities of provided video by two rate controller are very similar. This is outcome of perfect configuration of virtual buffer 2 and TBRE in the proposed CERCS.

The obtained simulation results show that the proposed splicing method and rate control system can provide standard compliant video bit streams for IPDC over DVB-H with good average quality and with minimum tune-in time and complexity.

5. CONCLUSIONS

In this paper we proposed a video encoding and splicing method which minimizes the tune-in time of channel zapping in IP datacasting (IPDC) over DVB-H. 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. The proposed rate control system can utilize many available normal video rate controllers. 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 relative small degradation in quality of spliced video.

6. REFERENCES

[1] ETSI, “Digital Video Broadcasting (DVB): Transmission systems for handheld terminals,” ETSI standard, EN 302 304 V1.1.1, 2004.

[2] J.M. Boyce, A.M. Tourapis, “Fast efficient channel change [set-top box applications],” IEEE International Conference on Consumer Electronics (ICCE), 8-12 Jan. 2005.

[3] 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.

[4] 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.

[5] S. Takamura, N. Kobayashi, “MPEG-2 one-pass variable bit rate control algorithm and its LSI implementation,” *IEEE Int. Con. on Image Processing*, pp. 942–945, 2002.

[6] M. Rezaei, S. Wenger, M. Gabbouj, "Video Rate Control for Streaming and Local Recording Optimized for Mobile Devices," *IEEE Int. Sym. on Personal Indoor and Mobile Radio Communications*, Berlin, September 2005.