AN ITERATIVE METHOD FOR HYPOTHETICAL REFERENCE DECODER

Z. G. Li1, Nam Ling2, Susanto Rahardja1, Xiao Lin1 and Ping Li1

1 Media Division, Institute for Infocomm Research
21 Heng Mui Keng Terrace, Singapore 119613
Emails:{ezgli, rsusanto, linxiao, liping}@i2r.a-star.org.sg
2 Department of Computer Engineering, Santa Clara University
Santa Clara, CA 95053 USA
Email: nling@scu.edu

ABSTRACT

In this paper, we propose a simple iterative method to verify whether a coded bitstream conforms to a hypothetical reference decoder (HRD). A concept of maximum tolerate delay (MTD) is introduced to study the possible low-delay operation such that we can bound the degree of incorrect motion rendition caused by the variation in end to end delay in the neighborhood of big pictures. Our method can be used in the design of a rate control algorithm to improve the possibility for the coded bitstream that conforms to the HRD.

1. INTRODUCTION.

When a digital video is compressed, the coded bit rate may vary significantly over time. The bitstream is usually transmitted over a channel at a constant bit rate (CBR). The buffer size of an encoded picture buffer (EPB) associated with an encoding process and that of a coded picture buffer (CPB) associated with the corresponding decoding process are finite, and hence the encoder must constrain the bit-rate variation such that a hypothetical reference decoder (HRD) with a predefined buffer size would decode the bitstream without suffering from overflow or underflow [1]. The HRD operates as follows: data associated with coded pictures flow into the CPB according to a specified arrival schedule generated by the stream scheduler. The data associated with each coded picture is removed and decoded instantaneously by the decoding process at CPB removal time. Normally, the coded data can be removed at the computed removal time. The decoding and display times preserve a fixed frame rate. This is the classical constant-delay mode. However, the coded data is removed at a delayed removal time when the amount of coded data for a frame is so large that it prevents removal at the computed removal time. This is an aspect of the low-delay operation [2].

The existing HRD schemes focused on the constant-delay mode of operation [1]. The EPB and the CPB are complementary in this case. However, the HRD can also operate in a low-delay mode (e.g. for video conferencing). Thus the low-delay mode should also be studied when the bitstream is coded.

In this paper, we propose a simple iterative method to verify whether a coded bitstream conforms to the HRD. Our method is applicable regardless of the constant-delay mode. To maintain the degree of incorrect motion rendition caused by the variation in end to end delay in the neighborhood of big pictures, a concept of maximum tolerate delay (MTD) is introduced to bound the possible delayed removal time. Our method can be used in the design of a rate control algorithm to improve the possibility for the coded bitstream that conforms to the HRD.

The rest of this paper is organized as follows. Our iterative method is presented in the following section. It is analyzed in Section 3. An application of our method is given in Section 4. Finally, some concluding remarks are given in Section 5.

2. METHOD FOR BITSTREAM CONFORMANCE TO HRD

In this section, we shall provide a simple iterative method to verify whether a coded bitstream conforms to the HRD.

Let \( \{(s_i, b_i)\} (i = 0, 1, 2, \ldots) \) be the encoder schedule, i.e. at frame time \( s_i \), the encoder instantaneously codes frame \( i \) into \( b_i \) bits and pours the bits into the CPB with buffer size as \( B_{CPB} \). In the CBR case, the EPB drains its accumulated bits into the network at a fixed bit rate \( R \) after an initial \( CPB_{removal\_delay\_offset} = \Gamma_2 \cdot 90000 \) [2]. The encoder should add enough bits to the EPB often enough so that there is no underflow of the EPB. On the other hand, the encoder must not add too many bits to the EPB too frequently so that there is no overflow of the EPB. Let \( t_e \) denote the clock tick, \( s_i \) is computed by [2]

\[
s_i = s_0 + t_e \cdot \text{CPB\_removal\_delay}(i)
\]

The coded bits enter the CPB with buffer size as \( B_{CPB} \) after transmission over the network. For the simplicity, we assume that the transmission delay from the EPB to the CPB is constant. After an initial \( CPB_{removal\_delay} = \Gamma_2 \cdot 90000 \) [2], the decoder starts the decoding process. At frame time \( t_i \), the decoder instantaneously extracts \( b_i \) bits from the CPB and decompresses frame \( i \). \( \{(t_i, b_i)\} (i = 0, 1, 2, \ldots) \) is thus the corresponding decoder schedule. The relationship among \( B_{CPB}, \Gamma_1 \) and \( \Gamma_2 \) is given as follows:

\[
R(\Gamma_1 + \Gamma_2) = B_{CPB}
\]

where \( R, \Gamma_1 \) and \( \Gamma_2 \) are the constant bit rate, the EPB buffer delay
and the CPB buffer delay, respectively.

Let $\Delta_0$ and $\hat{t}_i$ denote the transmission delay for the first packet from the encoder to the decoder and the computed removal time for frame $i$ on the basis of the encoder schedule, respectively. $\hat{t}_i$ is given by

$$\hat{t}_i = s_i + \Gamma_1 + \Gamma_2 + \Delta_0$$

(3)

Let $T_{\min}$ be the lower bound of the difference between two consecutive removal times and $T_{\min}/t_c$ is an integer [2]. Without loss of generality, we assume that

$$\min \{ s_i - s_{i-1} \} \geq T_{\min}$$

(4)

If $t_i$ is fixed at $\hat{t}_i$ as in [1], $B_{se}$ is equal to $B_{sd}$. The EPB and the CPB are complementary. If $t_i$ can be postponed because of a large frame, $B_{se}$ should be much larger than $B_{sd}$. The EPB and the CPB may not be complementary any more.

To maintain the degree of incorrect motion rendition caused by the variation in end to end delay in the neighborhood of big pictures, a concept of MTD is introduced to bound the possible delayed removal delay. The MTD is assumed to be $\Gamma_3$.

For any actual transmission delay $\Delta'$, we assume that

$$|\Delta' - \Delta_0| \leq \Gamma_4$$

(5)

For the simplicity, we assume that $\Gamma_3$ is not less than $\Gamma_4$, $\Gamma_3/t_c$ and $\Gamma_4/t_c$ are two integers.

The EPB size is given by

$$B_{se} = B_{sd} + R\Gamma_3 - R\Gamma_4$$

(6)

and $t_i$ is required to satisfy that

$$\hat{t}_i \leq t_i \leq \hat{t}_i + \Gamma_3$$

(7)

Given an EPB with parameters $(R, B_{se}, \Gamma_1)$ and a CPB with parameters $(R, B_{sd}, \Gamma_2)$, the coded bitstream is said to conform to the HRD if the following four requirements are satisfied: 1) No overflow of the EPB, no underflow and no overflow of the CPB; 2) Removal time consistency with the predefined accuracy; 3) Big picture removal time satisfying (7); and 4) Maximum removal rate from the CPB consistency with the predefined upper bound [2].

The problem studied in this paper is formulated as follows:

Given an EPB with parameters $(R, B_{se}, \Gamma_1)$ and a CPB with parameters $(R, B_{sd}, \Gamma_2)$, a coded bitstream is said to conform to the HRD if the following four requirements are satisfied:

1. No overflow of the EPB, no underflow and no overflow of the CPB;
2. Removal time consistency with the predefined accuracy;
3. Big picture removal time satisfying (7); and
4. Maximum removal rate from the CPB consistency with the predefined upper bound [2].

The relationship among $s$, $t$, $\Gamma$, $R$ and $L$ is given by the following three equations:

$$s_i = s_{i-1} + \Gamma_3 - \Gamma_4$$

(8)

$$L = R(s_1 - s_0) - R\Gamma_1 + R\Gamma_3$$

(9)

$$U = R\Gamma_2 - R\Gamma_4$$

(10)

**Verification** After coding the $i$th frame, we first check whether the coded bitstream conforms to the HRD.

If $(b_i > B_{sd})$ or $(b_i > U + R\Gamma_3)$, Flag is set as 0 and terminate the verification. Otherwise, the possible delayed removal time $t_i$ is set according to the following two cases:

**Case 1** When $b_i > U$, the final arrival time $t_i^{final}$ exceeds $\hat{t}_i$, the delayed removal time $t_i$ is set as

$$t_i = \max \{ t_{i-1} + T_{\min}, \hat{t}_i \}$$

(11)

where the time instant $\hat{t}_i$ is computed by

$$\hat{t}_i = \hat{t}_i + \left[ \frac{b_i - U}{Rt_c} \right] t_c$$

(12)

with $\left[ a \right]$ denoting the smallest integer greater than $a$.

**Case 2** When $b_i \leq U$, $t_i$ is set as

$$t_i = \max \{ t_{i-1} + T_{\min}, \hat{t}_i \}$$

(13)

$L$ and $L$ are iteratively reset according to the following two cases:

**Case 1** When $(b_i < L)$, there is underflow of the EPB. $L$ and $U$ are reset by

$$L = R(s_{i+2} - s_{i+1}) + R\Gamma_3$$

(14)

$$U = B_{sd} - R\Gamma_4$$

(15)

**Case 2** When $(b_i \geq L)$, there is no underflow of the EPB. $L$ and $U$ are reset by

$$L = L + R(s_{i+2} - s_{i+1}) - b_i$$

(16)

$$U = U + R(s_{i+1} - s_i) - b_i$$

(17)

$L$ and $U$ can be used to bound the bit rate of the next coded frame to improve the possibility of the coded bitstream to conform to the HRD. This will be discussed in Section 4.

**Remark 1** Clearly, the decoding and display times do not preserve a fixed frame rate because of (11).

### 3. ANALYSIS OF OUR METHOD

Two cases are studied when our iterative method is analyzed, one is that there is no underflow of the EPB and the other is that there is underflow of the EPB.

The available channel bandwidth is supposed to be maximally utilized. In other words, the bits in the EPB will be sent out continuously if there is no underflow of the EPB.

**Case 1.** There is no underflow of the EPB.

Let $s_i^{initial}$, $s_i^{final}$, $t_i^{initial}$ and $t_i^{final}$ denote the time that the first bit of frame $i$ leaves the EPB, the time that the last bit of frame $i$ leaves the EPB, the time that the first bit of frame $i$ arrives to the CPB, and the time that the last bit of frame $i$ arrives to the CPB, respectively. The relationship between the $s_i^{initial}$ and $s_i^{final}$ is given by the following three equations:

$$s_0^{initial} = \Gamma_1 + s_0$$

(18)

$$s_i^{final} = s_i^{initial} + \frac{b_i}{R}$$

(19)

$$s_{i+1}^{initial} = \max \{ s_{i+1}^{final}, s_{i+1} \}$$

(20)

The relationship among $s_i^{initial}$, $s_i^{final}$, $t_i^{initial}$ and $t_i^{final}$ is de-
scribed by
\[
\begin{align*}
    t_{i,initial}^i &= s_{i,initial}^i + \Delta t_{i,initial}^i \\
    t_{i,final}^i &= s_{i,final}^i + \Delta t_{i,final}^i
\end{align*}
\]  
(21)  
(22)

Evolution of the EPB and the CPB can be described in terms of buffer fullness. Let \(B_{i,\text{EPB}}^+, B_{i,\text{EPB}}^-, B_{i,\text{CPB}}^+, B_{i,\text{CPB}}^-\) denote the buffer fullness of the EPB just before the \(i\)th frame is added to the EPB, that of the CPB just before the \(i\)th frame is added to the CPB, and that of the CPB just after the \(i\)th frame is removed from the CPB, respectively. The dynamic of the EPB is described by
\[
\begin{align*}
    B_{i,e}^- &= \begin{cases} 
    B_{i-1,e}^+ - R(s_i - s_{0,\text{initial}}) & i \leq i_0 \\
    B_{i-1,e}^+ - R(s_i - s_{i-1}) & i > i_0 + 1 
    \end{cases} \\
    B_{i,e}^+ &= B_{i,e}^- + b_i \\
    B_{0,e}^- &= 0
\end{align*}
\]
where \(i_0\) is defined as follows:
\[
i_0 = \arg\max \{ s_k \} \text{ such that } (s_i - s_0) \leq \Gamma_1
\]  
(23)
and the dynamic of the CPB is represented by
\[
\begin{align*}
    B_{i,d}^- &= B_{i-1,d}^+ + R(t_i - t_{i-1}) \\
    B_{i,d}^+ &= B_{i,d}^- - b_i \\
    B_{0,d}^- &= \Gamma_2
\end{align*}
\]
To fully utilize the available channel bandwidth, \(s_{i+1,initial}^i\) is equal to \(s_{i,final}^i\) for each \(i\). A lower bound is then computed as
\[
b_i \geq R(s_1 - s_0) - R\Gamma_1 + \sum_{j=0}^{i-1} (R(s_{j+2} - s_{j+1}) - b_j); \forall i
\]  
(24)
To guarantee that there is no overflow of the EPB, an upper bound is computed as
\[
b_i \leq B_{se} - R\Gamma_1 + \sum_{j=0}^{i-1} (R(s_{j+1} - s_j) - b_j); \forall i
\]  
(25)
To guarantee that there is no underflow of the CPB if the \(i\)th frame is removed at \(t_i\), \(t_i\) should be greater than \(t_{i,final}^i\) for all \(i\). Another upper bound is then computed by
\[
b_i \leq \Gamma_2 - R\Gamma_4 + \sum_{j=0}^{i-1} (R(s_{j+1} - s_j) - b_j) \leq U; \forall i
\]  
(26)
Since the removal time \(t_{i,final}^i\) may exceed \(t_i\) due to a frame with a large size, its size must be such that it can be removed from the CPB without overflow. To guarantee this, \(b_i\) is required not to be greater than \(B_{se}\).
To guarantee that there is no overflow of the CPB at the possibile delay removal time \(t_i\), the coded bitstream should satisfy that
\[
b_i \geq R(s_1 - s_0) - R\Gamma_1 + R\Gamma_3 + \sum_{j=0}^{i-1} (R(s_{j+2} - s_{j+1}) - b_j) \equiv L; \forall i
\]  
(27)

**Case 2** There is underflow of the EPB.
Suppose that the first underflow of the EPB appears during the interval \([k_1 - 1, k_1]\). Similar to Case 1, there is no overflow and no underflow of the EPB within the interval \([0, k_1 - 1]\) if
\[
b_i \leq B_{se} - R\Gamma_1 + \sum_{j=0}^{i-1} (R(s_{j+1} - s_j) - b_j);
\]  
(28)
\[
b_i \geq R(s_1 - s_0) - R\Gamma_1 + \sum_{j=0}^{i-1} (R(s_{j+2} - s_{j+1}) - b_j)
\]  
(29)
hold for all \(i\)’s from 0 to \((k_1 - 2)\) and
\[
b_{k_1-1} < R(s_1 - s_0) - R\Gamma_1 + \sum_{j=0}^{k_1-2} (R(s_{j+2} - s_{j+1}) - b_j)
\]  
(30)
To guarantee that there is no underflow of the CPB if the \(i\)th frame is removed at \(t_i\) and no overflow because of the possible delayed removal time, \(b_i\) should satisfy the following two inequalities:
\[
b_i \leq R\Gamma_2 - R\Gamma_4 + \sum_{j=0}^{i-1} (R(s_{j+1} - s_j) - b_j) \equiv U
\]  
(31)
Meanwhile, to guarantee that there is no overflow of the CPB at the possible delay removal time \(t_i\), the coded bitstream should satisfy that
\[
b_i \geq R(s_1 - s_0) - R\Gamma_1 + R\Gamma_3 + \sum_{j=0}^{i-1} (R(s_{j+2} - s_{j+1}) - b_j) \equiv L; \forall i
\]  
(32)
There is neither underflow nor overflow of both the EPB and the CPB within time \([k_1, k_1 + 1]\) if \(b_{k_1}\) satisfies that
\[
R(s_{k_1+1} - s_{k_1}) \leq b_{k_1} \leq B_{sd}
\]  
(33)
Suppose that the second underflow of the EPB occurs within the interval \([k_2 - 1, k_2]\). The dynamic of the EPB is described by
\[
\begin{align*}
    B_{i,e}^- &= B_{i-1,e}^+ - R(s_i - s_{i-1}) - \sum_{j=0}^{i-1} (R(s_{j+2} - s_{j+1}) - b_j); k_1 \leq i \leq k_2 - 2 \\
    B_{i,e}^+ &= B_{i,e}^- + b_i; k_1 \leq i \leq k_2 - 1 \\
    B_{0,e}^- &= 0
\end{align*}
\]
and the dynamic of the CPB is represented by
\[
\begin{align*}
    B_{i,d}^- &= B_{i-1,d}^+ + R(t_i - t_{i-1}) \\
    B_{i,d}^+ &= B_{i,d}^- - b_i \\
    B_{0,d}^- &= B_{sd}
\end{align*}
\]
Similarly, to guarantee that there is no underflow of the CPB if the \(i\)th frame is removed at \(t_i\) and no overflow because of the
possible delayed removal time, \( b_i \) should satisfy that
\[
\begin{align*}
  b_i & \leq B_{sd} - R \Gamma_4 + \sum^{i-1}_{j=k_1} (R(s_{j+1} - s_j) - b_j) \leq U \quad (34) \\
  b_i & \leq B_{sd} \quad \forall i = k_1 + 1, \ldots, k_2 - 2 \quad (35)
\end{align*}
\]
To guarantee that there is no overflow of the CPB at the possible delay removal time \( t_i \), the coded bitstream should satisfy that
\[
\begin{align*}
  b_i & \geq R(s_1 - s_0) + R \Gamma_3 + \sum^{i-1}_{j=k_1} (R(s_{j+2} - s_{j+1}) - b_j) \geq L \quad \forall i \\
  b_i & \geq R(s_{k_1+1} - s_{k_1}) + \sum^{k_1-2}_{j=k_1} (R(s_{j+2} - s_{j+1}) - b_j) \quad (38)
\end{align*}
\]
where \( k_i = k_1 + 1, \ldots, k_2 - 2 \). To improve the possibility for the coded bitstream that conforms to the HRD, the target bitrate \( T_i \) is further bounded as follows:
\[
T_i = \max \{ \bar{L}, \min \{ T_i, \bar{U} \} \} \quad (42)
\]
where \( \bar{L} \) and \( \bar{U} \) are given by
\[
\begin{align*}
  \bar{L} & = (1 + \omega)L \\
  \bar{U} & = (1 - \omega)U
\end{align*}
\]
In the above equations, \( \omega \) is a constant which depends on the choice of a basic unit [5]. All macroblocks in a basic unit share a quantization parameter. The basic unit can thus be used to obtain a good tradeoff between the coding efficiency and the bits fluctuation. A large \( \omega \) can be chosen for a large basic unit. [6] discusses the details on rate control part and the concept of basic unit and includes extensive experimental results on our scheme.

5. CONCLUSION

A simple iterative method has been proposed in this paper to verify whether a coded bitstream conforms to a hypothetical reference decoder (HRD). Our method is applicable regardless of the constant-delay mode. A concept of maximum tolerate delay (MTD) is introduced to bound the degree of incorrect motion rendition caused by the variation in end to end delay in the neighborhood of big pictures. Our method can be used in the design of a rate control algorithm to improve the possibility for the coded bitstream that conforms to the HRD.

References