J. Vis. Commun. Image R. 34 (2016) 204–218

Contents lists available at ScienceDirect

J. Vis. Commun. Image R.

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

Content-adaptive parameters estimation for multi-dimensional rate control ☆,☆☆

Mingliang Zhou [a], Bo Li [a,b], Yongfei Zhang [a,b,*]

[a] Beijing Key Laboratory of Digital Media, School of Computer Science and Engineering, Beihang University, Beijing 100191, China
[b] State Key Laboratory of Virtual Reality Technology and Systems, Beihang University, Beijing 100191, China

ARTICLE INFO

Article history:
Received 8 September 2015
Accepted 16 November 2015
Available online 2 December 2015

Keywords:
Multi-dimensional rate control
Q-R-TQ
Video content-adaptive
Parameter estimation
Model coefficient
Update duration
Sliding window
Initial frame rate

ABSTRACT

Multi-dimensional rate control schemes have been recently utilized to adapt video streams to dynamic network conditions and heterogeneous devices. However, current multi-dimensional rate control methods, which estimate the model coefficients using fixed update duration, usually yield inaccurate parameters for dynamically changing video content. To address this problem, a content-adaptive parameters estimation scheme is proposed for multi-dimensional rate control. Firstly, we propose to estimate the parameters using dynamical update duration based on video content and the update duration of the model coefficients is determined by jointly considering the varying picture complexity and feedback information from the actual encoding results, which can improve the model parameter estimation accuracy. Secondly, a coarse-to-fine initial parameter calculation method is proposed to refine the initial frame rate according to the channel condition and the video sequence characteristics. Extensive experimental results show that the proposed solutions outperform the state-of-the-art schemes, especially for video sequences with high temporal and spatial complexity. Furthermore, our algorithm also slightly reduces the computational complexity as compared to related algorithms.

1. Introduction

1.1. Multi-dimensional rate control

The video streaming market is growing rapidly in recent years, partly because of the emergence and popularity of more and more heterogeneous devises and services. Therefore, service providers have to fit the video stream to variable bandwidth communication channels. Rate control plays a critical role in video coding and transmission. In conventional rate control schemes [1–6], only the quantization parameter (QP) is considered, wherein an increasing QP leads to a decrease in bit rate and an increase in quantization errors.

As the bandwidth of dynamic network conditions and heterogeneous devices are quite limited and change violently, conventional rate control schemes are no longer suit for these applications. Take the extremely low bitrate applications as an example, adjusting only the QP still cannot achieve the target bitrate. On the other hand, research has shown that frame skipping or image resizing might help in these scenarios [7].

Recently, several new algorithms related to multi-dimensional rate control scheme have been proposed. Multi-dimensional rate control schemes, which adjust not only the quantization parameter as in conventional rate control schemes, but also the frame rate or frame size, can better adapt to large-range dynamically changing networks conditions.

At sender, a pre-processor is applied to determine the optimal frame rate (FR), frame size (FS) or quantization parameters (QP), so as to get the best video quality under given total bit rate budget. It usually takes the sub-sampling method to get best frame size, and frame skipping to get the best frame rate. At the receiver end, the lower resolution sequences are converted into higher resolution through up-sampling (using the 6-tap half-sample interpolation filter) and frame repeating [7,8].

1.2. Related works

In this section, the state-of-the-art multi-dimensional rate control schemes are briefly reviewed.

To evaluate video quality, a proper assessment metric is necessary [9,10]. PSNR, SSIM and SAE [11] do not adequately reflect

☆ This paper has been recommended for acceptance by M.T. Sun.

☆☆ This work was partially supported by the National Science Fund for Distinguished Young Scholars (No. 61125206), National Hi-Tech Research and Development Program (863 Program) of China (2014AA015102) and National Natural Science Foundation of China (61272502).

* Corresponding author at: Beijing Key Laboratory of Digital Media, School of Computer Science and Engineering, Beihang University, Beijing 100191, China.

E-mail address: zhangyf.ac@gmail.com (Y. Zhang).

http://dx.doi.org/10.1016/j.jvcir.2015.11.011
1047-3203/

perceived video quality when changes in spatial resolution and frame rate are involved, and are therefore not adequate for assessing quality in a multi-dimensional rate control scheme. Several studies have investigated the impact of frame rate on perceptual video quality [12–15]. Feghali et al. proposed the Quality Metric (QM) [16] to assess the video quality which is formulized as a function of frame rate and PSNR, the average PSNR of each skipped frame reconstructed by repeating the previous frame. QM has been proved to be consistent with the subject quality and used to assess video quality in numerous studies [17,18]. Regarding to up and down conversions of frame rate, sophisticate interpolation methods instead of simple repeating were proposed in the literature [19–21]. Chien et al. [22] proposed an applicable quality evaluation model which gives a parameter prediction method using scene change detection. However, this algorithm estimates the model coefficients by using fixed update duration, which means the update duration for coding parameters is fixed without fully considering the dynamically changing video content.

Two perceptual models were proposed in [23] to map FR and FS into quality. These models consider additional video content parameters such as spatial and temporal activity to improve the performance in terms of correlation to the subjective quality. But it doesn't consider how to adjust FR or FS in such a way that the overall quality is maximized, under given bandwidth.

All the methods mentioned above give a quality assessment model only, without discussion of the relationship between the rate and optimal rate-control parameters. How to use the quality model, together with a rate model, to determine the optimal frame rate and frame size under a target bit rate is still a critical problem in multi-dimensional rate control.

Some recent work [24–30] has investigated a subset of this problem, where only FR and QP are considered (with fixed FS), based on quality and rate models as functions of FR and QP. The rate and quality models are analytically tractable, each requiring only few content-dependent parameters. Both models fit the measured data very accurately, with high Pearson Correlation (PC) and small root mean square error (RMSE). However, these works estimate the rate and quality model coefficients using fixed update duration [26]. It usually collects the video characteristics during video encoding for each processing time unit (fixed update duration, like 25 frame). In reality, however, video content changes over the time and is not stationary and quality model coefficients derived using fixed update duration would cause sub-optimal coding parameters thus degraded video quality, especially for high spatial–temporal complexity video sequences. Another important factor influencing the multi-dimensional rate control performance, i.e. how to determine the initial frame rate, is not well addressed. Some prior works have explored the problem of initial QP selection [31–33]. To the best of our knowledge, no comparable works about initial FR calculation have ever been proposed. Traditionally, the initial FR is determined only depending on the bit rate constraint as Chien does [22]. Although this initialization scheme is simple, yet it is not accurate enough. A better initialization scheme should not only depend on bit rate constraint, but also on the content of the video content characteristics.

1.3. Contribution of this paper

The contribution of this paper is two folds. First, to improve the parameter estimation accuracy of the multi-dimensional rate control model, we propose a content-adaptive model coefficients estimation scheme. We investigate the problem of related work in [22–30]. In contrast with fixed update duration therein, we estimated dynamically the model coefficients according to the time-varying video content, which improves the model coefficients estimation accuracy, thus the video quality. Second, we focus on how to determine the initial parameters of the video coding, and propose a coarse-to-fine initial frame rate calculation method, which can significantly reduce the video quality fluctuation commonly incurred in existing schemes [22,26]. Experimental results demonstrate that our proposed algorithms can adapt to the video content changes very well. We believe such a study brings us a new knowledge into the topic of multi-dimensional rate control.

The rest of the paper is organized as follows. Some observations and motivations supporting our proposed algorithm are provided in Section 2. Section 3 describes the proposed content-adaptive model coefficients estimation scheme and the coarse-to-fine initial frame rate calculation method, respectively. Experiment results are presented in Section 4, and Section 5 concludes the paper.

2. Observations and motivations

Some observations and motivations related to our proposed approach are discussed in detail as follows.

2.1. Parameters estimation

Quality-Rate models are commonly used in rate control to represent the mathematical relationship between the required bits R, the temporal resolution (frame rate, FR) and quantization parameter (QP) in multi-dimensional rate control. One of such models is the well-known Quality-Rate-Temporal resolution and Quantization (Q-R-TQ) model proposed by Wang et al. in [24], which use the pair of (q, f) to design an encoded video stream with quantization q and frame rate f. Given a video stream at the full rate $R_{\max}$ and best quality $Q_{\max}$ corresponding to a minimum QP $q_{\min}$ and a maximum FR $f_{\max}$, where $Q_{\max}$ is normalized to 1, the quality and bitrate for a sub stream with $q \geqslant q_{\min}$ and $f \leqslant f_{\max}$ can be modeled as [24]:

$$Q(q,f) = Q_{\max} \frac{e^{-c\frac{q}{q_{\min}}}}{e^{-c}} \frac{1 - e^{-d\frac{f}{f_{\max}}}}{1 - e^{-d}} \tag{1}$$

$$R(q,f) = R_{\max} \left(\frac{q}{q_{\min}}\right)^{-a} \left(\frac{f}{f_{\max}}\right)^{b} \tag{2}$$

where a, b, c, d are four model coefficients, which can be estimated from features abstracted from the original video. f and q denote the optimal frame rate and optimal quantization parameter, respectively. The Q-R-TQ model is usually employed to achieve optimal assignment of the frame rate and quantization parameter based on the given bit-allocation.

To prevent from being confusing and misleading, we emphasize that f and q refer to coding parameters which need to be decided according to the video content, while a, b, c and d represent the model coefficients in the model shown as in Eqs. (1) and (2).

In Q-R-TQ model, the four model coefficients, i.e. a, b, c and d can be obtained through the prediction model $P = HF$, where F represents the content features, P denotes the model coefficients set $\{a, b, c, d\}$, and H contains only rows corresponding to the quality model coefficients, which need to be pre-designed for different coding structures. In this procedure, three features, δ_{FD}, u_{MVM}, $\eta(\mu_{MVM}, \mu_{MDA})$ can accurately predict the four quality model coefficients, where FD is frame difference, MVM is Motion vector magnitude and MDA is motion direction activity, $\eta(\mu_{MVM}, \mu_{MDA}) = \mu_{MVM}/\mu_{MDA}$ where μ_{MVM} represents the average value of MVM, μ_{MDA} represents the average value of MDA, $\eta(\mu_{MVM}, \mu_{MDA})$ represents the ratio between two values. δ_{FD} can be obtained from the original video while u_{MVM} and $\eta(\mu_{MVM}, \mu_{MDA})$ must be obtained after coding. Q-R-TQ algorithm estimates the model coefficients using fixed update duration [26]. In reality, video content may change over the time and is