Delay Constrained Fuzzy Rate Control for Video Streaming over DVB-H

[1]Mehdi Rezaei, [2]Moncef Gabbouj,[3]Imed Bouazizi,
[1, 2] Tampere University of Technology, [3]Nokia Research Center, Tampere, Finland
mehdi.rezaei@tut.fi, moncef.gabbouj@tut.fi, imed.bouazizi@nokia.com

Abstract

A novel fuzzy video rate control algorithm (RCA) with delay and quality constraints is proposed for video streaming over DVB-H (Digital Video Broadcasting for Handheld terminals) application. In this application a variable bit rate (VBR) bit stream is used to improve the quality and compression performance of encoded video at the expense of buffering delay while there is a holdup as channel changing delay. DVB-H uses a time-sliced transmission scheme to reduce the power consumption for the receiver. The time-slice scheme amplifies the channel changing delay. Two main factors in channel changing delay are the time until a media decoder is refreshed by a random access point such as an IDR (Instantaneous Decoder Refresh) picture in H.264/AVC and the delay to compensate the variation in bit rate. The average decoder refresh delay can decrease by frequent IDR pictures in the bit stream while it drops the quality and compression efficiency remarkably. On the other hand, a VBR video provides higher quality at the expense of higher buffering delay. Hence force, there is a multilateral contradiction between the quality and delay in this application. The proposed RCA has been optimized to provide high quality VBR video with minimum channel changing delay. Simulation results on H.264/AVC video coding standard show that encoded bit streams by proposed RCA strictly obey the delay and quality constraints.

1. Introduction

Digital Video Broadcasting for Handheld terminals (DVB-H) is an ETSI standard specification for bringing broadcast services to handheld receivers [1]. DVB-H is mainly based on the 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. 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 bit rate of audio-visual service. Time-slicing enables a receiver to stay active only a fraction of the time (Burst), while receiving bursts of a requested service [2].

Channel changing delay in DVB-H refers to the time between the channel switching and the start of the media rendering. The channel changing delay for newly-joined recipients consists of several parts including: delay until the start of the desired time-slice or burst, reception duration of a complete burst, delay to compensate the size variation of bursts, delay to compensate the variation of bit rate, delay to compensate the synchronization between the associated streams 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 channel changing delay is the time until a media decoder is refreshed to produce correct output frames, which can be minimized if burst is started with a random access point such as an IDR picture. In DVB-H, the content encoding and the encapsulation to bursts are implemented independently and it is hard to align the location of IDR pictures to the burst boundaries. The average decoder refresh delay can decrease by frequent IDR pictures in the bit stream.

Another main factor in channel changing delay is the required initial buffering delay to compensate the variation in bit rate. While in low delay video communications bit streams with constant bit rates are required, in streaming application more variation in bit rate is acceptable. Normally, a variable bit rate RCA operates closer to the constant quality optimum, which results in a better average quality.

We reviewed several variable RCAs and we proposed a RCA for mobile applications in our previous work [3]. Furthermore, we proposed a fuzzy RCA in [4] optimized for high delay streaming application. In this paper a new RCA optimized for streaming over DVB-H application is proposed.

Proceedings of the 2006 International Conference on Intelligent
Information Hiding and Multimedia Signal Processing (IIH-MSP'06)
0-7695-2745-0/06 $20.00 © 2006 **IEEE**

In video streaming over DVB-H application we need special bit streams. First, we want to utilize the advantage of VBR video to improve the quality and compression performance. Second, frequent IDR pictures are desired in the bit stream to decrease the decoder refresh delay. While in a similar quality an IDR picture consumes a bit budget from 5 to 10 times more than an inter-prediction (P) picture, using frequent IDR pictures drops the compression efficiency and quality remarkably. If we encode the IDR pictures with a quality similar to P pictures to use the advantage of VBR video, it needs a high initial buffering delay to compensate the variation in bit rate. On the other hand, if the IDR pictures are encoded with a quality much lower than P pictures, we do not use the advantage of variable bit rate video. Hence force, there is a strong multilateral contradiction between quality and delay in this application. We need to find an appropriate operating point for the video RCA in VBR between constant quality and constant bit rate to optimize the average quality and delay.

Considering the quality and delay constraints, a fuzzy RCA for video streaming over DVB-H application is proposed in this paper. The proposed RCA has been optimized such that it utilizes the advantage of VBR and also it minimizes the channel changing delay. Simulation results on H.264/AVC video coding standard show that the proposed RCA can provide VBR video with a high level average quality and with small delay. Section 2 of this paper presents details of proposed RCA. Simulation results are provided in Section 3. The paper is closed with conclusions in section 4.

2. Proposed Rate Control Algorithm

The proposed RCA has been optimized to provide VBR bit streams including frequent IDR pictures with a relative constant visual quality and low average buffering delay. Moreover, it has a small degree of complexity from the processing power point of view. The proposed RCA controls the bit rate by adjusting the quantization parameter (QP) on picture base. It utilizes a relative small size virtual buffer, a fuzzy controller and several other tools to calculate the QP for different types of pictures.

While frequent IDR pictures exist in the bit stream and the size of virtual buffer is relative small, the bit allocation to IDR pictures has a remarkable impact on the encoding results so the QP of IDR pictures should be computed carefully. Many theoretic and experimental results have been used in bit allocation to different video pictures. The key concept is to minimize the fluctuation in QP and thereafter in quality while a target average (on all random access points) buffering delay should be achieved.

The proposed RCA is divided in two main parts. The first part utilizes a fuzzy controller to compute the QP for P pictures and the second part uses more other tools to calculate the QP of IDR pictures. The IDR pictures at the scene cuts are handled differently from the normal predefined frequent IDR pictures. Details of proposed RCA are presented below.

2.1. QP of Inter-Prediction Pictures

The QP for a P picture is defined by a fuzzy rate control system. Figure.1 depicts the block diagram of proposed fuzzy rate control system. The fuzzy controller and virtual buffer are the basic parts of the proposed rate controller. Furthermore, a low pass filter (LPF) smoothes the feedback signal from the rate to the fuzzy controller. Moreover, a gain control block (G) tunes the gain of feedback loop adaptively according to the target delay as:

$$G = G_C \times TR / BS, \qquad (1)$$

where TR and BS denote the target bit rate and the buffer size respectively. G_C is a constant coefficient typically about 0.5. The QP for the current P picture (Q_P) is sum of the output of fuzzy controller (Q_F) and the QP used for encoding previous frame (Q):

$$Q_P(i) = Q(i-1) + Q_F(i). \qquad (2)$$

From the controlling point of view the main part of computed QP for P frames is a delayed version of previous used QP and the control or variation of QP is provided by the fuzzy controller.

The virtual buffer used by controller simulates the buffering process of the decoder in the receiving side of streaming and it nearly compromises with hypothetical reference decoder (HRD) models used in different video coding standards. The occupancy of buffer is updated after encoding each video frame as:

$$BF(i+1) = BF(i) - B(i) + (TR / FR), \qquad (3)$$

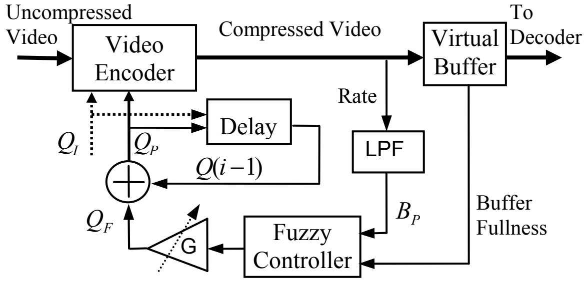

Figure 1. Block diagram of the fuzzy video RCA

Proceedings of the 2006 International Conference on Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP'06)
0-7695-2745-0/06 $20.00 © 2006 IEEE

Where BF denotes the buffer fullness and B shows the number of bits consumed by the previous encoded P or IDR frame. FR indicates the frame rate. The size of virtual buffer is one the most important parameters and it is defined according to the target average buffering delay on all random access points which is set by user. The fuzzy controller has been optimized such that the average buffer fullness during operating time is about 60% to 65% of buffer size. Therefore, the size of virtual buffer is computed as:

$$BS = TR \times TD / 0.65, \tag{4}$$

where BS denotes the size of virtual buffer, TD stands for the target average delay. More details about the fuzzy controller are presented in the sequel.

2.1.1. Fuzzy Controller

The fuzzy controller has basic modifications relative to proposed fuzzy RCA in [4]. The fuzzy controller has two input signals as:

$$x_1 = BF / BS, \tag{5}$$

$$x_2 = \frac{B_P \times FR}{TR}\left(1 + \frac{RX - 1}{IDRI}\right), \tag{6}$$

where B_P denotes the consumed bit budget by the previous encoded P picture. $IDRI$ is the interval of frequent IDR pictures in the bit stream. RX indicates the relative coding complexity of IDR pictures to P pictures and it computed as:

$$RX = \overline{B}_I / \overline{B}_P, \tag{7}$$

where $\overline{B}_I$ and $\overline{B}_P$ denotes the average consumed bit budgets by the all encoded IDR and P pictures respectively. If the previous encoded picture is an IDR picture, the value of B_P in (6) is set by the value of B_I / RX. Moreover, a simple low pass filter (LPF) is used to smooth the variation in B_P before input to fuzzy controller. The impulse response of used LPF is:

$$H(z) = m / \left(m + 1 - z^{-1}\right), \tag{8}$$

where m is a constant value and good results are obtained with m=1.2.

All the fuzzy rules are summarized in the table 1. The content of table 1 specifies the output of the controller. The letters H, L, M, V, X, U and S correspond to linguistic specifications of High, Low, Medium, Very, Extremely, Ultra and Super. As an example from the table it can be expressed as:

if (x_1 is VL and x_2 is M) *Then* (Output is VH).

The input signals are specified by their fuzzy membership functions (MSF). Numbers of 7 and 9 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 Figure 2. The desired central values for the output of fuzzy system correspond to VXL, XL, VVL, VL, L, ML, M, MH, H, VH, VVH, XH, VHX, UH and SH in the table 1 are -6, -5, -4, -3, -2, -1, 0, 1, 2, 3, 4, 5, 6, 7 and 8 respectively. We used a well-known fuzzy system with two inputs using "Product Inference Engine", singleton fuzzifier, and centre average defuzzifier which is:

TABLE 1. SUMMARIZATION OF THE IF-THEN FUZZY RULES

x_2 \ x_1	XL	VVL	VL	L	ML	M	MH	H	VH
VH	SH	UH	VXH	XH	VVH	VH	H	MH	M
H	UH	VXH	XH	VVH	VH	H	MH	M	ML
MH	VXH	XH	VVH	VH	H	MH	M	ML	L
M	XH	VVH	VH	H	MH	M	ML	L	VL
ML	VVH	VH	H	MH	M	ML	L	VL	VVL
L	VH	H	MH	M	ML	L	VL	VVL	XL
VL	H	MH	M	ML	L	VL	VVL	XL	VXL

$$f(x_1, x_2) = \frac{\sum_{i_1=1}^{N_1}\sum_{i_2=1}^{N_2} \bar{y}^{i_1 i_2} \mu_{A_1^{i_1}}(x_1).\mu_{A_2^{i_2}}(x_2)}{\sum_{i_1=1}^{N_1}\sum_{i_2=1}^{N_2} \mu_{A_1^{i_1}}(x_1).\mu_{A_2^{i_2}}(x_2)}, \tag{9}$$

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 $\{\mu_{A_2^{i_2}}(x_2)\}_{1 \le i_2 \le N_2}$ membership functions defined for inputs x_1 and x_2. The centre of output fuzzy set ($B^{i_1 i_2}$),denoted by $\bar{y}^{i_1 i_2}$, is chosen as the output desired value. More information about the fuzzy system (10) are presented in [5].

2.2. QP of IDR Pictures

The QP of IDR picture is computed based on the frame complexity, target bit rate, target average delay, buffer fullness and scene cut information. There are two types of IDR pictures, the normal frequent IDR pictures which are placed in locations with a constant frequency and the IDR pictures which are placed at scene cuts. The QP of all IDR pictures is computed as:

$$Q_I = Q_R + A_C(S + B + D), \tag{10}$$

where Q_I denotes the QP of IDR picture. A_C is a constant value in the range of (0_1) that is used for fine tuning of the rate control according to video content properties. A bigger A_C provides more aggressive control. Q_R is a reference value for the Q_I which is computed differently for the two types of

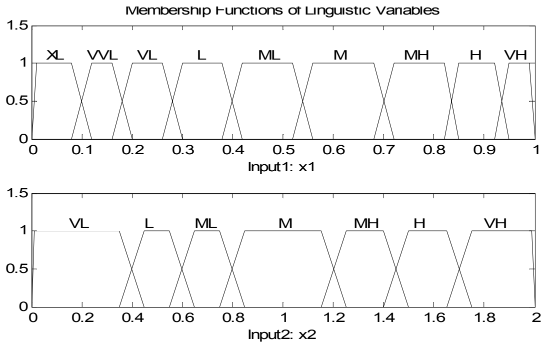

Figure 2. Membership function of the linguistic variables

IDR pictures. The main control of Q_I is imposed by three controlling signals including S, B and D. The signals S and B adapt the QP according to the coding complexity of video picture and buffer fullness respectively. Moreover, D adapts the QP according to the target average delay. While the reference QP defines a reference value for the QP, the controlling signals make some variation around the reference value. More details about the controlling signals and the reference QP are presented in the sequel.

2.2.1. Reference QP

The reference QP for two types of IDR pictures are handled differently. While the reference QP is the main part of final QP, it has an important rule from the quality and delay points of view. For the frequent IDR pictures, the idea is to have a quality as much as close to neighboring pictures. Implementing a low pass filter similar to (8) on the QP of encoded frames gives a local average value which is used as reference QP for the frequent IDR pictures. The IDR pictures at scene cuts may have correlation with previous encoded pictures or not. Therefore, any estimation independently of previous encoded pictures or only based on encoded pictures may lose the bit budget or the quality. From this point of view finding a fit QP for the IDR picture at scene cut is quite challenging. We proposed a method for calculation of QP for IDR pictures at scene cuts in [4]. The QP was computed based on target bit rate and the coding complexity using a rate-distortion model and a coding complexity measure. A simple alternative solution is proposed as:

$$Q_R = \left(\overline{Q} + Q_m\right)/2, \tag{11}$$

where $\overline{Q}$ indicates the average QP on all encoded pictures and Q_m is a constant QP in the middle range e.g. (28-32) for H.264/AVC. The average QP keeps the quality of picture close to quality of previous pictures when there is a correlation between the contents of consequent scenes. The Q_m guarantees a bit budget in the meddle range even there is no correlation between scenes.

2.2.2. Coding Complexity Adaptation

The complexity adaptation signal or S controls the QP of IDR picture according to the coding complexity of picture. Using a simple first-order rate-distortion model, it can be derived as:

$$S = A_S \times \overline{Q}_I \left(X / \overline{X} - 1 \right), \tag{12}$$

where $\overline{Q}_I$ denotes the average of QP on all encoded IDR pictures. A_S namely complexity adaptation factor is a constant value typically about 0.27. X denotes the coding complexity of the IDR picture and $\overline{X}$ is the average of X on all encoded IDR pictures. Different criteria such as variance or used measure in [4] can be used for the coding complexity.

2.2.3. Buffer Fullness Adaptation

The buffer fullness adaptation signal or B controls the QP of IDR picture according to the size and occupancy of buffer. The buffer fullness adaptation signal is calculated as:

$$B = 14 - 38r + 40r^2 - 14r^3, \quad B \le 8, \tag{13}$$

where $r = BF/BS$. Several points are considered in design of function above. First, a more aggressive control is provided when the buffer fullness is closer to critical conditions i.e. underflow or over flow and a looser control is provided when the buffer is far from critical conditions to prevent unnecessary variations in quality. Second, while an IDR frame can consume a bit budget several times more than an inter frame, a buffer close to overflow condition can go to normal state by an IDR frame with a normal QP so a fullness close to over flow is less critical than to under flow. Finally, it is asymmetric concerning to buffer fullness and buffer size to be appropriate to fuzzy controller.

2.2.4. Delay Adaptation

According to (4) the size of virtual buffer is set based on target bit rate and target average delay. The idea is while a smaller buffer size is used for smaller target delay, the allocated bits to IDR pictures should be smaller properly to prevent buffer underflow. The delay adaptation signal biases the QP of IDR picture with a constant value according to the target delay. Using a first-order rate-distortion model the delay adaptation signal can be derived as:

$$D = A_D \times \overline{Q}_P \left(TD_0 / TD - 1 \right), \tag{14}$$

where $\overline{Q}_P$ denotes the average of QP on all encoded P frames. A_D is a constant value typically about 0.055. TD_0 indicates a constant value as a reference point for the delay in which no adaptation is needed. E.g. $TD_0 = 0.75$ when $BS = 1.25 \times TR$.

3. Simulation Results

To evaluate the proposed RCA, a set of simulations were run on different video contents using H.264/AVC coding standard. We repeated a number of the known video sequences including Foreman, Carphone, News, Paris, Salesman, New York, Football, Hall and Glasgow to make suitable long (5 minutes) video sequences. Moreover, four other video sequences including: sport, news, music video and movie contents captured from TV were used for simulation. The video sequences were encoded for different target buffering delays including 0.2, 0.3, 0.4, 0.5, 0.6 and 0.7 second. For each buffering delay, simulations were repeated for different values for the frequency of IDR pictures from 0.25 Hz to 1 Hz. From the delay point of view simulation results show that the encoded bit streams strictly obey the delay constraint.

To evaluate the proposed RCA from the quality point of view we compared the results of proposed RCA with the results of constant QP case as constant quality case. To achieve the exact average target bit rate in the constant QP case, the video sequence is divided into sections and the two sections are encoded by two consequent QPs. The target bit rate is achieved by controlling the number of video frames in each section. A part of simulation results are presented in table2. The results are provided in two different values for the frequency of IDR pictures (1 and 0.5 Hz) for the bit rate of 300 kb/s and the frame rate of 15 f/s and QVGA picture format. While theoretically the constant QP case almost corresponds to the maximum average quality case, simulation results show that the average quality of encoded video by the proposed RCA has a PSNR about 0.22 dB higher than the constant quality case while the controlled bit streams need small buffering delays (about 0.4 s in average). Moreover, we measure the mean and variance of QP used for each case. Small values for the variance of QP (about 3.4 in average) means small variation in quality and it means a high level visual quality for compressed video. This measure for a constant bit rate video can be very high and for the described constant QP case it has an average value about 0.5. Moreover, similar simulations on the mentioned known video sequences for the frame rate of 30 f/s and QCIF picture format, were provided 0.11 dB enhancements in PSNR relative to constant QP case with 0.26 second buffering delay and the value of 2.22 for variance of QP.

TABLE 3. Comparison The Results of proposed RCA with The Constant QS Case in H.264/AVC Encoder with Target Rate of 300kb/s, 15 f/s, QVGA Picture Format.

Video Content	**IDR Pic. Fre.**	**Constant QP**		**Proposed RCA**			
		Mean QP	***PSNR (dB)***	***Mean QP***	***PSNR (dB)***	***Var. QP***	***Mean Delay (s)***
Sport	1.0	32.38	32.10	32.36	32.14	1.48	0.38
	0.5	31.93	32.32	31.98	32.34	1.66	0.42
News	1.0	27.70	35.75	26.86	36.28	3.37	0.47
	0.5	26.91	36.20	26.27	36.64	3.37	0.45
Movie	1.0	29.03	36.98	28.62	37.21	4.95	0.41
	0.5	28.53	37.33	28.10	37.55	5.15	0.35
Music Video	1.0	29.20	35.87	28.81	36.01	3.60	0.42
	0.5	28.86	36.06	28.34	36.26	3.66	0.41
Average		**29.32**	**35.33**	**28.92**	**35.55**	**3.40**	**0.41**

4. Conclutions

A novel fuzzy video RCA with delay and quality constraints is proposed for video streaming over DVB-H application. The proposed RCA were optimized such that it can utilize the advantage of variable bit rate video to improve the quality and compression performance and also it decreases the channel changing delay by optimal bit allocation to frequent IDR pictures in the bit stream and with optimal variation in the bit rate. Simulation results on H.264/AVC video coding standard show that it can provide variable bit rate video with a high level average quality and with a relative small delay.

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 The IEEE*, V. 94, No. 1, Jan. 2006

[3] M. Rezaei, S. Wenger and M. Gabbouj, “Video Rate Control for Streaming and Local Recording Optimized for Mobile Devices,” *IEEE Int. Sym. on Personal Indoor and Mobile Radio Communications (PIMRC’05)*, Berlin, Sep. 2005

[4] M. Rezaei, M. M. Hannuksela, M. Gabbouj, “Low-Complexity Fuzzy Video Rate Controller for Streaming, ” *IEEE Int. Con. on Acoustic, Speech and Signal Processing (ICASSP)*, Toulouse, France, May 2006.

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