Signal Processing: *Image Communication* 28 (2013) 20–33

Contents lists available at SciVerse ScienceDirect

Signal Processing: *Image Communication*

journal homepage: www.elsevier.com/locate/image

Rate control for consistent visual quality of H.264/AVC encoding ☆

Long Xu [a,b,]*, Sam Kwong [a], Hanli Wang [c,d], Debin Zhao [e], Wen Gao [f]

[a] Department of Computer Science, City University of Hong Kong, Kowloon, Hong Kong
[b] School of Automation and Electrical Engineering, University of Science and Technology Beijing, Beijing 10083, China
[c] Department of Computer Science and Technology, Tongji University, Shanghai 201804, China
[d] Key Laboratory of Embedded System and Service Computing, Ministry of Education, Tongji University, Shanghai 200092, China
[e] Department of Computer Science and Technology, Harbin Institute of Technology, Harbin 150001, China
[f] School of Electronics Engineering and Computer Science, Peking University, Beijing 100871, China

ARTICLE INFO

Article history:
Received 7 October 2011
Accepted 7 October 2012
Available online 29 October 2012

Keywords:
Video coding
Sliding window
Rate control
Video streaming
Consistent/smooth visual quality

ABSTRACT

In this paper, a novel rate control scheme with sliding window basic unit is proposed to achieve consistent or smooth visual quality for H.264/AVC based video streaming. A sliding window consists of a group of successive frames and moves forward by one frame each time. To make the sliding window scheme possible for real-time video streaming, the initial encoder delay inherently in a video streaming system is utilized to generate all the bits of a window in advance, so that these bits for transmission are ready before their due time. The use of initial encoder delay does not introduce any additional delay in video streaming but benefits visual quality as compared to traditional one-pass rate control algorithms of H.264/AVC. Then, a Sliding Window Buffer Checking (SWBC) algorithm is proposed for buffer control at sliding window level and it accords with traditional buffer measurement of H.264/AVC. Extensive experimental results exhibit that higher coding performance, consistent visual quality and compliant buffer constraint can be achieved by the proposed algorithm.

1. Introduction

In recent years, we have witnessed an exponential increase of real-time multimedia applications. The real-time requirements of these applications such as videophone, videoconference, Video On Demand (VOD), and TV broadcasting, pose a big challenge for their integration into IP networks. Due to limited network resources and buffer requirements, significant variations of picture quality may be observed when high motion object or scene change occurs. Thus, rate control should not only control bit rate, but also achieve perceptually consistent visual quality. Most of the popular rate control schemes of existing video coding standards are model-based, such as TM5 for MPEG-2 [1], TMN8 [2–4] for H.263 and VM8 for MPEG-4 [5]. For H.264/AVC, the rate control scheme [7] was adopted as the reference algorithm. The key of these model-based rate control schemes is to find the relation between rate and distortion, namely Rate-Distortion (*R-D*) model. Traditionally, both rate and distortion models are formulated as a function of Quantization Parameter (QP), so that traditional *R-D* models can be expressed as functions of QP, such as the conventional *R-D* model quadratic in D inverse [6]. Another novel *R-D* model is the ρ-domain *R-D* model [10], which directly models the relation between bit rate and the percentage of zero coefficients after quantization.

Traditional rate control schemes are concentrated on the accuracy of bit rate controlling between the target bit

☆ The work was partially supported by a grant for the Research Grants Council of the Hong Kong (Project CITYU115109), the Major State Basic Research Development Program of China (973 Program) under Grant 2009CB320905, the Shanghai Pujiang Program under Grant 11PJ1409400, the National Natural Science Foundation of China under Grants 61202242, 61102059, the Fundamental Research Funds for the Central Universities under Grant 0800219158.

* Corresponding author at: Department of Computer Science, City University of Hong Kong, Kowloon, Hong Kong. Tel.: +852 34425253.
E-mail addresses: lxu@jdl.ac.cn, xulong@ee.cuhk.edu.hk (L. Xu).

0923-5965/$ - see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.image.2012.10.001

rate and the final coded bit rate. In [7–9], the *R-D* model quadratic in D inverse and the linear prediction of Mean Absolute Distortion (MAD) are presented for H.264/AVC. Meanwhile, Hypothetical Reference Decoder (HRD) is being considered in the process of rate control. Thus, the match of bit rate controlling and compliant buffer constraint can be both satisfied. However, there is usually a significant picture quality fluctuation by such methods, the reason of which lies in the unavailability of the information of future frames for bit allocation, actual MAD values for calculating QPs before encoding the corresponding basic units, as well as of the existing buffer control strategy.

Conventionally, in addition to bit rate match and buffer constraint, rate control should also produce consistent visual quality [11–16] under given bit rate and buffer constraints. In [13], a new H.263+ rate control algorithm was proposed to support the Variable Bit Rate (VBR) encoding through frame rate adjustment. It changes the frame rate adaptively based on the motion information in a sliding window to reduce the image quality variation between adjacent frames. In [14], the sliding window concerned how many frames previously coded were used to update the rate-distortion model and model parameters. If the video content complexity changes significantly, a smaller window with more recent data would be used. In VBR rate control, the temporally consistent visual quality can be achieved by adjusting the output bits to the varying characteristics of video content. One of the typical VBR applications is the Digital Versatile Disk (DVD) [17,18]. Such an application relies on a VBR encoding system to achieve maximally consistent visual quality for the entire sequence, under the constraint of total storage capacity. It is a challenge because the video content may vary significantly from one frame to another. In practice, the researchers tried to solve this problem by two-pass or multi-pass encoding processes [17–20], where the first pass or several preceding passes are used to track the variation of characteristics of video content and thus the available bit resources can be distributed appropriately to various video segments, e.g., more bits are allocated to complex scenes or pictures. For the storage of video data without real-time requirement and buffer constraint, such as DVD, the coding scheme with multi-pass encoding is preferred for optimal coding efficiency and consistent visual quality. However, multi-pass encoding is not suitable for real-time encoding applications, so some one-pass VBR rate control algorithms are proposed [21–24]. On the other hand, in the client-server architecture of video streaming applications over networks, such as VOD and TV broadcasting, the videos may have been compressed already and are provided to users on demand, so there is no real-time requirement on encoding. But the buffer/time delay constraints should be considered due to the limited channel bandwidth, buffer capacity of network devices as well as the maximum tolerated time-delay requirement of terminal devices or end-users. The video streaming poses a big challenge to rate controls due to unstable networks and unexpected error of transmissions. And, rate controls for video streaming applications [25–28] have been becoming more and more popular in both research and industry fields.

In this paper, we first propose a new rate control scheme with sliding window basic unit for video streaming. A sliding window refers to a segment of video sequence which includes several successive frames or several Group-of-Pictures (GOPs). In addition, it moves forward by one frame each time along a video sequence. The initial encoder delay T_e inherently in video streaming systems is utilized to generate the encoding bits of a window for transmission purpose in advance, which makes the proposed rate control scheme possible for real-time video streaming applications. Then, a Sliding Window Buffer Checking (SWBC) method is proposed for elegant buffer control at sliding window level instead of frame level by traditional methods. It should be noted that the usage of sliding window in this work is different from those in [13,14]. The proposed algorithm in this work is also different from [21] which proposed a theoretical model handling the tradeoff between buffer constraint and picture quality fluctuation, but lacked a practical buffer control mechanism.

The rest of this paper is organized as follows. Section 2 gives a brief introduction of HRD and the Leaky Bucket (LK-B) model in H.264/AVC. In Section 3, the proposed SWBC and sliding window rate control algorithm are presented in detail. The simulation results of the proposed algorithm are given in Section 4, and the last section concludes the paper.

2. HRD in H.264/AVC

For video streaming applications, the buffer constraint is required, which is realized by HRD for H.264/AVC buffer management [29–31]. The goal of HRD is to ensure that the coded bitstream neither overflow nor underflow the decoder buffer under a given channel rate. As the heart of HRD, the LK-B [29] model is represented by a triple parameter (R,F,B), where F denotes the initial buffer fullness with F_d representing for decoder and F_e for encoder, B is the bucket capacity, and R is the leaking bit rate of bucket. The LK-B is actually a direct metaphor for the decoder's input buffer or encoder's output buffer, i.e., the queue between the decoder/encoder and the communication channel. At the encoder, we assume that the coded bits of the ith frame are poured into LK-B at the time s_i. After that, the coded bits are transmitted through communication channel with the given bit rate and enter into the decoder buffer. And then, at time $t_i = s_i + \Delta$ (Δ is the channel delay from server to client), the decoder removes the bits of the ith frame from the decoder buffer and decompresses it.

According to the LK-B model, the time scheduling including initial arrival time, removal time and decoding time of each frame can be deduced from the triple parameters (B,F,R) or (D,T_d,R). The relation between T_e/F_e, T_d/F_d and D/B is illustrated in Fig. 1, where the vertical axis shows the accumulated bits of encoder/decoder buffer and the horizontal axis represents time. The staircases in Fig. 1 are due to the behavior of bits arrival and removal of