EXTENDING G.1070 FOR VIDEO QUALITY MONITORING

Niranjan D. Narvekar, Tao Liu, Dekun Zou, and Jeffrey A. Bloom
Dialogic Media Labs, 12 Christopher Way, Suite 104, Eatontown, NJ 07724
{first.last@dialogic.com}

ABSTRACT

In 2007, the International Telecommunication Union (ITU) standardized a multimedia quality assessment model as ITU-T Recommendation G.1070. The video quality estimation model proposed in this document uses the encoded bit rate and frame rate of the compressed video, along with the expected packet loss rate of the channel, to predict the subjective video quality. The model was designed as a video quality planning tool and requires prior knowledge of or assumptions about the video and channel parameters. This paper considers the use of the G.1070 video quality estimation model for monitoring applications, where the input parameters themselves must be estimated from the observed bitstream. We address the issues that arise when estimating these video and channel parameters from a real-time video stream.

Index Terms— ITU-T Recommendation G.1070, quality monitoring, bitrate, frame rate, and packet loss rate.

1. INTRODUCTION

As video is becoming increasingly popular for entertainment and day-to-day communications, there is an ever increasing need to assess the Quality of Experience (QoE) of the end users. Numerous models have been proposed to predict the subjective quality of video in various scenarios and applications [1]. One popular work is the multimedia quality model standardized by ITU in its ITU-T Recommendation G.1070 [2].

In this ITU recommendation, a framework for assessing the multimedia quality is proposed. It consists of three models: a video quality estimation model, a speech quality estimation model, and a multimedia quality integration model. The video quality estimation model (which we will loosely refer to as the G.1070 model in this paper) uses the bit rate and frame rate of the compressed video, along with the expected packet loss rate of the channel, to predict the subjective quality of the video subject to compression artifacts and transmission error artifacts. Details of the G.1070 models, including equations, can be found in [2].

Since its standardization, the G.1070 model has been widely used, studied, extended, and enhanced. Yamagishi and Hayashi proposed to use G.1070 in the context of IPTV quality assessment [3]. Since the G.1070 model is codec-dependent, Belmudez and Mo□ller extended the model, originally trained for H.264 and MPEG4 video, to MPEG-2 content [4]. Joskowicz and Ardao enhanced the G.1070 with both resolution- and content-adaptive parameters [5].

However, the use of G.1070 is still limited to planning. The G.1070 model cannot be readily used within a network or at a video player for the purpose of real-time video quality monitoring. This is because the three inputs to the G.1070 model; bitrate, frame rate, and packet loss rate of the encoded video bitstream; are not immediately available. To overcome this limitation, these quality parameters can be estimated from the bitstream. As we will discuss in this paper, this is sometimes easier said than done. In this paper, we propose estimation methods that allow G.1070 to be extended from a planning tool to a real-time video quality monitoring tool. Specifically, we describe methods for real-time estimation of these three quality parameters in a typical video streaming environment. Some of the practical issues are discussed. Based on simulation results, we analyze the performance of the proposed system in estimating these parameters.

2. ESTIMATING THE G.1070 PARAMETERS

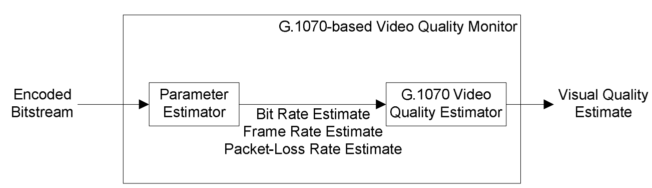

Figure 1. A Video Quality Monitor using the estimated quality parameters.

Figure 1 shows a block diagram of a proposed video quality monitor. It consists of a parameter estimator followed by a G.1070 video quality estimator. The input to the Video Quality Monitor is an encoded bitstream, packetized using any of the standard packetization formats such as RTP, MPEG2-TS etc.

The function of the parameter estimator is to estimate bit rate, frame rate, and packet loss rate from the received

978-1-61284-350-6/11/$26.00 ©2011 IEEE

bitstream. These parameters are then used by a G.1070 Video Quality Estimator, which calculates the video quality based on the video quality estimation function [2].

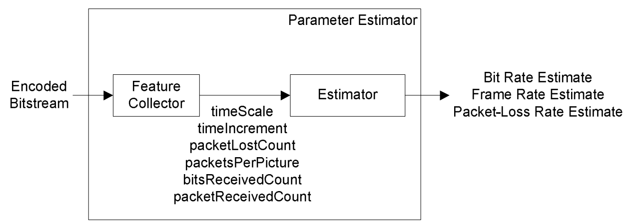

Figure 2. The Parameter Estimator is composed of a Feature Collector and an Estimator.

Figure 2 shows the sub-components of the Parameter Estimator: a Feature Collector and an Estimator. The Feature Collector extracts the necessary information from the bitstream. The Estimator uses this information to estimate the G.1070 input parameters.

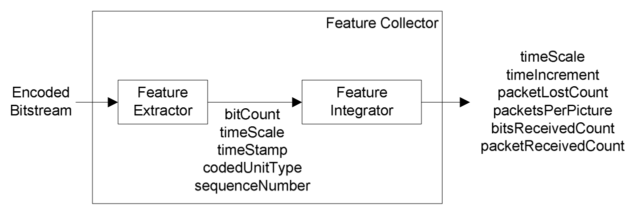

Figure 3. The Feature Collector is composed of a Feature Extractor and a Feature Integrator.

The details of the Feature Collector are shown in Figure 3. It consists of a Feature Extractor and a Feature Integrator. The Feature Extractor collects the following information from each packet in the bitstream:

- timeScale – the reference clock frequency of the transport format. For example, let us consider the transport of video over RTP, the standard clock frequency in this case is 90 KHz.
- timeStamp – the display time of the frame to which the packet belongs.
- bitCount – the number of bits in the packet.
- codedUnitType – the type of data in the packet. For example, in H.264 the coded unit type corresponds to the NAL-unit type.
- sequenceNumber – the sequence number of the input packet.

In order to estimate the bit rate, frame rate, and packet loss rate, statistics are collected over a number of frames. An N-frame sliding window is used. This sliding window mechanism is implemented by the Feature Integrator. The Feature Integrator takes the information collected by the Feature Extractor and generates the following estimates over a window of N frames:

- timeScale – same as described earlier.
- timeIncrement – the time interval between two adjacent video frames in display order.
- bitsReceivedCount – the number of video coding layer bits received over the N-frame window. The determination of whether the bits belong to the video coding layer is based on the input codedUnitType. For example, in H.264 the SPS and PPS NAL-units do not belong to video coding layer and hence are not included in the calculation.
- packetReceivedCount – the number of packets received over the N-frame window.
- packetLostCount – the number of packets lost over the N-frame window. This can be determined by counting the discontinuities in the sequence number information.
- packetsPerPicture – the number of video coding layer packets per picture.

The estimates of timeIncrement, bitsReceivedCount, and packetsPerPicture are prone to be affected by packet losses. Hence, care must be taken to either carefully calculate these estimates or compensate for errors. The calculation of bit rate is based on the bitsReceivedCount for which packet loss errors can be compensated. This is explained later. However, the estimation of timeIncrement and packetsPerPicture can be performed such that they are not affected by packet losses as explained below.

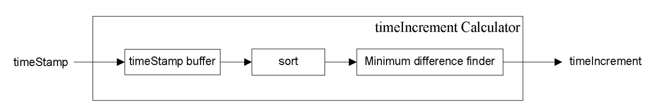

Figure 4. timeIncrement Estimator

Figure 4 shows the timeIncrement Calculator. The estimation of the timeIncrement between the frames in display order is complicated by the fact that almost all state-of-the-art encoding standards use a highly predictive structure. Because of this, the coding order is not the same as display order and hence the received timestamps are not monotonically increasing. Also packet losses can lead to frame losses which can cause missing timestamps. In order to overcome these issues, the timeIncrement estimator buffers the timestamps over N frames, sorts them in ascending order, and then calculates the minimum timestamp difference which is taken as the timeIncrement. The sorting makes sure that the timestamps are monotonically increasing and calculating the minimum timestamp difference helps in discarding the negative impact of any frame losses for the timeIncrement calculation.

The packetsPerPicture estimate is calculated by taking into consideration the fact that packet losses can interfere with an accurate estimation. A packetsPerPicture estimate is calculated for each picture. For those frames that are affected by packet loss, the corresponding packetsPerPicture estimates are discarded.

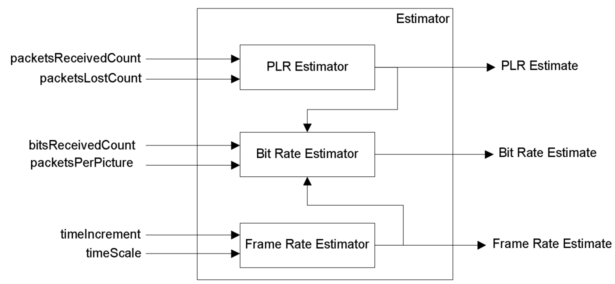

Figure 5. The Estimator sub-component of the Parameter Estimator.

At this point the Feature Collector component has collected all the necessary information to calculate the input parameters of the G.1070 Video Quality Estimation model. The actual calculation of the input parameters is performed in the Estimator sub-component of the Parameter Estimator as shown in Figure 5.

The Packet Loss Rate Estimator takes the packetReceivedCount and the packetLossCount as input and calculates the packet loss rate (PLR) as follows:

$$PLR = \frac{packetsLostCount}{(packetsLostCount+packetsReceivedCount)} \quad (1)$$

The Frame Rate Estimator takes the timeIncrement and timescale as input and calculates the frame rate (FR) as follows:

$$FR = timescale \,/\, timeIncrement \quad (2)$$

The bit rate (BR) is estimated from the bitsReceivedCount, the packetsPerPicture, the estimated packet loss rate, and the estimated frame rate. In order to make the calculation of BR robust to packet loss, this calculation varies based on the estimated number of packets per picture. When each frame is transmitted in a single packet, i.e. packetsPerPicture = 1, no correction factor is needed and the BR is calculated as follows:

$$BR = FR \times bitsReceivedCount \,/\, N \quad (3)$$

However, if a frame is broken into multiple packets, i.e. packetsPerPicture > 1, the bitrate estimate is normalized by the percentage of packets received.

$$BR = FR \times \frac{bitsReceivedCount}{N \times (1\text{-}PLR)} \quad (4)$$

Finally, the estimates of bit rate, frame rate, and packet loss rate are provided to a standard G.1070 Video Quality Estimator which calculates the corresponding video quality. Note that the parameters are estimated over a window of N frames. This means that the quality estimate at a frame is obtained from the statistics of the N preceding frames. The proposed system generates a video quality estimate for each frame, except during the initial buffering of N frames. No quality measurement is generated for lost frames.

3. PERFORMANCE RESULTS

The parameter estimates are evaluated using the EPFL-PoliMI video quality assessment database [6]. The bitstreams available in [6] are RTP packetized H.264 bitstreams that were encoded using the JM encoder [7]. The encoded bitstreams do not have the correct timestamp information; the time increments are a constant value of 1000. JM inserts these constant timestamps regardless of the video frame rate. The bitstreams used in our evaluation were preprocessed to correct the timestamp information to comply with the RTP standard clock frequency of 90 KHz.

A JM decoder is instrumented to collect the bitstream information necessary for estimating these parameters.

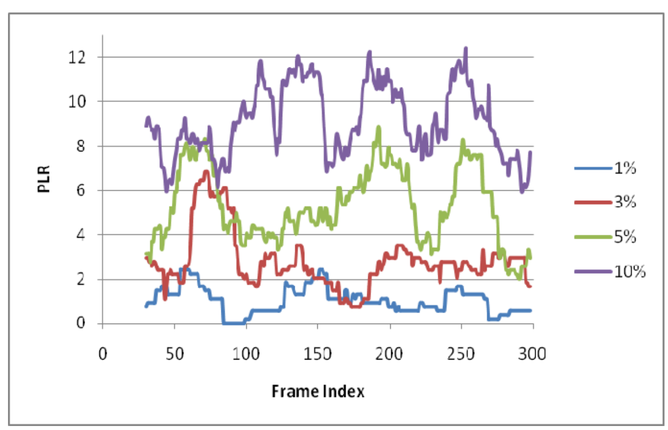

Figure 6. Estimated Packet Loss Rate vs. frames for various average packet loss rates.

Figure 6 shows a plot of the estimated packet loss rate vs. frames for the 30 fps foreman sequence of CIF resolution having average packet loss rates of 1%, 3%, 5%, and 10%. Note that the instantaneous packet loss rate estimates fluctuate around the average packet loss rate value. This is because the proposed system implements a sliding window mechanism and the estimated packet loss rate is over that window. In this case a window size of N = 30 was used. The average estimated packet loss rates are 1.03%, 2.81%, 5.29%, and 9.09% for average packet loss rates of 1%, 3%, 5%, and 10% respectively.

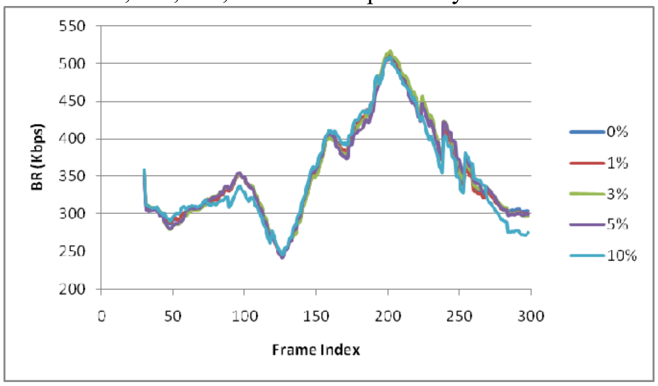

Figure 7. Estimated Bit Rate vs. Frame Index for various average packet loss rates.

Figure 7 shows a plot of the estimated bit rate vs. frames for 30 fps foreman sequence of CIF resolution having average packet loss rates of 0%, 1%, 3%, 5%, and 10%. Note that this sequence was encoded at a bit rate of 353 kbps. From the figure it can be noticed that the estimated bit rates for various packet loss rates are very close to the 0% loss case implying that the bit rate estimate is quite robust to packet losses. The 10% packet loss rate case seems to be slightly off. This is not unexpected, as the accuracy of the bit rate estimate is expected to decrease at high packet loss rates. The average estimated bit rates are 355 kbps, 354.8 kbps, 356 kbps, 354.2 kbps, and 351.8 kbps for average packet loss rates of 0%, 1%, 3%, 5%, and 10% respectively. Note that the average estimated bit rates are very close to the encoded bit rate of 353 kbps. The variation in the instantaneous bit rates is because of the sliding windowed implementation of the proposed system and is also dependent on video content.

The plot of estimated frame rate vs. frames is omitted here because the system accurately estimates it at exactly 30 fps for all the considered packet loss rates of 1%, 3%, 5%, and 10%.

4. CONCLUSION

The G.1070 video quality model is intended for planning applications. In order to extend the model for a monitoring application, the bit rate, frame rate, and packet loss rate must be accurately estimated from an encoded bitstream. In this paper, we presented methods for estimating of the three G.1070 input parameters. Preliminary evaluation suggests that the methods are reasonably accurate and reasonably robust to packet loss.

One of the next steps in this work is a larger scale study of the accuracy and robustness of the three estimation methods. The methods should be evaluated on a larger data set covering a variety of content, bit rates, and frame rates. In addition to obtaining more conclusive statistics, it is important to understand any dependencies on the video content, the actual frame rate and bit rate, the codec used, the resolution of the video, etc.

Another next step is to understand how sensitive the G.1070 algorithm is to errors in the bit rate, frame rate, and packet loss rate estimates. This can only be done through correlations with subjective MOS scores as there is no G.1070 ground truth with which to compare. For example, a bit rate of 350 kbps might be the target bit rate at which a video is encoded and this value of 350 kbps would be the value used for planning purposes to obtain a G.1070 estimate of the resulting video quality. However, even using a constant bit rate encoding, the instantaneous bit rate will likely deviate from 350 kbps and the resulting video quality will likely vary over time. For constant QP encoding, the deviation is more significant. This is why the G.1070 value obtained as a planning estimate should not be used as ground truth to assess a monitoring quality estimate. Subjective MOS scores should be used.

5. REFERENCES

[1] S. Winkler, Digital Video Quality: Vision Models and Metrics, John Wiley & Sons, 2005

[2] Recommendation ITU-T G.1070, “Opinion Model for Video-telephony Applications”, April 2007

[3] K. Yamagishi and T. Hayashi, “Parametric Packet-Layer Model for Monitoring Video Quality of IPTV Services”, IEEE International Conference on Communications, May 2008

[4] B. Belmudez, and S. Mo□ller, “Extension of the G.1070 video quality function for the MPEG2 video codec”, QoMEX, 2010

[5] J. Joskowicz, and J. Ardao, “Enhancements to the opinion model for video-telephony applications”, Proceedings of the 5th International Latin American Networking Conference, 2009

[6] F. Simone, M. Naccari, M. Tagliasacchi, F. Dufaux, S. Tubaro, and T. Ebrahimi, "Subjective assessment of H.264/AVC video sequences transmitted over a noisy channel", QoMEX 2009

[7] Joint Video Team (JVT), “H.264/AVC reference software version JM16.2,” downloadable at http://iphome.hhi.de/suehring/tml/.