

A Novel Scheme of Video Quality Estimation for the ROI Transcoding Video

Ting Yao¹.

Abstract. Because of the limitation of bandwidth and display screen size on the mobile devices, the experience of watching videos on the mobile device is extremely limited, especially in real-time video conferencing. This issue can be solved by the ROI (Region of Interest)-based transcoding. To verify the result of ROI-based transcoding scheme, we need an effective method to assess the ROI-based transcoding video quality. In our system, we used the ITU-T Recommendation G.1070 model to assess the video quality. To use the model, we had to estimate twelve parameters in G.1070 firstly because these parameters used in our system are not in the Annex A of the G.1070. By analyzing the result of the subjective video quality assessment, we estimated the twelve parameters. Based on the twelve parameters, we estimated two weighting factor to assess the ROI-based transcoding video quality.

Keywords: H.264; mobile device; ROI; G.1070; real-time videoconferencing.

1 Introduction

According to the Google I/O 2013 conference, the activated Android devices reached 900 million [9]. More and more people tend to use the smart phone to deal with the daily life and their work. Real-time interactive multimedia applications have become common programs on smart phones. However, these high quality real-time video applications bring up challenges to mobile devices due to the limitation of the bandwidth and the small display size. Then we should find the trade off of between the video quality and these limitations. We find a novel scheme based on ROI transcoding to solve these challenges [8]. In our system, we need to estimate the video quality in real-time to adapt our transcoding scheme.

To video quality assessment, there are subjective video quality and Objective video quality assessment. Subjective video quality assessment methods are based

¹ T. Yao (✉)

School of Computer Science and Engineering, Beihang University, Beijing, China
e-mail: yaoting@nlsde.buaa.edu.cn

on groups of trained/untrained users viewing the video content, and then providing ratings for quality [4]. But it is obviously that the subjective video quality assessment could not provide instantaneous measurement of video quality not even real-time video and they are time consuming, laborious and expensive. The Objective video quality assessment methods can overcome these shortages. The multimedia quality model that is standardized by ITU-T in its Recommendation G.1070 in 2007[5] is a widely used Objective video quality assessment model.

There are twelve parameters in the Recommendation G.1070. The parameters are for the MPEG-2 or MPEG-4 in the Annex A of the Recommendation G.1070. The new parameters and relation proposed are calculated for H.264/AVC [2] in bit ranges from 100kb/s to 2Mb/s in this paper. We extend the ITU-T Recommendation G.1070 so that the model can take into account the video content and measure the video quality more accurately in real-time.

This paper is organized as follows: We elaborate our experiment in section 2 and analyze the results in section 3. Finally, we conclude this paper and point the further work in section 4.

2 Video Quality Assessment framework for ROI-based Mobile Video Adaptation

2.1 Video quality function in the ITU-T Recommendation G.1070 model

The ITU-T Recommendation G.1070 is an objective quality assessment model for video telephony or video conferencing. Its framework consists of three functions: video quality estimation, speech quality estimation [3], and multimedia quality integration [7]. Video quality estimation is V_q calculated as shown in the following equations:

$$V_q = 1 + I_{coding} e^{\frac{P_{ph}}{D_{ph}}} \quad (1)$$

The V_q represents the video quality, in the range from 1 to 5. Different video quality scores (VQS) represent the video quality level showing in the table 1.

Table.1. VQS to video perceived quality relation

| Quality | Bad | Poor | Fair | Good | Excellent |
|---------|-----|------|------|------|-----------|
| VQS | 1 | 2 | 3 | 4 | 5 |

$$I_{coding} = I_{Ofp} e^{\frac{(\ln(Fr_V) - \ln(O_{fr}))^2}{2D_{Fr}^2}} \quad (2)$$

I_{coding} represents the basic video quality, which affected by the coding distortion under a combination of video bit rate B_{rv} , and video frame rate F_{rv} .

$$O_{fr} = v_1 + v_2 B_{rv}, \quad 1 \leq O_{fr} \leq 30 \quad (3)$$

$$I_{Ofr} = v_3 - \frac{v_3}{1 + \left(\frac{B_{rv}}{v_4} \right)^{v_5}}, 0 \leq I_{Ofr} \leq 4 \quad (4)$$

I_{Ofr} is the optimal quality with a given bit rate and the frame rate to optimal quality is denoted as O_{fr} .

$$D_{FrV} = v_6 + v_7 B_{rv}, \quad 0 \leq D_{FrV} \quad (5)$$

The degree of video quality robustness due to frame rate (Frv) is denoted as D_{frV} .

$$D_{Pplv} = v_{10} + v_{11} e^{\frac{FrV}{v_8}} + v_{12} e^{\frac{BrV}{v_9}}, 0 < D_{Pplv} \quad (6)$$

D_{pplv} represents the packet loss robustness caused by the packet loss where P_{plv} means packet loss Rate (PLR).

v_1, v_2, \dots, v_{12} are the twelve constants to be determined. The video quality V_q without packet loss is expressed as follow:

$$V_q = 1 + I_{coding} \quad (7)$$

From equation (7), the V_q is related to v_1, v_2, \dots, v_7 coefficients. According to the ITU-T Recommendation G.1070, coefficients v_1, v_2, \dots, v_7 are dependent on code type, video format, key frame interval, and video display size, and must be calculated with subjective video quality tests. The Appendix I of the Recommendation G.1070 [5] only for MPEG-4 in QVGA (320×240 pixels) and QQVGA (160×120 pixels) video formats. Belmudez et al [1] proposed a new set of parameters for the MPEG-2 and Yamagishi et al [6] proposed another set of parameters for the H.264 codec. However, in our system, the value of these parameters for H.264 codec and other display size need to be determined. As the ITU-T Recommendation G.1070, the model coefficients are not dependent on the video content and there is only one set of coefficient for all video contents.

2.2 ROI based video quality assessment method

Fig. 1 shows our ROI (Region of Interest) -enabled transcoding scheme. Firstly, the frame that captured by the HD camera is larger than the resolution of the mobile client. Therefore, the mobile client only displays part of the whole frame. The mobile client displays part of the HD frame and if the server launches the transcoding scheme, the frame displayed on the smart phone will include the ROI region. Under the low bandwidth condition, the transcoding server transcodes the frame by coding the non-ROI with higher QP (quantization parameter) value and ROI with lower QP value. As we know, the QP value is lower, the video quality is higher. Therefore, the video perceived quality is relatively good but bandwidth is not so high.

To estimate the video quality in our system, we redefine the method for calculating the overall video quality. The overall video quality estimation V_q is calculated as shown in Equation (8).

$$V_q = w_1 V_{qBase} + w_2 V_{qROI}, w_1 + w_2 = 1 \quad (8)$$

V_{qBase} represents the video quality estimation of the frame without ROI and V_{qROI} is the video quality estimation of the ROI region. The coefficients w_1 and w_2 are the weighting factors.

3 Video Quality Experiments

We conduct two viewing-only subject quality experiments. One is for the precision value of the coefficients in Recommendation G.1070, and the other is for estimating the weighting factors w_1 and w_2 in Equation (8).

In our experiment, the video formats are VGA and CIF. One of our video content displays the head and shoulder, and we can see in Fig. 2. Each test condition is determined by a combination of bit rate, frame rate and packet loss rate. Thirty conditions for a null packet loss rate are tested in our experiment with six different frame rates and five bit rates. To the subjective quality assessment, we reference to the ITU-T Recommendation P.910 [3]. We invite twenty test participants age between 20-30 balances in gender in the two experiments. They are not experts in multimedia and are not concerned with the multimedia quality as part of their work. And before they beginning the experiment, the test participants follow a training phase. Table 2 shows the experiment conditional parameters: bit rates, frame rates, codec type, display formats and so on.

The second experiment is about viewing-only subjective experiment. We invite the same twenty users to score these video sequences. Then we collect the MOS of two kinds of video sequences. The two kinds of video sequences are base frame video (frame without ROI transcoding), ROI video with higher quality, and the ROI based transcoding video. The ROI based transcoding video and base frame video are under the same bit rate. Once we collect the MOS of the two kinds of video sequences, we can calculate the two parameters in equation (8).

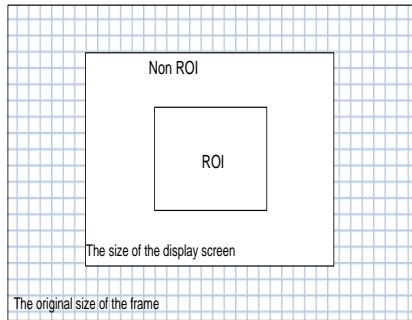


Fig. 1. The ROI-enabled transcoding scheme.



Fig.2. One of the test scenes.

Table.2. Parameters of experiment

| | |
|--------------------------|--------------------------|
| Video Codec | H.264 |
| Display format | VGA, CIF |
| Key frame interval (sec) | 1 |
| GOP Pattern | IPBPBP |
| Video Bit Rate (kbs) | 128, 256, 512, 768, 1024 |
| Frame rate (fps) | 5, 10, 15, 20, 25, 30 |
| Packet loss rate (%) | 0, 1, 2, 5, 10 |
| Packet loss pattern | Poisson |

4 Experiment Result Analysis

4.1 Parameters estimating

According to the Annex A of G.1070 [5], we model the subjective data and extract the optimized parameters (v_1 to v_7) which are described in equations (1) to (5). We first study the video quality affected by the codec with a null packet loss rate ($P_{plv}=0$). It is elaborated in equation (7). In Fig. 3, we can see that the data set fits quite well to the proposed model. We can notice that the MOS reaches to a maximum which corresponds to the optimal frame rate for a given bit rate, and then decreases to higher frame rates. To the bit rate 128kbps and 256 kbps, the optimal frame rate is about 6 - 7fps, but to the bit rate 768kbps, 1024kbps and 2048kbps, the optimal frame rate is about 15fps or higher. It can be explained by the tradeoff between bit rate and frame rate. If the frame rate is too low or too high, the perceived video quality may not be the best.

4.2 Video quality influenced by codec

From the video quality fitting, we get a relationship among B_r , F_r , V_q . Therefore, we get the corresponding values of O_{fr} , I_{Ofr} and D_{frv} for each bit rate. Parameters v_1 to v_7 can be extracted by the equations (1) to (5).

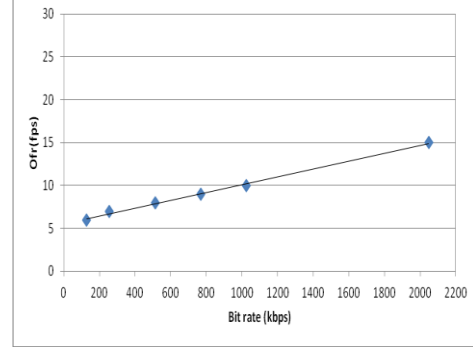
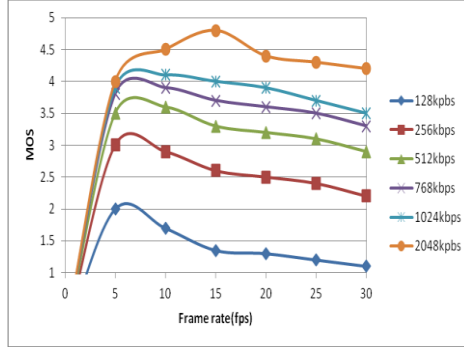


Fig. 3 Perceptual video quality for different bit rates **Fig. 4** Optimal frame rate (O_{fr}) affected by the bit

In Fig. 4, the O_{fr} (optimal frame rate that maximizes the video quality at each video bit rate) increases in a linear way. And in Fig. 5, the value of I_{Ofr} linear increases starts at about 400kbps, so the lowest bit rate is excluded from further fitting. The fitting of I_{Ofr} (the maximum video quality at each bit rate) uses equation (4). And then, we can see the relationship between D_{frv} in equation (5) and the bit rate in Fig.6. In Fig.6, the line approximately increases except the lowest bit rate 128kbps. Therefore, we exclude it again for fitting the proposed model.

According to the Annex A of ITU-T G.1070 model, we use a linear regression algorithm based on the least square approximation (LSA) to perform the data fitting the formulas (2), (3), (5) and a non-linear regression algorithm based on the LSA to perform the data fitting the equation (4). Therefore, we derive coefficients v_1 to v_7 . The values of the seven coefficients are in table 4.

4.3 Video quality influenced by packet loss

To control the overall amount of test condition for the subjective experiment, we choose four different packet loss rates for different combinations of bit rates and frame rates. The combinations are shown in Table 3.

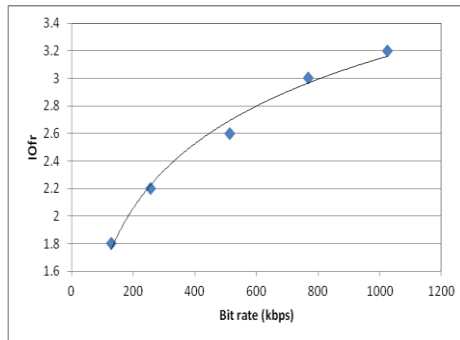


Fig. 5. I_{Ofr} affected by the bit rate

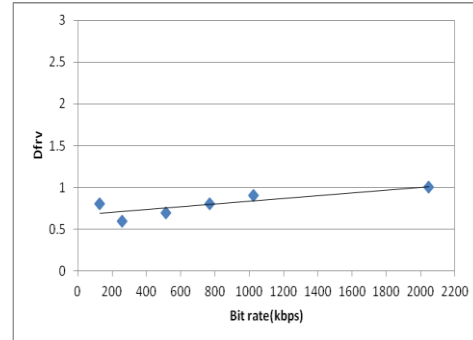


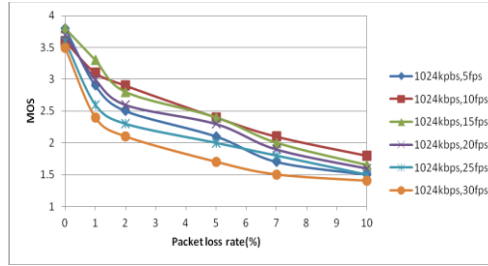
Fig. 6. D_{frv} affected by the bit rate

Table.3. Combinations among bit rates, frame rates, and PLR

| | |
|----------------------|-----------------------|
| Bit Rate (kbps) | 2048, 1024, 512 |
| Frame Rate (fps) | 5, 10, 15, 20, 25, 30 |
| Packet Loss Rate (%) | 1, 2, 5, 10 |

The Fig. 7 represents the video quality influenced by packet loss, bit rate and the frame rate. In the figure, we can observe that the video quality decreases very fast at first. And then with increasing the packet loss rate, the video quality converges slowly around 1.6.

We use a non-linear regression algorithm based on the LSA to perform the data fitting the formulas (1) and (6). In Fig. 7, we can figure out that the data fitting using an exponential function is not optimal.

**Fig. 7.** Perceptual video quality affected by packet loss

Finally, we discuss the accuracy of these coefficients in the ITU-T Recommendation G.1070 model. We use the subjective quality test cases to verify the validity of the formulas (1) to (7). The optimized coefficients v_1, v_2, \dots, v_{12} for the video quality estimation are list in Table 4. The Pearson correlation computed between the estimated and observed MOS value is about 0.978 on average.

4.4 Weighting parameters estimating for ROI based transcoding

In this part, we conduct a real-time video conferencing to extract the two weighting parameters w_1 and w_2 . Under a given bandwidth, if the quality is not enough high, our system will start the transcoding scheme. The transcoding server transcodes one frame with different quantization parameters (QP) and then combines the frame with different QP. The perceived quality of ROI region is higher than the non-ROI region in the frame. Therefore, we can obtain lower bit rate but higher perceived video quality.

Table.4. The optimized parameters

| Coefficients | H264/VGA | H264/CIF |
|--------------|----------|----------|
| v_1 | 8.061 | 3.988 |
| v_2 | 0.007 | 0.013 |
| v_3 | 3.083 | 3.625 |
| v_4 | 80.74 | 89.25 |
| v_5 | 1.14 | 1.125 |
| v_6 | 1.043 | 0.713 |
| v_7 | 0.002 | 0 |
| v_8 | 2.116 | 1.542 |
| v_9 | 647.4 | 245.5 |
| v_{10} | 2.436 | 3.011 |
| v_{11} | 15.28 | 39.31 |
| v_{12} | 10.27 | 16.67 |

In Fig.8, we can observe that, the overall perceived video quality of the frame with ROI (in red rectangle) after transcoding in (a) is higher than that of the frame without ROI in (b). In the figure, the ROI is marked in red rectangle. The resolution of ROI is 352×288 (CIF).

Finally, we conduct another viewing-only subjective quality experiment. One video sequence is about the video with ROI transcoding and the other one is with-



(a) One frame with ROI after transcoding

(b) One frame without ROI before transcoding

Fig.8. Comparison between the original and the transcoded frame

out ROI region under the same bit rate 768kbps. Therefore, we have two set of data which are about the MOS value of the frame with ROI and that of frame without non-ROI, respectively. Associating with the formula (8), we can compute the weighting parameters. The weighting parameters w_1 and w_2 are 0.44 and 0.56 respectively.

5 Conclusion

This paper proposes a model to estimate the objective video quality in real-time. In this paper, we also extract a new set of optimized parameters so that we can use the ITU-T Recommendation G.1070 in our system to estimate the video quality. The experiment result proves good estimation accuracy by using these coefficients. When the bit rate is higher than 256kbps, our data can fit well. Finally, we also extract the weighting parameters in our proposed algorithm.

References

1. B Belmudez, S Moller (2010), Extension of the G.1070 video quality function for the MPEG2 video codec, In: International Workshop on Quality of Multimedia Experience (QoMEX) IEEE, 2010, 7-10. DOI=<http://dx.doi.org/10.1109/QOMEX.2010.5518225>.
2. ITU-T Rec. H.264 & ISO/IEC 14496-10 (2008), Advanced Video Coding for Generic Audio-visual Services, 2008.
3. ITU-T Rec. P.910-200804-I (1998), Subjective video quality assessment methods for multimedia applications. International Telecommunication Union, Geneva, Sep. 1998.
4. ITU-R Recommendation BT.500-11(2006), Methodology for the subjective assessment of the quality of television pictures, 2006.
5. ITU-T Rec. G1070 (2007), Opinion model for video-telephony applications, 2007.
6. K Yamagishi, T Hayashi(2006). Opinion model for estimating video quality of videophone services. In: Global Telecommunications Conference (San Francisco, CA, Nov. 27-Dec.1, 2006). GLOBECOM'06.IEEE,1-5.
7. J Joskowicz, J Ardao (2009). Enhancements to the opinion model for video-telephony applications. In: LANC'09 Proceedings of the 5th International Latin American Networking Conference, LANC'09,ACM, New York, NY, 87-94. DOI= <http://dx.doi.org/10.1145/1636682>.
8. Y Lou, T Yao, Y Chen and W. Wu (2013). A novel scheme of ROI detection and transcoding for mobile devices in high-definition video conferencing. In: MoVid '13 Proceedings of the 5th Workshop on Mobile Video. MoVid'13, ACM, New York, NY,31-36.DOI=<http://dx.doi.org/10.1145/2457413.2457421>.
9. <https://developers.google.com/events/io/sessions/324958375>.