Power saving control for the mobile DVB-H receivers based on H.264/SVC standard

Eugeniy Belyaev, Vitaly Grinko and Anna Ukhanova
Saint-Petersburg State University of Aerospace Instrumentation, Russia

Abstract—This paper discusses the utilization of scalable extension of H.264/AVC standard in digital video broadcasting for handheld devices. In this area the problem of mobile receiver power consumption is critically important. This paper amplifies the well-known idea of the time-slicing and allows the receiver to control the trade-off between video quality and power saving in the receiver depending on the priorities.

I. INTRODUCTION

One of the main goals in DVB-H [1] development was the idea of minimizing receiver power consumption by time-slicing, where transmission occurs in bursts [2], [3]. The main idea of it is that receiver operates for a short time interval, and during it the part of the video data is received. Then the receiver powers off radio parts completely while video is constantly decoded. The received data also contains information when to power on again to receive next part of the data. The degree of power consumption depends on the parameters such as peak bit rate, burst bit rate and service bit rate.

In the existing DVB-H systems the MPEG-2 [4] transport stream is used to transmit H.264/AVC [5] encoded video content. There, fixed parameters peak bit rate and burst bit rate define the level of power consumption. Thus, it does not allow to control the level of power consumption at the receiver side. In this paper it is proposed to use the scalable extension of H.264/AVC standard to provide adaptive method to perform trade-off between power consumption and video quality in a mobile hand-held receiver.

As an example, mature technologies are used to demonstrate the benefit of utilizing progressive video codes in digital video broadcasting. DVB-H along with the scalable extension of H.264/AVC standard for video encoding allows receiver to control the level of power consumption depending on the priorities. In the case of single-layer H.264/AVC or MPEG-2 video stream the type of power consumption is set by the codec, but in the proposed case the receiver is responsible for the type of power consumptions. For example, either if there is no charge in the receiver, or the transmission of the video data will continue for a long time, or the high level of power saving is needed, the receiver can switch to the energy economy work conditions. But if the battery is full or we are connected to the power supply, receiver could work with the highest quality of the video data.

Paper is organized as follows. Section II gives a description of the power saving in DVB-H standard. Section III shortly describes the scalable extension of the H.264/AVC standard and compares it with H.264/AVC standard in single-layer mode. Section IV shows how scalable video coding could be used for widening time-slicing approach for power saving. In the following sections practical results of the usage of this idea are shown and the conclusions are made.

II. POWER SAVING IN DVB-H STANDARD

Let’s now take a closer look on the time slicing operation as illustrated in Fig. 1. It can be described with following parameters: Burst Size $B_s$ refers to the number of bits within a burst; Burst Bit rate $B_b$ is the bit rate used by stream inside a burst. Constant Bit rate $C_b$ is the average video bit rate required by the stream when it is not time-sliced. Delta-t is the time instant before second burst cannot occur.

Power saving calculations are presented in [6]. Burst duration $B_d$ is calculated as $B_d = B_s/(0.96B_b)$ (assuming 4% overhead from packet and section headers [2], [3], [6]). Off-time $O_t$ is the time between bursts $O_t = B_s/(0.96C_b) - B_d$. Then, taking into account synchronization time $S_t$ and Delta-t jitter $D_j$, the power saving $P_s$ can be given as

$$P_s = \left(1 - \frac{C_b}{B_b} - 0.96\frac{C_b}{B_b} \left(S_t + \frac{3}{4}D_j\right)\right) \cdot 100\%.$$  \hspace{1cm} (1)

Usually all the parameters - $C_b$, $B_b$, $B_s$, $S_t$ and $D_j$ - depend on the transmission system and there is no chance for the receiver to influence on them. So, regardless of the individual parameters like battery charge all the users receiving data from the same TV channel have equal level of power consumption and, consequently, the quality level of the received video data.
III. SCALABLE EXTENSION OF THE H.264/AVC STANDARD

Scalable extension of the H.264/AVC standard is a highly attractive solution to the problems posed by the characteristics of modern video transmission systems. “Scalability” in this paper means the removal of parts of the bit stream to adapt it to the different needs or preferences of end users as well as to the network conditions.

The main idea of scalable coding is that coder forms the bit stream from several layers: base layer and enhancement layers. The base layer of a bit stream is always coded in compliance with a non-scalable profile of H.264/AVC (single-layer coding). For the next enhancement layers encoding the previous layers (that may include base layer) is needed. Each layer is characterized by its own bit rate and visual quality. Thus, receiver could decode the necessary layer to provide with the necessary bit rate and visual quality dependence.

There exist few ways of the video data processing to form the stream has the properties described above:

- **Temporal scalability.**
- **Spatial scalability.**
- **SNR-scalability.**
- **Combined scalability.**

The next few subchapters will describe each one.

A. Temporal scalability

Temporal scalability provides by using hierarchical coding structures with B-pictures [7], [8]. The pictures of the temporal base layer are only predicted from previous pictures of this layer. The enhancement layer pictures can be bidirectionally predicted by using the two surrounding pictures of a lower temporal layer as references. A picture of the temporal base layer and all temporal refinement pictures between the base layer picture and the previous base layer picture form a group of pictures (GOP). In each GOP, the frame at the lowest level is called the key frame and it is encoded as I- or P- frames.

Each temporal layer is marked by an additional identifier \( T \). \( T \) is equal to 0 for pictures of the temporal base layer and is increased by 1 from one temporal layer to the next.

Fig. 2 shows an example of building hierarchical B-picture structure for the case of GOP containing 8 frames. In this case base temporal layer \( T = 0 \) consists only of the single key (frame 8) of this GOP. Next layer \( T = 1 \) consists of single B-picture (frame 4) that needs two reference frames in forward and backward directions (frame 0, frame 8) from layer \( T = 0 \). In the same manner B-picture (frame 2) in the layer \( T = 2 \) requires also two reference frames (frame 0, frame 4) from layers \( T = 0 \) and \( T = 1 \) accordingly. The following steps are done in the similar manner.

B. Spatial scalability

Spatial scalability allows to decode the streams with the various frame resolution. Each spatial layer corresponds to a supported spatial resolution and is marked by dependency identifier \( D \) [7]. The dependency identifier \( D \) for the base layer is equal to 0, and it is increased by 1 from one spatial layer to the next (see Fig. 3). As for single-layer coding in each spatial layer motion-compensated prediction and intra-prediction are employed. Besides this, so-called inter-layer prediction is also used. Lets take a closer look to the three types of it.

1) **Inter-Layer Motion Prediction:** for spatial enhancement layers a new macroblock type is included. For this macroblock type only a residual signal but no additional side information such as intra-prediction modes or motion parameters is transmitted. When the reference layer macroblock is inter-coded, the enhancement layer macroblock is also inter-coded.

2) **Inter-Layer Residual Prediction:** can be employed for all inter-coded macroblocks. There are two positions. The first one is used when the corresponding reference layer residual signal is up-sampled and used as a prediction for the residual signal for the current macroblock, so that only the corresponding difference signal is coded. The up-sampling of the reference layer residual is done on a transformed block basis in order to ensure that no filtering across transform block boundaries is applied, which could induce visually disturbing signal components.

3) **Inter-Layer Intra-Prediction:** when the macroblock is coded using the reference layer skip mode and the derived prediction mode specifies intra-picture coding, the prediction signal is generated by up-sampling the co-located reconstructed intra-signal of the reference layer.
C. SNR-scalability

H.264/AVC scalable extension supports two types of the SNR-scalability: coarse-grain quality scalable coding (CGS) and medium-grain quality scalability (MGS) [7].

Coarse-grain quality scalable coding can be considered as a special case of spatial scalability with identical picture sizes for base and enhancement layer, so the dependency identifier D is also used here. The same inter-layer prediction mechanisms as for spatial scalable coding are employed, but without using the corresponding upsampling operations and the inter-layer deblocking for intra-coded reference layer macroblocks. When utilizing inter-layer prediction for CGS scalability, a refinement of texture information is typically achieved by requantizing the residual texture signal in the enhancement layer with a smaller quantization step size relative to that used for the preceding CGS layer.

Medium-grain quality scalability is a variation of the CGS approach, allowing to divide it into more numbers of quality layers with the different quality identifier Q.

D. Combined scalability

The general idea for combining spatial, quality, and temporal scalability is illustrated in Fig. 4. For bit stream extracting with necessary tradeoff between bit rate and visual quality decoder need to be informed of the set of the identifiers (D,T,Q) [9]. Note, that to extract the stream with the required parameters all other layers with lower values of the identifiers should have been already extracted.

E. Comparison of single-layer and scalable modes

It is necessary to underline that visual quality of reconstructed video data at the same bit rates for the scalable coding mode is a little bit worse than while using only the main part of the H.264/AVC (see example at Fig. 5). However, for wireless video transmission over error-prone channel scalable video stream could be effectively joined with the unequal error protection. As a result, the quality of the received video data would be better than for the single-layer mode.

IV. PROPOSED POWER-SAVING CONTROL SCHEME

In this paper, the following time-slicing scheme is proposed. Receiver is choosing the necessary (D,T,Q) identifiers. These parameters define the necessary video bit rate C_v. Therefore according to (1) receiver controls trade-off between power saving and video quality [10].

The scheme described above was illustrated for test video sequence “foreman”. Fig. 6 shows time-slicing diagrams for four receiving modes. Table I demonstrates characteristics of received video stream and level of power saving for each mode. Depending on the priorities it is possible to choose receiving modes with high level of power saving (modes a and b) or receiving modes with high level of visual quality (modes c and d).

<table>
<thead>
<tr>
<th>TABLE I</th>
</tr>
</thead>
<tbody>
<tr>
<td>Receiving modes example for test video sequence &quot;foreman&quot;.</td>
</tr>
<tr>
<td>Receiving mode, (D,T,Q)</td>
</tr>
<tr>
<td>a) (D0,T4,Q0)</td>
</tr>
<tr>
<td>b) (D0,T4,Q0–Q1)</td>
</tr>
<tr>
<td>c) (D0,T4,Q0–Q2)</td>
</tr>
<tr>
<td>d) (D1,T4,Q0–Q3)</td>
</tr>
</tbody>
</table>
In addition in proposed scheme it is possible to control relation between the frame rate and signal-to-noise ratio for fixed level of power saving. Therefore, user can tune relation between these parameters to provide the most acceptable visual quality. Thus, the way presented in this paper allow the receiver to control the trade-off between video quality and power saving in the receiver depending on the priorities.

It is also necessary to mention that scalable video decoders require 10-50% more computational resources than single-layer H.264/AVC decoders for the same target resolution and bit rate [12]. But the decoder complexity depends on the number of layers. The minimization of the encoder complexity overhead for scalable coding has become an active research area in the video coding community. So the proposed in the paper approach in addition allows to control the video decoder power consumption on mobile receiver.

ACKNOWLEDGMENT

This article is prepared within the scope of Finnish-Russian University Cooperation Program in Telecommunications (FRUCT, http://www.fruct.org) and the authors would like to thank all experts and organizers of the FRUCT program for their help and contribution.

REFERENCES


Fig. 7. Relation between average Y-PSNR and power saving for video sequences “Foreman”, “Hall” and “Akiyo” in case of burst rate of 3 Mbps.

V. PRACTICAL RESULTS AND CONCLUSION

For our experiments was used the Joint Scalable Video Model (JSVM) reference software v.9.15 [11] which has formed video stream in SNR and temporal scalable modes. Practical results were obtained for three test video sequences “foreman”, “hall” and “akiyo” with 352 × 288 resolution.

Fig. 7 shows relations between peak signal-to-noise ratio and power saving for different receiving modes in case of burst rate of 3 Mbps.