J. Vis. Commun. Image R. 25 (2014) 1199–1208

Contents lists available at ScienceDirect

J. Vis. Commun. Image R.

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

Content-adaptive bitstream-layer model for coding distortion assessment of H.264/AVC networked video

Honglei Su *, Fuzheng Yang

State Key Laboratory of Integrated Service Networks, Xidian University, China

ARTICLE INFO

Article history:
Received 6 August 2013
Accepted 5 April 2014
Available online 18 April 2014

Keywords:
No reference
Video quality assessment
Bitstream-layer model
Coding distortion
H.264/AVC
Networked video
Spatial complexity
Temporal complexity

ABSTRACT

Bitstream-layer models are designed to use the information extracted from both packet headers and payload for real-time and non-intrusive quality monitoring of networked video. This paper proposes a content-adaptive bitstream-layer (CABL) model for coding distortion assessment of H.264/AVC networked video. Firstly, the fundamental relationship between perceived coding distortion and quantization parameter (QP) is established. Then, considering the fact that the perceived coding distortion of a networked video significantly relies on both the spatial and temporal characteristics of video content, spatial and temporal complexities are incorporated in the proposed model. Assuming that the residuals before Discrete Cosine Transform (DCT) keep to the Laplace distribution, the scale parameters of the Laplace distribution are estimated utilizing QP and quantized coefficients on the basis of the Parseval theorem firstly. Then the spatial complexity is evaluated using QP and the scale parameters. Meanwhile, the temporal complexity is obtained using the weighted motion vectors (MV) considering the variations in temporal masking extent for high motion regions and low motion regions, respectively. Both the two characteristics of video content are extracted from the compressed bitstream without resorting to a complete decoding. Using content related information, the proposed model is able to adapt to different video contents. Experimental results show that the overall performance of CABL model significantly outperforms that of the P.1202.1 model and other coding distortion assessment models in terms of widely used performance criteria, including the Pearson Correlation Coefficient (PCC), the Spearman Rank Order Correlation Coefficient (SROCC), the Root-Mean-Squared Error (RMSE) and the Outlier Ratio (OR).

1. Introduction

Recently, with the development of advantage multimedia and network technologies, networked video services such as mobile video conference, videophone, and Internet Protocol Television (IPTV) have gained significant popularity in our daily life. However, the quality of these applications cannot be guaranteed in IP network due to its best-effort delivery. It is therefore crucial to establish objective models for networked video quality assessment targeting system design, Quality of Service (QoS) planning and quality monitoring.

From the viewpoint of the input information, objective video quality methods can be classified into parametric models, packet-layer models, bitstream-layer models, media-layer models and hybrid models [1,2]. The parametric model employs parameters form the network or the application. Parameters from the network or application usually include the packet loss rate, the delay information, the coding bit rate, frame rate and so on. This model is initially designed for Quality of Experience (QoE) planners to formulate a crude estimation of the end-to-end service quality. The packet-layer model exploits the packet headers to obtain information about the service quality. Apart from the parameters employed in a parametric model, other information useful for estimating quality can be also employed for a packet-layer model. Such extra information includes the frame type, the bit-rate of each frame, the ratio of the bit-rate for I frames and P frames, the detailed positions of lost packets in a video and so on [3,4], which make it well suited for in-service non-intrusive monitoring. The bitstream-layer model predicts the networked video quality using media related payload information extracted from the encoded bitstream in addition to the packet layer information [5]. This model is dedicated to in-service non-intrusive monitoring, with more weight placed on the accuracy issue. The media-layer model, which has been studied in the past two decades [6–12], utilizes media signals to estimate video quality. This model requires the

* Corresponding author.
E-mail address: hlsu@mail.xidian.edu.cn (H. Su).

http://dx.doi.org/10.1016/j.jvcir.2014.04.004

reconstructed pixels of the video and therefore is also known as the pixel-based model. The hybrid model uses a combination of information from both bitstream and pixel domains, and therefore has an advanced performance due to combined features of the other models [13–15].

Generally speaking, extracting more information from the bitstream usually means an increase in computational complexity. Therefore, it becomes a critical problem to make a good trade-off between accuracy and computational complexity of quality assessment models. On one hand, due to their non-accessing to payload information, the parametric model and the packet-layer model cannot acquire a higher accuracy. For another, the media-layer model and the hybrid-layer model are time-consuming since they need to decode networked video completely to obtain pixel-based information. Moreover, pure pixel-based methods cannot employ any packet-related and coding-related parameters both of which are extremely helpful for networked video quality assessment. Consequently, the bitstream-layer model is an eclectic metric to get both a comparatively higher accuracy and a lower computational complexity, which is suitable for real-time quality assessment.

The definition and standardization of the bitstream-layer model of multimedia quality assessment has been determined by ITU-T as P.1202 [16]. The P.1202 has two modes: Mode 1 is a parsing and lower complexity mode. The model in this mode does not completely decode the payload. Without utilizing the pixel-domain information, any kind of analysis of the payload can be applied. While Mode 2 is the full decoding mode in which the model can partially or fully decode all of the video sequence, and the pixel information can be used for quality assessment. Perceptibly, when full decoding is applied, Mode 2 is similar to the hybrid model, which needs the input of the bitstream and the reconstructed video.

Considering the lossy nature of video coding and characteristics of practical networks, quality degradation or distortion in networked video consists of two main parts: coding distortion caused by quantization and distortion caused by packet loss. It is obvious that coding distortion assessment is fundamental in video quality assessment, because it can be used to assess the quality in absence of channel errors. When channel errors are present, coding distortion assessment still serves as a critical part for quality assessment of networked video. In this paper, therefore, bitstream-layer coding distortion assessment of H.264/AVC networked video is addressed.

Bitstream-layer models are usually utilized to estimate the peak signal-to-noise ratio (PSNR) of networked video [17–22], however, many studies have shown that PSNR is poorly correlated with perceived coding distortion since it does not take visual masking effects of Human Visual System (HVS) into consideration, which depend intensively on video content. Therefore, PSNR-based bitsteam-layer metrics considering the characteristic of HVS are studied in [23,24], whereas these PSNR-based metrics resort to full decoding to obtain the spatial characteristic of video content from pixel domain information. Moreover, the metrics in [25] use the pixel information to estimate spatial information (SI) and temporal information (TI) of a video sequence, described in P.910 [26], to make a more accuracy performance of video quality estimation. These bitstream-based models in [23–25], therefore, are categorized as bitstream-layer models with complete decoding which are similar to Mode 2 in P.1202. However, complete decoding usually means an increase in computational complexity.

A number of bitstream-layer models without resort to complete decoding have been studied in [5,17–22,27–31]. Metrics in [17–22] concentrate on the PSNR estimation of networked video and their disadvantages have been discussed in the above paragraph. The NR QANV-PA model is proposed in [5] where quantization parameter (QP) and bitrate are employed to estimate perceived coding distortion of video sequences. This approach provides a better performance and is suitable for real-time quality monitoring for MPEG4 coded networked videos. A bitstream-based approach considering temporal features is proposed in [27] where various motion-based factors as well as bitrate are utilized to estimate video quality, yet the spatial features have not been considered. In [28], the Mean Square Error (MSE) has been predicted using DCT coefficients and their corresponding quantizing steps, and then the spatial activity, estimated by computing the standard deviation of the values of the DCT coefficients contained in I frames, is combined using a logistic function to evaluate the coding distortion of video sequences. Nevertheless, the proposed method does not take the temporal activity into consideration. Models considering both spatial and temporal features have been proposed in [29,30]. However, the spatial feature in [29] which is the index of the spatial complexity are calculated utilizing the residual errors in P or B frames which also reflect the temporal feature of videos in a way [5], so the spatial feature estimation in this way may loss efficacy to some extent. The formula in [30] for deriving the final quality show a linear relationship between the average QP value and the objective score, while the linear function may not work well for H.264/AVC networked video [31]. In [31], the relationship between QP and the values of Mean Opinion Score (MOS) for different video sequences with various contents is shown as an inverse sigmoid curve. This model is aiming at the quality estimation of H.264/AVC coded videos. The content features of video sequences, however, are not taken into consideration in the proposed metric, therefore it cannot give a higher accuracy for individual networked videos.

In this paper, a content-adaptive bitstream-layer (CABL) model is proposed for assessing coding distortion of H.264/AVC networked video, and a diagram of the CABL framework is shown in Fig. 1. Without resorting to inverse quantization and its following operations in decoding, the proposed model employs frame type information, quantization parameters, motion vectors (MV) and quantized coefficients to estimate the coding distortion of networked video. Since quantization is the central reason for coding distortion, the fundamental relationship between QP and coding distortion is determined. Then, the quantized coefficients and their corresponding QPs extracted from intra-coded frames are used to calculate the spatial complexity, and MVs extracted from inter-coded frames are used to calculate the temporal complexity. Finally, the spatial and temporal complexities of video content are incorporated together with QP to obtain the perceived video quality, which make the proposed model content-adaptive. The current work as reported in this contribution targets lower resolution applications, which are widely used for real-time wireless video services due to the bandwidth limitations, such as mobile TV [16].

The remainder of this paper is organized as follows. Section 2.1 describes the relationship between perceived coding distortion and QP for H.264/AVC networked video. How to obtain the content-related characteristics, the spatial complexity and the temporal complexity, are described in Sections 2.2 and 2.3, respectively. Content-adatptive bitstream-layer model for quality assessment is proposed in Section 2.4. Section 3 provides the experimental results and discussions, and conclusions are drawn in Section 4.

2. Content-adaptive bitstream-layer assessment for coding distortion of H.264/AVC networked video

2.1. Perceived coding distortion and quantization parameter

Blocking artifacts are the main coding distortion in reconstructed video frames due to the block-based DCT compression methods. It is defined as the boundaries discontinuity of coding blocks, which is