Fuzzy Joint Encoding and Statistical Multiplexing of Multiple Video Sources with Independent Quality of Services for Streaming over DVB-H

[1]Mehdi Rezaei, [2]Imed Bouazizi, [3]Moncef Gabbouj
[1,3]*Tampere University of Technology,* [2]*Nokia Research Center, Finland*
Mehdi.Rezaei@tut.fi, Imed.Bouazizi@nokia.com, Moncef.Gabbouj@tut.fi

Abstract

A joint video encoding and statistical multiplexing (StatMux) method for broadcasting over DVB-H (Digital Video Broadcasting for Handhelds) channels is proposed to decrease the channel changing delay. DVB-H uses a time-sliced transmission scheme to reduce the power consumption used for radio reception in DVB-H receivers. Channel changing delay, i.e. changing from one audio-visual service to another, increases due to the time slicing scheme in DVB-H. The proposed method decreases the buffering delay as a part of channel changing delay by dynamically distribution of available bandwidth between the video sources according to their temporal bit rates. Moreover, the proposed method in combination with the time-slicing decreases the overall channel changing delay by other means.

1. Introduction

Digital Video Broadcasting for Handheld terminals (DVB-H) is an ETSI specification for delivering broadcast services to battery-powered handheld receivers [1], [2]. DVB-H is mainly based on the DVB-T specification for digital terrestrial television. However, it adds a number of features designed to consider the limited battery life of handheld devices and the particular environments in which such receivers operate. Services used in mobile handheld terminals require relatively low bit rates. The estimated maximum bit rate for streaming video using advanced compression technology like H.264/AVC is in the order of a few hundred kilobits per second. A DVB transmission system usually provides a bit rate of 10 Mbps or more. This provides a possibility to significantly reduce the average power consumption of a DVB-H receiver by introducing a scheme based on time division multiplexing. This scheme is called Time-slicing. 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 small fraction of the time, while receiving bursts of a requested service. It significantly reduces the power consumption used for radio reception. DVB-H also employs additional forward error correction to further improve mobile and indoor reception performance of DVB-T.

Channel changing delay or tune-in time in DVB-H refers to the time between the start of switching and the start of the media rendering for new channel. Channel changing delay includes several parts: *Arrival Delay* (delay to arrival of desired burst), *Reception Delay* (reception duration of desired burst), *Decapsulation Buffering Delay*, *Decoder Refresh Delay* (delay to the first random access point), *Decoder Buffering Delay* (Initial buffering period of *Coded Picture Buffer*). The decapsulation buffering delay includes two buffering delays for the *Multi-protocol Decapsulation Buffer* (MDB) and *RTP (Real Time Protocol) Decapsulation Buffer* (RDB). The decapsulation buffering delay is required to compensate the variation of burst size and the decoder buffering delay is needed to compensate the variation of bitrate. Moreover, another delay is needed for synchronization between the associated streams (e.g. audio and video) of the streaming session which is not shown in the graph.

One of the significant factors in channel changing delay is the arrival delay. The arrival delay depends on the time-slicing parameters that define the power consumption of DVB-H receiver. Lower the receiver power consumption, higher is the arrival delay. Another factor in channel changing delay is the required delay to compensate the variation in bit rate. For video streaming over DVB-H application the advantages of variable bit rate video is exploited. For most video contents, a variable bit rate video can provide better visual quality and coding efficiency than a constant bit rate video. A higher quality and compression performance can be obtained by more variations in bit rate and higher buffering delay.

StatMux in DVB-H can reduce the cannel changing delay by two ways. First, when the multiplexed services are encapsulated to one burst, which is completely received by receiver, channel changing between two multiplexed services can be decreased considerably. In this case, the arrival delay and the reception delay are minimized to zero. The decoder refresh delay and decapsulation buffering delay decrease and depending on the time of switching even they can be minimized to zero. An average reduction about 50% for decapsulation delay and decoder refresh delay is expected. Second, a joint encoding and StatMux method can decrease the decoder buffering delay by providing variable bandwidths proportional to temporal variations in the bit rates of bit streams.

The major problem of joint video encoding and StatMux is how to allocate the available bit budget among the video sources that share the common channel bandwidth and are jointly encoded. All the studied methods follow two main approaches: *Forward Analysis* and *Modeling Strategy*. In forward analysis a preprocessing is performed on video sources to gather statistics about the coding complexity. The real coding process can operate based on the statistics obtained by the pre-analysis. In the modeling approach first it is attempted to model the performance of video encoder and the coding complexity of video sources and then the allocated bits to video sources is controlled based on provided models while the models are updating during encoding. See the proposed methods in [3], [4] and [5] as examples for the two approaches. The system presented in [3] consists of several preprocessors and video encoders. Each preprocessor analyzes a video source and derives picture statistics. Using these statistics, a joint rate controller calculates dynamically the bit rate for each encoder based on the relative complexities of the sources. Another bit allocation method for joint coding of multiple video sources is presented in [4]. In this method, the input video sources are divided into *Super GOPs* (a number of GOP) and *Super Frames* (a collection of frames, one from each program) and then, the bit budget is distributed hierarchically between the video super GOPs, super frames and frames according to their relative complexities while the encoder and decoders' buffers are prevented from overflowing and under flowing. Finally, using a rate-distortion model, a quantization parameter is calculated for each frame according to the allocated bits to the frame. A similar approach to that in [4] is presented in [5]. In this paper we propose a new fuzzy joint encoding and StatMux method to decrease the decoder buffering delay as a major part of channel changing delay in DVB-H application.

The paper is organized as follows: Section 2 provides detailed information about the proposed joint encoding and StatMux method. Simulation results of proposed method are provided in Section 3. The paper is closed with conclusions in Section 4.

2. Proposed Joint Video Encoding and Statistical Multiplexing Method

The proposed joint encoding and StatMux method is implemented by a joint video rate control system. Fig. 1 shows the simplified block diagram of the rate control system. The rate control system utilizes a fuzzy controller to decrease the buffering delay of all encoded bit stream and thereafter the channel changing delay. The proposed system is a real-time controller without any look a head. Utilizing the fuzzy controller, it has a very low degree of complexity in comparison to the previous studied methods. Unlike the previous methods, in proposed method the multiplexed bit streams can have independent quality of services and different bit rates. Although the proposed method allows short-term exchanges of bit budgets between bit streams, it prevents long-term exchange of bit budget between bit streams. Therefore the average quality of encoded bit stream maintains constant in comparison to independent encoding case. An independent video rate controller (IRC) is used for encoding each video bit stream to guarantee a variable bit rate bit stream with an average bit rate and a limited delay. The encoded bit streams are multiplexed and move to a virtual joint buffer. The data is removed from the joint buffer with a constants bitrate appropriate to the transmission channel. The occupancy of joint buffer is used as a feedback signal by a joint rate controller (JRC). The JRC controls the variations in bitrate of joint stream to guarantee a limited delay for the joint stream. Less variation in the bitrate of joint bit stream means smaller buffering delays for the each video bit stream at the receiver. This structure not only decreases the decoder buffering delay by allocating variable bandwidth instead of a fixed band width to each bit streams, but also , it decreases the buffering delay that is required before start of encapsulating at the IP Encapsulator of DVB-H network. These two delays are related to each other and any reduction in decoder buffering delay means a similar reduction in buffering delay before the IP encapsulator.

The video encoders are controlled by the Quantization Parameter (QP) per picture basis. The QP is mainly controlled by the IRCs while the JRC may add a small positive or negative value to the QP according to the occupancy of joint buffer.

$$Q_n = Q_{IRCn} + \Delta Q_J \,, \tag{1}$$

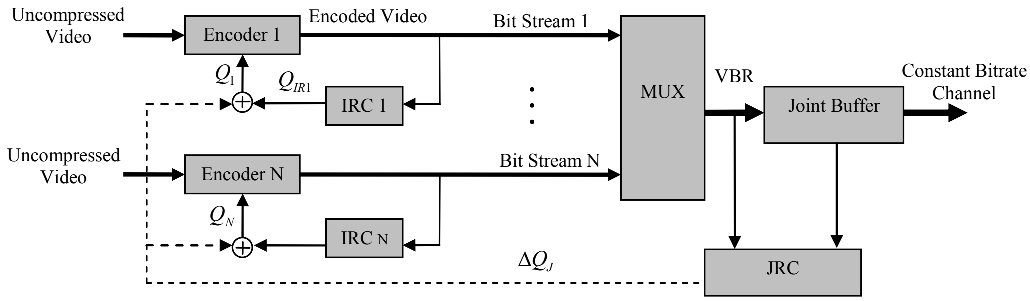

Fig. 1. Block diagram of proposed joint encoding system

where Q_n denotes the used QP by the *n*th encoder. Q_{IRCn} is the QP calculated by the *n*th independent rate controller and ΔQ_J stands for the out put of JRC. Any variable bitrate rate control algorithm such as our presented algorithms in [6] can be used in IRC blocks. The output of JRC is computed according to the occupancy of joint buffer and according to the total bit rate. More details about the proposed control system are presented in the sequel.

A. Virtual Joint Buffer

Considering the virtual buffer at the receiver side, the buffer occupancy is updated after encoding a series of corresponding pictures (m th of each source) as:

$$O_B(m) = O_B(m-1) - \sum_{i=1}^{N} B_i + R \quad , \tag{2}$$

where O_B denotes the buffer occupancy. B_i stands for the consumed bits by the encoded picture of *i*th source.

B. Joint Rate Controller

The output of JRC is computed by a fuzzy controller. While each bit stream uses an independent controller with a buffer constraint, without any other control the multiplexed bit stream is constraint to a joint buffer with a size equal to sum of size of the buffers used by IRCs. The idea is to use a virtual buffer with the size of S_B as smaller as possible and then use the JRC to operate only when the buffer condition is critical. According to the block diagram, we need a loose controller to provide a small quantized output based on rough measurements on the occupancy of joint buffer and the bitrate of multiplexed bit stream. These properties make it fit to a fuzzy controller with low resolution inputs and output.

The fuzzy controller has been designed such that the JRC has a non-zero value only when the buffer conditions are critical. This minimizes the interaction between encoders and also it minimizes the variation in the quality of each bit stream. The fuzzy controller has two input signal as:

$$x_1 = O_B / S_B \quad , \tag{3}$$

$$x_2 = \frac{F \sum_{i=1}^{N} B_i}{\sum_{i=1}^{N} \left(R_i / \left(1 + \frac{X_{IPi} - 1}{I_{Ii}} \right) \right)} \quad , \tag{4}$$

where R_i denotes the target bit rate of the *i*th bit stream. I_{Ii} indicates the interval of frequent IDR pictures in the *i*th bit streams. F stands for the frame rate of bit streams and X_{IPi} indicates the relative coding complexity of IDR pictures to P pictures in the *i*th bit stream.

All the fuzzy rules are summarized in Table I. The content of Table II specifies the output of the controller. The letters H, L, M and V correspond to linguistic specifications of High, Low, Medium and Very. As an example from the table it can be expressed as: *if* (x_1 is VL and x_2 is M) *Then* (Output is H).

The input signals are specified by their fuzzy membership functions (MSF). Numbers of 7 and 3 MSFs for the two inputs x_1 and x_2 were considered. The linguistic fuzzy rules and MSFs were designed based on some theoretic and experimental results and provided experiences in [3] and [4]. Furthermore, an optimization process was performed for fine tuning of the fuzzy MSFs. The distributions of MSFs are shown in the Fig. 2. The desired central values for the output of fuzzy system correspond to VL, L, ML, M, MH, H and VH in the Table I are -3, -2, -1, 0, 1, 2 and 3 respectively. We used a well-known fuzzy system with two inputs using "Product Inference Engine", singleton fuzzifier, and centre average defuzzifier which is:

TABLE I. SUMMARIZATION OF THE IF-THEN FUZZY RULES

x_2 \ x_1	VL	L	ML	M	MH	H	VH
H	VH	H	MH	M	M	M	ML
M	H	MH	M	M	M	ML	L
L	MH	M	M	M	ML	L	VL

$$f(x_1,x_2)=\frac{\sum_{i_1=1}^{N_1}\sum_{i_2=1}^{N_2}\bar{y}^{i_1 i_2}\mu_{A_1^{i_1}}(x_1).\mu_{A_2^{i_2}}(x_2)}{\sum_{i_1=1}^{N_1}\sum_{i_2=1}^{N_2}\mu_{A_1^{i_1}}(x_1).\mu_{A_2^{i_2}}(x_2)}, \quad (5)$$

where $f(x_1,x_2)$ denotes the output and $\{A_i^1, A_i^2, ..., A_i^{N_i}\}_{i=1,2}$ are fuzzy sets with $\{\mu_{A_1^{i_1}}(x_1)\}_{1\le i_1 \le N_1}$ and $\{\mu_{A_1^{i_2}}(x_2)\}_{1\le i_2 \le N_2}$ membership functions defined for inputs x_1 and x_2. The centre of output fuzzy set, denoted by $\bar{y}^{i_1 i_2}$, is chosen as the output desired value. More information about the used fuzzy system are presented in [7]. Moreover, the output of fuzzy system is tuned adaptively according to buffer size as:

$$\Delta Q_A = \alpha \cdot f(x_1,x_2)\cdot R/S_B, \quad (6)$$

where α is a constant coefficient typically about 0.3.

3. Simulation Results

To evaluate the proposed joint video encoding and multiplexing method a simulation was run. A number of 4 long (60 seconds) video sequences with the frame rate of 15 fps, QVGA picture format and different contents captured from the TV were encoded by two methods for a target bit rate of 300kb/s for each. First, the sequences were encoded independently by a rate controller such as in [6]. Second, they were encoded with the proposed method using the same rate controller as IRC for each and the proposed JRC. Then, encapsulating, transmission and reception of DVB-H were simulated on the two sets of encoded bit streams for a constant bit rate channel with a bandwidth of 1200 kb/s. The required decoder buffer size, decoder buffering delay and PSNR of luminance component were measured for two sets. Results of simulation are presented in the Table II. The proposed method provides 38% reduction in the required decoder buffer size and 62% reduction in the decoder buffering delay at the expense of 0.02 dB degradation in quality. Simulation results show that JRC feedback loop can decrease the buffering delay considerably without any cost by very small changes in the QP of encoded bit streams in comparison to independent encoding.

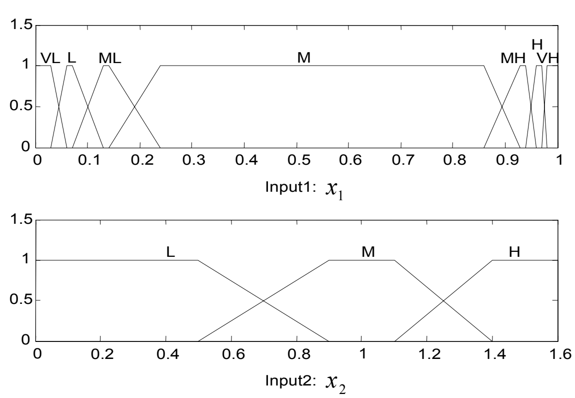

Fig. 2. Membership function of the linguistic variables

TABLE II. SIMULATION RESULTS ON 4 VIDEO SEQUENCES(S1, S2, S3 AND S4) , 300 KB/S FOR EACH , 15 FPS, QVGA, 1200 KB/S CHANNEL.

	Independent Encoded			Joint Encoded		
Sequence	*PSNR "dB"*	*Delay "s"*	*Buffer "kbit"*	*PSNR "dB"*	*Delay "s"*	*Buffer "kbit"*
S1	38.86	0.56	269	38.83	0.23	170
S2	41.41	0.37	190	41.41	0.24	180
S3	37.02	0.39	211	37.00	0.25	176
S4	40.08	1.25	477	40.07	0.27	184
Average	**39.35**	**0.65**	**287**	**39.33**	**0.25**	**178**

4. Conclusion

A method for joint video encoding and StatMax of multiple video bit streams was proposed which decreases the decoder buffering delay and hence the channel changing delay. The proposed method can be used in conjunction with the time-slicing scheme in DVB-H to decrease the channel changing delay even more. The multiplexed bit stream can have different independent quality of services.

5. References

[1] ETSI, "Digital Video Broadcasting (DVB): Transmission Systems for Handheld Terminals," *ETSI standard*, EN 302 304 V1.1.1, 2004.

[2] G. Faria, J. A. Henriksson, E. Stare, P. Talmola, "DVB-H: Digital Broadcast Services to Handheld Devices," *Proceeding of IEEE*, V. 94, No. 1, Jan. 2006

[3] L. Boroczky, A.Y. Ngai, E.F. Westermann, "Joint Rate Control with Look-ahead for Multi-program Video Coding," *Circuits and Systems for Video Technology, IEEE Transactions on*, Vol 10, No. 7, pp. 1159-1163, Oct. 2000.

[4] L. Wang. A. Vincent, "Bit allocation and Constraints for Joint Coding of Multiple Video Programs," *Circuits and Systems for Video Technology, IEEE Transactions on*, Vol. 9, No. 6, pp. 949-959, Sep. 1999.

[5] H. Xiong, J. Sun, S. Yu, C. Luo, J. Zhou, "Design and Implementation of Multiplexing Rate Control in Broadband Access Network TV Transmission System," *Consumer Electronics, IEEE Transactions on*, Vol. 50, No. 3, pp. 849-855, Aug. 2004.

[6] M. Rezaei, M. Gabbouj, I. Bouazizi, "Delay Constrained Fuzzy Rate Control for Video Streaming over DVB-H," *IEEE Int. Con. Intelligent Information Hiding and Multimedia Signal Processing (IIHMSP)*, Pasadena, California, USA, Dec. 2006.

[7] L. X. Wang, *Adaptive Fuzzy System and Control: Design and Stability Analysis*. Englewood Cliffs, NJ: Prentice-Hall, 1994.