

Tune-in Time Reduction in Video Streaming Over DVB-H

Mehdi Rezaei, Miska M. Hannuksela, and Moncef Gabbouj

Abstract—A novel method is proposed which minimizes the tune-in time of “channel zapping”, i.e. changing from one audio-visual 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 standard. 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 analysed 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.

Index Terms—Channel change delay, Digital Video Broadcasting-Handheld (DVB-H), IP datacasting (IPDC), mobile tv, splicing, video coding.

I. INTRODUCTION

DIGITAL Video Broadcasting for Handheld terminals (DVB-H) is an ETSI standard specification for bringing broadcast services to battery-powered handheld receivers [1],

Manuscript received June 27, 2006; revised November 24, 2006. This work was supported in part by Nokia and in part by the Academy of Finland, Project No. 213462 (Finnish Centre of Excellence program 2006–2011).

M. Rezaei and M. Gabbouj are with the Institute of Signal Processing, Tampere University of Technology, FI-33720, Finland (e-mail: Mehdi.Rezaei@tut.fi; Moncef.Gabbouj@tut.fi).

M. M. Hannuksela is with the Nokia Research Center, Tampere, FI-33720, Finland (e-mail: Miska.Hannuksela@nokia.com).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TBC.2006.889682

[2]. 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 Fig. 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 is typically 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. Fig. 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 encapsulator. The connection between the server and an IP network may be a fixed-line private network.

The IP encapsulator encapsulates IP packets into Multi-Protocol Encapsulation (MPE) Sections which are further packetized into MPEG-2 Transport Stream packets. The IP encapsulator optionally uses MPE Forward Error Correction (MPE-FEC) based on Reed-Solomon codes. An IPDC system over DVB-H further includes at least one radio transmitter which is

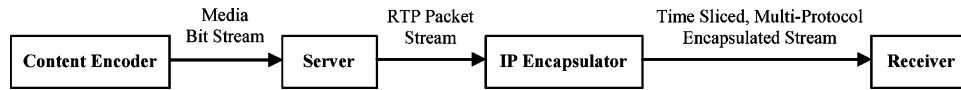


Fig. 1. Simplified block diagram of a conventional IP datacasting system.

not essential for the operation of the proposed splicing system and it is not discussed further.

To reduce power consumption in handheld terminals, the service data is time-sliced and then sent to 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 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. A method for fast channel zapping in Set-Top Box applications has been presented in [3] in which an auxiliary bit stream including frequent low quality IDR pictures is sent to the receiver in parallel to the main bit stream. When channel change the receiver replaces an Inter picture from the main bit stream with an IDR picture from the auxiliary bit stream. Although this method decreases the average tune-in time in IPDC over DVB-H, it cannot minimize the tune-in time, as the IDR picture intervals in the auxiliary bit stream may not match with time-slice intervals. Furthermore, the auxiliary bit stream consumes transmission bandwidth to a considerable amount. Moreover, some modifications in the receiver are required relative to the DVB-H specifications to switch between two bit streams. Finally, the employed low quality IDR pictures degrade the quality of following pictures to the next normal IDR picture.

In our previous work [4], we proposed a video splicing method, which minimizes the decoder refresh time in IPDC over DVB-H without any increase in bandwidth and modification on the receiver. We proposed 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 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 investigated the propagated error to verify that it is subjectively negligible. In proposed splicing method, it had to be ensured that standard compliancy of the resulting bit stream is maintained. We proposed a video rate control system in [5] to guarantee the standard compliancy of the spliced bit stream. The proposed rate control system is implemented in both the content encoder and IP encapsulator. A part of details related to rate control system at the content encoder was presented in [5] and some details related to rate control system at the IP encapsulator were presented in [6]. Furthermore, we presented some details about the buffering of spliced video in [7]. This paper provides an overview on whole research work and more investigation results.

The paper is organized as follows: Section II presents details of the proposed splicing method. The propagated error is investigated in Section III. Section IV presents the details of the proposed rate control system. Simulation results are provided in Section V. The paper is closed with conclusions in Section VI.

II. PROPOSED VIDEO SPLICING METHOD

A simplified block diagram of the proposed IPDC system is depicted in Fig. 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 spliceable bit stream 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 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 decoder refresh bit stream contains only intra or IDR pictures corresponding to spliceable pictures and with a picture quality similar to corresponding spliceable pictures. The Decoder refresh bit stream and the spliceable bit stream are transmitted from the server to the IP encapsulator. The IP encapsulator composes

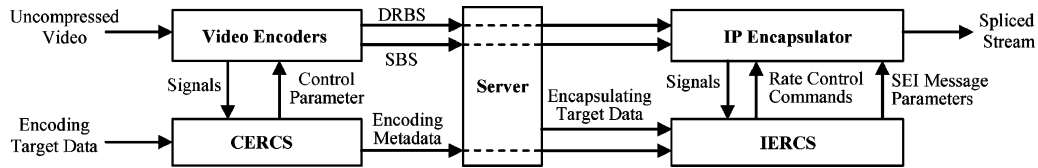


Fig. 2. Block diagram of proposed splicing and rate control method.

MPE-FEC frames, in which the first picture in decoding order is an IDR picture from the decoder refresh bit stream and the other pictures are from the spliceable bit stream. The 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 spliceable bit stream 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 to avoid mismatch altogether 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 [8]. The Extended profile of H.264/AVC is not allowed in the current DVB-H standard [9]. A similar solution was proposed for stream switching in streaming servers by Farber and Girod in [10]. They used a kind of S picture more specifically, to enable switching between bit streams. Unlike SP pictures in [4], S pictures introduce mismatch error while switching between bit streams. To minimize mismatch error, the quantization parameter used for S pictures should be kept small. While in our application we do not need to switch between multiple bit streams, instead of high quality switching frames, the spliceable pictures and corresponding IDR pictures are proposed to be encoded with a higher quality than other pictures to decrease the mismatch error. These alternative solutions for the Decoder Refresh bit stream are studied in this paper and it is shown that it reduces the mismatch error remarkably.

In addition to the mismatch of pixel values, the proposed method involves two buffering related challenges. Firstly, the standard HRD (Hypothetical Reference Decoder) compliancy of the spliced stream is hard to verify in the IP encapsulator. Secondly, the initial delay for coded picture buffering in the HRD is hard to derive in the IP encapsulator. More details about these technical challenges are discussed in the sequel.

III. ERROR PROPAGATION IN SPLICED VIDEO

According to the proposed splicing method the spliceable pictures in spliceable bit stream are replaced with corresponding IDR pictures in decoder refresh bit stream. 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

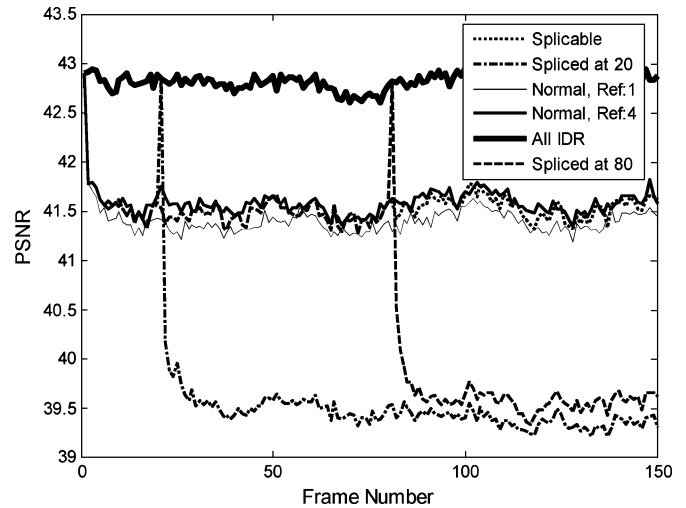


Fig. 3. Quality comparison of different bit streams for the Foreman video sequence, QP = 20.

Rate [11]. Bjontegaard proposed a method for calculating the average difference between two rate-distortion curves in which each curve is interpolated by four points and then the difference between two curves is calculated in terms of PSNR or bitrate. The Bjontegaard Delta PSNR indicates a weighted average PSNR difference in dB over the whole range of bitrates and Bjontegaard Delta Rate indicates the weighted average bitrate difference in percentage over the whole range of PSNRs.

The graphs depicted in Fig. 3 show an example of simulation results. In this example different bit streams are compared in terms of PSNR on the luma component while all of them were encoded with a constant quantization parameter of 20. A bit stream including only IDR frames is compared with spliceable bit stream, two normal bit streams and spliced bit streams. The normal bit streams differ in the maximum number of used reference frames (Ref: 1, Ref: 4). Moreover, the PSNR curves of two spliced bit streams are depicted in which replacing of spliceable picture with IDR picture started in different locations (at frame number 20 and 80). 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 spliceable bit stream depicted in Fig. 3.

As an important result, the degradation in quality of spliced streams in comparison to the spliceable bit streams saturates to a constant value after a small number of frames. This means

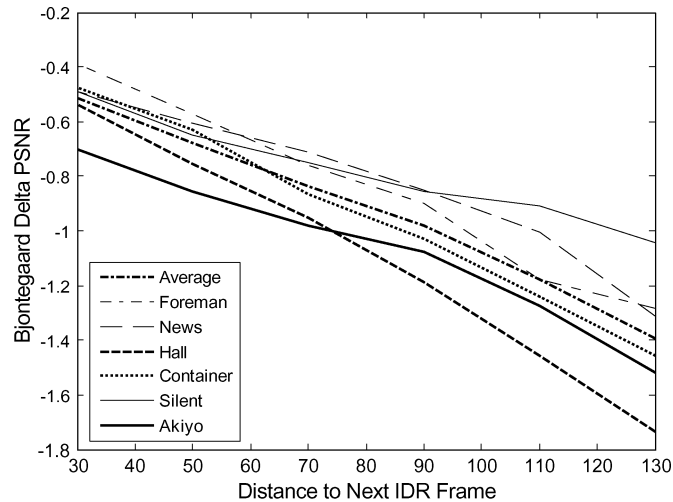


Fig. 4. Bjontegaard Delta PSNR between spliceable and spliced bit streams with one spliced IDR picture.

that the degradation in quality is independent of frequency of IDR frames in spliced bit stream. Hence, using just one IDR picture in each 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, the plots show that the quality degradation of spliced video relative to the normal video is comparable with the quality difference between IDR and following inter frames in normal bit streams which are encoded with a constant quantization parameter.

The graphs depicted in Fig. 4 show the Bjontegaard Delta PSNR between spliceable bit stream 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 consider as variable. Almost linear curves indicate that the average quality degradation in each sequence is proportional to the number of video pictures which are affected by the propagated error and the propagation error is saturated to a constant value. In other words the error level for all pictures after spliced picture is similar in each sequence and there is no accumulation error. The average quality degradation (on different video sequences) between spliceable bit stream 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. Although this approach can improve the average quality of reconstructed video, very high quality spliceable pictures similar to high quality S pictures in [10] can not be used here. Very high quality spliceable pictures increase the tune-in time indirectly by increasing variation in bitrate and required initial buffering period.

While the error propagation degrades the quality of reconstructed video by decoding spliced stream in comparison to spliceable bit stream, the quality of decoded spliceable bit stream 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 spliceable bit stream, the degradation in quality of spliceable bit stream 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 spliceable bit stream has been compared with corresponding frame in spliced bit stream in Fig. 5, while they have been encoded with a constant quantization parameter of 28 and the splicing error has propagated over 14 video frames.

Coding of frequent and periodical IDR pictures into the transmitted bit stream is an alternative means to reduce the tune-in time. To compare the proposed tune-in time reduction method with frequent IDR picture coding, we performed a simulation on 10 video sequences to measure the quality degradation resulting from increased frequency of IDR pictures and hence degraded compression efficiency compared to the proposed method. The results of simulation for the bitrate of 64 kb/s and frame rate of 15 f/s are depicted in Fig. 6. As an example, if the IDR period is changed from 30 to 5 to decrease the average tune-in time from 1.0 to 0.17 seconds respectively, the quality of encoded video degrades about 4.7 dB in average luma PSNR.

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.

IV. VIDEO RATE CONTROL SYSTEM FOR SPLICED VIDEO

A. Video Rate Control System

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 as depicted in Fig. 2 is proposed which is implemented



Fig. 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.

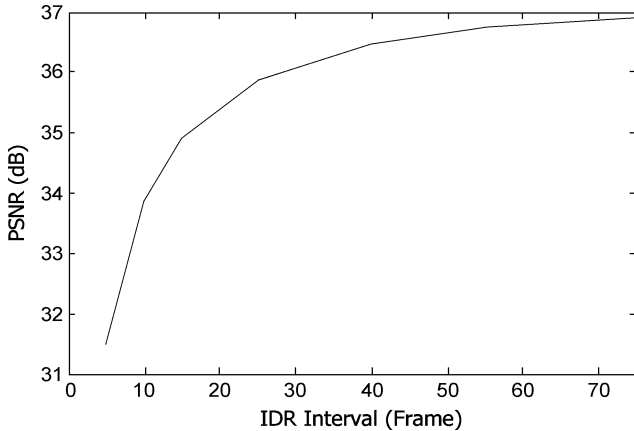


Fig. 6. Effect of frequency of IDR pictures on average quality, bitrate 64 kb/s, frame rate 15 f/s.

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 H.264/AVC Supplemental Enhancement Information (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 is 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

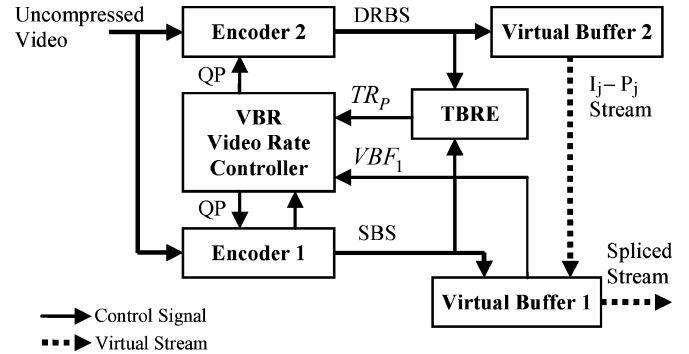


Fig. 7. Block diagram of proposed rate control system at content encoder level (CERCS).

spliced stream and IDR frequency of spliced stream. The details of proposed rate control systems are presented in the sequel.

B. Content Encoder Rate Control System

Fig. 7 illustrates the block diagram of the proposed CERCS. The CERCS is configured to control 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 [12] that minimizing the overall distortion is roughly equivalent to minimizing the variation in quality or in encoding parameters. In the system depicted in Fig. 7, two separate video encoders, encoder1 and encoder2, encode the uncompressed video to provide spliceable bit stream and decoder refresh bit stream, 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 algorithms in [13] and [14]. 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.

1) *Virtual Buffers*: A receiver buffer model is proposed for the virtual buffer 1. The fullness of buffer 1 is updated after encoding each picture in spliceable bit stream 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 the i th picture i.e. P_i . TR_S and FR refer to the target bit rate and frame rate of the 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 j th 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 (according to (2)) and then the difference is moved gradually to virtual buffer 1 during the whole IDR period (according to formula 3). The occupancy of the virtual buffer 2 is updated after encoding each frame in spliceable bit stream. 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.

2) *Target Bitrate Estimator*: The TBRE adjusts the target bit rate of spliceable bit stream, as a control signal smoothly, according to target bit rate of spliced stream and according to results of encoding decoder refresh bit stream as:

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

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

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

$$H(z) = \frac{m}{m+1-z^{-1}}, \quad (7)$$

where TR_P denotes the target bit rate of spliceable bit stream. R_{I-P} stands for extra bitrate result of replacing j th spliceable picture with corresponding IDR picture. \tilde{R}_{I-P} refers to low pass filtered version of R_{I-P} provided by a low pass filter such as $H(z)$. m is a constant value larger than one and good results are obtained with $m = 2$. The low pass filtering minimizes the fluctuation of target bit rate and thereafter encoding parameter to provide high average quality for encoded video.

C. IP Encapsulator Rate Control System

Fig. 8 illustrates the block diagram of the proposed IERCS. The IERCS utilizes a fuzzy rate controller and a Virtual Buffer. The fuzzy controller controls the bit rate of spliced stream by controlling the frame rate and the type of pictures. It may drop a number of pictures from an MPE-FEC frame to decrease the bit rate or it may replace one or more extra spliceable pictures by IDR pictures to increase the bit rate. However, as another alternative, stuffing NAL units or SEI message to the bit stream can be used instead of using extra IDR picture to increase the

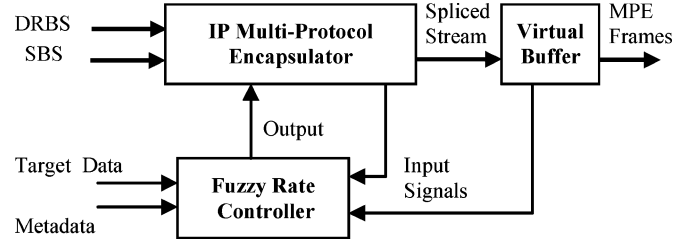


Fig. 8. Block diagram of proposed IERCS.

bit rate. The fuzzy controller and virtual buffer operate based on MPE frame. The size of virtual buffer can be computed according to metadata and encapsulating target data.

The fuzzy controller is configured to provide control on the bit rate of the spliced stream by minimum variation in the frame rate. It minimizes the number of dropped pictures and also it prevents unnecessary IDR pictures. The output of controller is an integer number. A positive number shows the number of pictures that should be dropped from the start or end of MPE frame and a negative number shows the number of extra spliceable pictures which can be replaced by IDR pictures in MPE frame. Locations of extra IDR pictures are distributed uniformly along the MPE frame. The fuzzy controller uses two following signals as inputs:

$$INPUT1 = FB \times FR/TR_S - F, \quad (8)$$

$$INPUT2 = (BF - I_j + P_j)/BS, \quad (9)$$

where FB denotes the total number of bits consumed by the current MPE frame before any dropping or extra IDR picture. BF and BS refer to the fullness and size of virtual buffer respectively. TR_S and FR are the target bit rate and frame rate of spliced stream. F denotes the target number of pictures in one MPE frame in the target frame rate. I_j and P_j stand for the bit budgets consumed by the j th replaced IDR picture and corresponding spliceable picture respectively.

All the defined fuzzy rules are summarized in the Table I. The content of table specifies the output of controller. The letters H, L, M, V, X and S correspond to linguistic specifications of High, Low, Medium, Very, Extremely and Super. The desired central values for the output of fuzzy system correspond to VL, L, ML, M, MH, H, VH, XH, SH are $-3, -2, -1, 0, 1, 2, 3, 4$ and 5 . The distributions of fuzzy membership functions are shown in Fig. 9.

We used a well-known and simple fuzzy system with two inputs using “Product Inference Engine”, singleton fuzzifier and center average defuzzifier which is

$$f(x_1, x_2) = \frac{\sum_{i_1=1}^{N_1} \sum_{i_2=1}^{N_2} \bar{y}^{i_1 i_2} \mu_{A_1^{i_1}}(x_1) \cdot \mu_{A_2^{i_2}}(x_2)}{\sum_{i_1=1}^{N_1} \sum_{i_2=1}^{N_2} \mu_{A_1^{i_1}}(x_1) \cdot \mu_{A_2^{i_2}}(x_2)}, \quad (10)$$

where $f(x_1, x_2)$ denotes the estimated output and $\{A_1^1, A_1^2, \dots, A_1^{N_1}\}_{i=1,2}$ are fuzzy sets with $\{\mu_{A_1^{i_1}}(x_1)\}_{1 \leq i_1 \leq N_1}$ and $\{\mu_{A_2^{i_2}}(x_2)\}_{1 \leq i_2 \leq N_2}$ membership functions defined for inputs x_1 and x_2 respectively. The center of output fuzzy set ($B^{i_1 i_2}$), denoted by $\bar{y}^{i_1 i_2}$, is chosen

TABLE I
SUMMARIZATION OF IF-THEN FUZZY RULES

INPUT1	XL	M	M	M	M	ML	L	VL
	VL	M	M	M	M	M	ML	L
	L	M	M	M	M	M	M	ML
	ML	M	M	M	M	M	M	M
	M	MH	M	M	M	M	M	M
	MH	H	MH	M	M	M	M	M
	H	VH	H	MH	M	M	M	M
	VH	XH	VH	H	MH	M	M	M
	XH	SH	XH	VH	H	MH	M	M
	VL	L	ML	M	MH	H	VH	

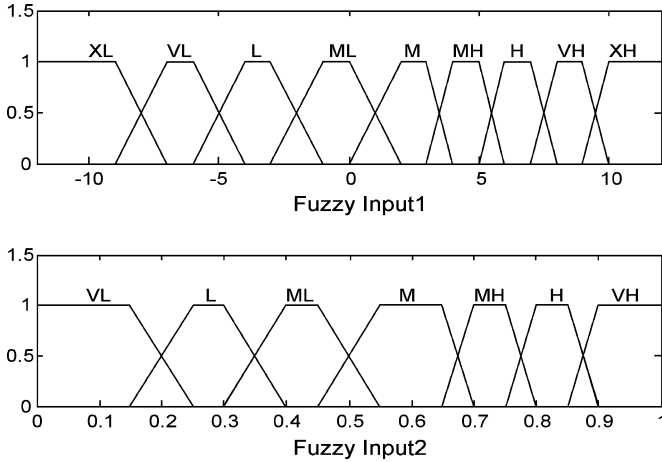


Fig. 9. Fuzzy membership functions of linguistic variables.

as output desired value using our practical experiences. More information about the above fuzzy system (10) is presented in [15].

D. Buffer Size Calculations

In proposed rate control system we utilize three virtual buffers: two buffers in the content encoder rate control system and one buffer in the IP encapsulator rate control system. From the standard compliancy point of view the buffer sizes should be computed according to size of the Coded Picture Buffer (CPB) of the HRD.

From the bitrate control point of view and from the quality point of view several points should be considered in implementation of proposed rate control system. First, the size of virtual buffer which is running at CERCS is close to size of HRD CPB which is constraint by the video coding standard. Second, according to results in [7], the size of virtual buffer in IERCS should be larger than the size of virtual buffer1 running at the CERCS. A larger buffer in IERCS minimizes the number of dropped pictures. On the other hand a HRD CPB with higher occupancy increases the delay result of initial buffering period at the decoder. Finally, a larger virtual buffer at the CERCS can increase the quality of compressed video by allowing more variation in the bitrate and a smaller virtual buffer decrease the quality of encoded video by less variation in bitrate and more variation in quality. Therefore, we should find optimal values

for the size of buffers such that they conform to all the constraints above and also maximize the quality of encoded video and minimize the number of dropped pictures and delay. A practical solution for the problem was proposed in [7]. A general rule for the computation of buffer sizes is proposed as follow:

$$\begin{aligned} & (\text{Encoder Buffer Size} + \text{Average of IDR Picture Size}) \\ & \leq BS \ll \text{Max Standard Size of HRD CPB} \end{aligned}$$

where the encoder buffer size means the capacity of virtual buffer 1 in Fig. 7 and BS denotes the capacity of virtual buffer which is running at the IP encapsulator.

E. Offline Applications

In offline applications if the IDR frequency used in the CLRCS is equal to the average IDR frequency used by the ELRCS, the target rate of the spliced stream at the encapsulator is similar to the target rate of the spliced stream at the encoder. Otherwise, there is a small drift from the target bit rate resulting from a drift in frequency of IDR pictures. In this case the new target bit rate can be computed by the server as:

$$TR_2 = TR_1 \cdot \frac{F_1}{F_2} \cdot \frac{BR_P(F_2 - 1) + BR_I}{BR_P(F_1 - 1) + BR_I}, \quad (11)$$

where TR_1 and TR_2 denote the target bit rate of the spliced stream at the encoder and the encapsulator, respectively. F_1 and F_2 indicate the IDR period used by the CLRCS and the ELRCS, respectively. BR_P and BR_I denote the average bit rate of the spliceable bit stream and decoder refresh bit stream respectively.

V. SIMULATION RESULTS

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 a normal bit stream and the average IDR picture interval is equal to the time-slice interval, the expected tune-in time 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 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. Some detailed results were presented in Section III. The main results related to the error propagation are summarized as follow. The results of simulations show that the propagation error is saturated to a constant relative small value after several frames. This means the average degradation in quality is independent of the IDR period or MPE-FEC frame size. In other words using one IDR frame in each MPE-FEC frame minimizes the tune-in time and the bitrate simultaneously at the expense of a constant degradation in quality of spliced video. According to the simulation results, the average degradation in quality of spliced bit stream is mainly a

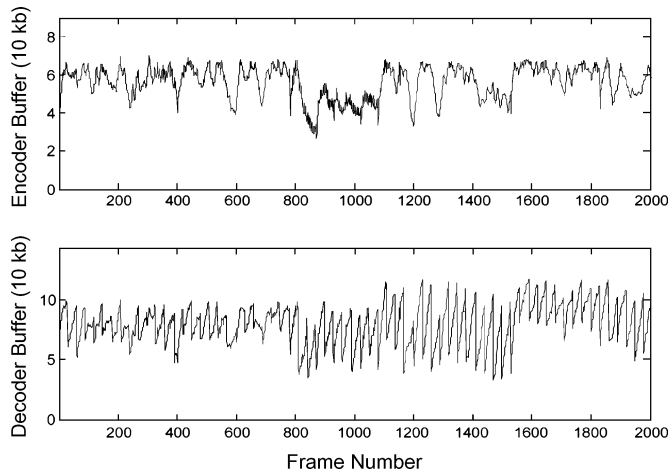


Fig. 10. Simulation of encoder and decoder buffer fullness for proposed video splicing method.

function of quantization parameter and video content. The average simulation results on 6 different video content in 5 quantization parameters in the range of 20 to 40 show that the degradation in quality of spliced bit stream relative to a normal bit stream is about 1.6 dB in luma PSNR which is almost independent of the frequency of IDR pictures. When we encode spliceable pictures and corresponding IDR pictures with a little higher quality than other pictures, the average degradation in quality decreases to 1.1 dB. Despite of PSNR drop, the subjective impact is hardly noticeable. If we try to decrease 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 whereas it can not minimize the tune-in time.

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 picture frequencies and frame rates. We concatenated and mixed several known video sequences which are used for the standardization process to provide long video sequences suitable for our simulations. 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 in spliced stream is not known for encoders. Furthermore, the percentage of dropped-pictures is very small (less than 0.3%) even if the frequency of IDR pictures in the spliced stream is different from the average IRD picture frequency which is used for the rate control of primary streams. Moreover, the simulation results show that with a common content encoder rate control system, without the proposed fuzzy rate controller in the IP encapsulator, the percentages of dropped pictures and extra IDR picture can be much higher than when we use the proposed fuzzy rate controller. Fig. 10 shows the simulation results of encoder and decoder buffers for a video sequence in the proposed video splicing and rate control system. While the spliced bit stream in decoder buffer looks very different from the primary bit streams in the encoder buffer, still it is a standard

compliant bit stream with controlled bit rate. The used video sequence in the experiment above is a concatenated version of several known video clips including Foreman, Carphone, News, Salesman, Hall, Paris, Container, Silence, Akiyo, New York, Suzie and Sailboat (not all the frames are shown in the figure).

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

VI. CONCLUSION

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. The proposed method propagates a reference mismatch error in the spliced video bit stream. The error propagation was studied and a comprehensive rate control system was proposed to ensure that the spliced bit stream is compliant with the Hypothetical Reference Decoder specification of the video coding standard. The proposed rate control system is implemented in both the content encoder and IP encapsulator. 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.

REFERENCES

- [1] *Digital Video Broadcasting (DVB): Transmission Systems for Handheld Terminals*, ETSI standard, EN 302 304 V1.1.1, ETSI, 2004.
- [2] G. Faria, J. A. Henriksson, E. Stare, and P. Talmola, "DVB-H: Digital broadcast services to handheld devices," *Proceeding of the IEEE*, vol. 94, no. 1, Jan. 2006.
- [3] J. M. Boyce and A. M. Tourapis, "Fast efficient channel change [set-top box applications]," in *IEEE Int. Con. on Consumer Electronics (ICCE)*, Jan. 8–12, 2005.
- [4] M. M. Hannuksela, M. Rezaei, and M. Gabbouj, "Video encoding and splicing for tune-in time reduction in IP datacasting (IPDC) over DVB-H," in *IEEE Int. Symp. on Broadband Multimedia Systems and Broadcasting*, Las Vegas, Apr. 2006.
- [5] M. Rezaei, M. M. Hannuksela, and M. Gabbouj, "Video encoding and splicing for tune-in time reduction in IP datacasting (IPDC) over DVB-H," in *IEEE Int. Conf. on Multimedia & Expo (ICME)*, Toronto, Canada, Jul. 2006, pp. 601–604.
- [6] —, "Video splicing and fuzzy rate control in IP multi-protocol encapsulator for tune-in time reduction in IP datacasting (IPDC) over DVB-H," in *IEEE Int. Conf. on Image Processing (ICIP)*, Atlanta, USA, Oct. 2006, pp. 3041–3044.
- [7] —, "Spliced video and buffering considerations for tune-in time minimization in mobile TV," in *IEEE Int. Symp. on Personal Indoor and Mobile Radio Communications (PIMRC)*, Helsinki, Finland, Sep. 2006.
- [8] M. Karczewicz and R. Kurceren, "The SP- and SI-frames Design for H.264/AVC," *Circuits and Systems for Video Technology, IEEE Trans. on*, vol. 13, no. 7, Jul. 2003.
- [9] *Specifications for the Use of Video and Audio Coding in DVB Services Delivered Directly Over IP Protocols DVB*, ETSI Standard, TS 102 005, ETSI, Nov. 2005.
- [10] N. Farber and B. Girod, "Robust H.263 compatible video transmission for mobile access to video servers," in *Proc. IEEE Int. Conf. on Image Processing (ICIP)*, Santa Barbara, CA, USA, Oct. 1997, vol. 2, pp. 73–76.
- [11] G. Bjontegaard, "Calculation of average PSNR differences between RD-curves," in *ITU-T Video Coding Experts Group (VCEG)*, Austin, Texas, USA, Apr. 2–4, 2001, Document VCEG-M33.

- [12] S. Takamura and N. Kobayashi, "MPEG-2 one-pass variable bit rate control algorithm and its LSI implementation," in *IEEE Int. Conf. on Image Processing*, 2002, pp. 942–945.
- [13] M. Rezaei, S. Wenger, and M. Gabbouj, "Video rate control for streaming and local recording optimized for mobile devices," in *IEEE Int. Sym. on Personal Indoor and Mobile Radio Communications (PIMRC)*, Berlin, Sep. 2005.
- [14] M. Rezaei, M. M. Hannuksela, and M. Gabbouj, "Low-complexity fuzzy video rate controller for streaming," in *IEEE Int. Conf. on Acoustic, Speech and Signal Processing (ICASSP)*, Toulouse, France, May 2006.
- [15] L. X. Wang, *Adaptive Fuzzy System and Control: Design and Stability Analysis*. Englewood Cliffs, NJ: Prentice-Hall, 1994.



Mehdi Rezaei received the B.S. degree in electronics engineering from Amir Kabir University of Technology (Polytechnic of Tehran), in 1992 and he received the M.S. degree in electronics engineering from Tarbiat Modares University of Tehran in 1996.

He was academic member of Electrical Engineering Department, University of Sistan & Baluchestan, Iran, from 1997 to 2003. He has been a researcher of Institute of Signal Processing, Tampere University of Technology, Finland, since 2003. His research interests include video signal

processing, variable bit rate video, video rate control, video splicing, video multiplexing, region of interest (ROI) video coding, scalable video coding (SVC), video enhancement, video streaming and communication, Mobile TV and DVB-H. He has published several papers in these fields.

Mr. Rezaei was the recipient of the Nokia Foundation Award in 2005 and also in 2006.



Miska M. Hannuksela received the M.S. degree in engineering from Tampere University of Technology, Tampere, Finland, in 1997.

He is currently a Senior Research Manager heading visual technologies research in Nokia Research Center, Tampere, Finland. From 1996 to 1999, he was a Research Engineer with Nokia Research Center in the area of mobile video communications. From 2000 to 2003, he was a Project Team Leader and Specialist in various mobile multimedia research and product projects in Nokia Mobile Phones. From

2003 to 2006 he was a Research Manager heading Video/Image Coding and Transport as well as Video/Image Transport and Systems groups of Nokia Research Center.

Mr. Hannuksela has been an active participant in the ITU-T Video Coding Experts Group since 1999 and in the Joint Video Team of ITU-T and ISO/IEC since its foundation in 2001. He has co-authored more than 100 technical contributions to these standardization groups. Mr. Hannuksela has also contributed to several other multimedia standards, such as IP datacasting over DVB-H and 3GPP multimedia services. His research interests include video error resilience, scalable video coding, and video communication systems. He has co-authored several tens of papers in these fields.



Moncef Gabbouj received his BS degree in electrical engineering in 1985 from Oklahoma State University, Stillwater, and his MS and PhD degrees in electrical engineering from Purdue University, West Lafayette, Indiana, in 1986 and 1989, respectively.

Dr. Gabbouj is currently Professor and Head of the Institute of Signal Processing at Tampere University of Technology, Tampere, Finland. Dr. Gabbouj is the co-founder and past CEO of SuviSoft Oy Ltd. From 1995 to 1998 he was a professor with the Department of Information Technology of Pori School of Tech-

nology and Economics, and during 1997 and 1998 he was a Senior Research Scientist with the Academy of Finland. From 1994 to 1995 he was an Associate Professor with the Signal Processing Laboratory of Tampere University of Technology, Tampere, Finland. From 1990 to 1993 he was a senior research scientist with the Research Institute for Information Technology, Tampere, Finland. His research interests include multimedia content-based analysis, indexing and retrieval; nonlinear signal and image processing and analysis; and video processing and coding.

Dr. Gabbouj is a Honorary Guest Professor of Jilin University, China (2005–2010). Dr. Gabbouj is a Distinguished Lecturer for the IEEE Circuits and Systems Society and Past-Chairman of the IEEE-EURASIP NSIP (Nonlinear Signal and Image Processing) Board. He was chairman of the Algorithm Group of the EC COST 211quat. He served as associate editor of the IEEE TRANSACTIONS ON IMAGE PROCESSING, and was guest editor of the European journal Applied Signal Processing (Image Analysis for Interactive Multimedia Services, Part I in April and Part II in June 2002) and Signal Processing, special issue on nonlinear digital signal processing (August 1994).

He is the past chairman of the IEEE Finland Section and past chair of the IEEE Circuits and Systems Society, Technical Committee on Digital Signal Processing, and the IEEE SP/CAS Finland Chapter. He was also Chairman of CBMI 2005, WIAMIS 2001 and the TPC Chair of ISCCSP 2004, CBMI 2003, EU-SIPCO 2000, NORSIG 1996 and the DSP track chair of the 1996 IEEE ISCAS. He is also member of EURASIP AdCom. He also serves as Publication Chair and Publicity Chair of IEEE ICIP 2005 and IEEE ICASSP 2006, respectively.

Dr. Gabbouj is the Director of the International University Programs in Information Technology and vice member of the Council of the Department of Information Technology at Tampere University of Technology. He is also the Vice-Director of the Academy of Finland Center of Excellence SPAG, Secretary of the International Advisory Board of Tampere International Center of Signal Processing, TICSP and member of the Board of the Digital Media Institute. He serves as Tutoring Professor for Nokia Mobile Phones Leading Science Program (1998–2001). He is a member of Eta Kappa Nu, Phi Kappa Phi, IEEE SP and CAS societies.

Dr. Gabbouj was the recipient of the 2005 Nokia Foundation Recognition Award and co-recipient of the Myril B. Reed Best Paper Award from the 32nd Midwest Symposium on Circuits and Systems and co-recipient of the NORSIG 94 Best Paper Award from the 1994 Nordic Signal Processing Symposium. He is co-author of over 280 publications.

Dr. Gabbouj has been involved in several past and current EU Research and education projects and programs, including ESPRIT, HCM, IST, COST, Tempus and Erasmus. He also served as Evaluator of IST proposals, and Auditor of a number of ACTS and IST projects on multimedia security, augmented and virtual reality, image and video signal processing.