No-reference Perceptual Quality Assessment for Streaming Video Based on Simple End-to-end Network Measures

Shengke Qiu[1,2] Huaxia Rui[2,3] Le Zhang[2,3]

[1]*Research Institute of Information Technology, Tsinghua University, Beijing, China*
[2]*Department of Automation, Tsinghua University, Beijing, China*
[3]*Network Research Center, Tsinghua University, Beijing, China*
qsk03@mails.tsinghua.edu.cn

Abstract

Video compression and transmission are essential to IPTV service providers. Errors in transmission may introduce degradation on the quality of the video sequences. In this paper we focus on the transmission process and propose an analytical framework based on no-reference perceptual quality assessment to evaluate the impact of packet loss on the images of video sequences. Two network parameters are employed to characterize the packet loss pattern of an IP network. The regression model established in this paper demonstrated a strong correlation between the network conditions and the quality of video sequences. The method developed in this paper enables the choice of network parameters to estimate users' QoE, which is very useful in IPTV network evaluation.

1. Introduction

With the increasing popularity of video communication over best-effort IP networks, it is crucial to detect and monitor the perceptual quality degradation, which is introduced directly or indirectly by the distortions due to errors in effects of compression and transmission-related distortions.

Users want to maximize their Quality of Experience (QoE), a measure of the actual quality perceived by an end-user. In today's emerging IPTV industry, QoE is the customer perception of how good of a job the service provider is doing in delivering the service, and it can be defined as a combination of the quality of the encoding components, the networking components, the decoding components, and the human factors per unit time [12], which is related to the *Quality of Service* (QoS) of the overall video communication system.

The perceptual quality degradation can be classified into two categories. The first is caused by data error introduced by the video compression process, such as blocking edge artifacts, mosaic pattern effect, and blurring. Figure 1 shows typical blocking edge artifacts. The second is caused by packet loss during the video transmission process of compressed video. Figure 2 gives an example of artifacts caused by packet loss.

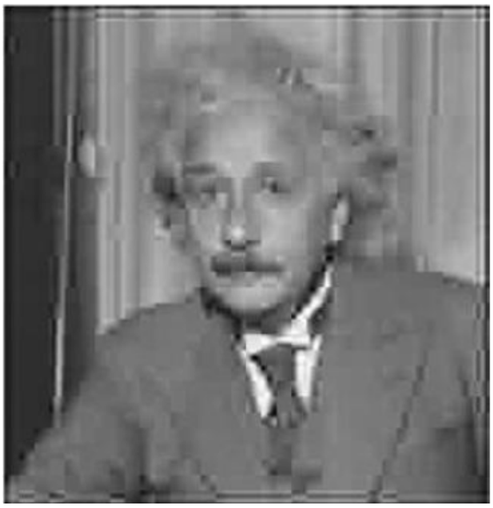

Figure 1. Blocking artifacts caused by compression

For the first category of perceptual quality degradation, a lot of models and evaluating works have been going on for quite a long time. M. Yuen and H.R. Wu reviewed and classified different coding distortions and various evaluation metrics [2].

Figure 2. Impairment caused by packet loss

Reported works on perceptual quality degradation in the second category are much less than the first one in the literature, whereas accurate evaluation of the packet loss impairment on the perceptual quality of compressed video is extremely valuable because it contributes significantly to various video applications.

X. Lu, R.O.Morando, and M.E. Zarki developed an adaptive perceptual video quality control mechanism based on an application-level perceptual video quality scheme [3]. At the same time, N. Feamster and H. Balakrishnan quantified the effects of packet loss on the quality of MPEG-4 video, and proved the effectiveness of their proposed adaptive system [4]. S. Tao and R. Guerin later studied on the dynamical path selection problem of streaming video from the perspective of video quality [5]. Several other researchers used quality degradation metrics to compare error concealment strategy and system performance in the presence of errors [6, 7, 8].

To video service providers, the online real-time QoE measures can be used as the basis to adjust video services. The key point is to easily draw out useful real-time information from a series of video sequences. Under end-to-end transmission environments, if video providers want to know QoE of end-users, they currently have to fetch the video sequences from end-users side and compare the transmitted and received video sequences, which costs a great deal of resources including network bandwidth [1]. In this paper, we propose a simple QoE evaluation method that does not require a reference stream for quality comparison. The method may save significant resources and investments for video service providers.

No-reference (NR) quality assessment is a relatively new research direction with promising applications. A NR metric can be used to estimate QoE of a multimedia presentation without using original audiovisual media streams for reference. Making objective quality assessment without source videos/images is difficult, and so far only a few algorithms have been brought forward. The NR evaluation method used in this paper, originally as a NR perceptual quality assessment for JPEG compressed image, is proposed by Z. Wang, H.R. Sheikh, and A.C. Bovik [9]. As the model considers the assessment of quality degradation due to errors in both compression and transmission, it is modified in our research to measure only the quality impacted caused by transmission errors.

The study presented in this paper considers compressed video being transmitted as our source video, and focus on the distortions introduced by transmission errors, with the network condition as the only factor to influence QoE. In the Section 2, we will illustrate the NR evaluation method and the parameters representing network condition. In Section 3, we will discuss the relationship between No-Reference evaluation values and the parameters of network condition. In Section 4 we will propose the future work and possible usage of this method in the industry.

2. No-Reference quality assessment and parameters of network condition

Firstly, we will analyze the structure of MPEG-2 video sequences, which is illustrated in Figure 3. It illustrates the unit of structure of a whole image/frame. A frame is composed of an inerratic 8×8 pixels blocks. We denote the whole image/frame signal as $x(m,n)$, $m \in [1,M]$, $n \in [1,N]$ (M and N stand for the number of rows and columns), and calculate a differencing signal along each horizontal line,

$$d_h(m,n) = x(m,n+1) - x(m,n), \quad n \in [1, N-1]$$

Then we will calculate the features mentioned above from horizontal and vertical directions.

$$B_h = \frac{1}{M(\lfloor N/8-1 \rfloor)} \sum_{i=1}^{M} \sum_{j=1}^{\lfloor N/8-1 \rfloor} |d_h(i,8j)|$$

$$B_v = \frac{1}{N(\lfloor M/8-1 \rfloor)} \sum_{j=1}^{N} \sum_{i=1}^{\lfloor M/8-1 \rfloor} |d_v(8i,j)|$$

Where B_h and B_v stand for the estimation of average differences across block boundaries. Let

$$B = \frac{B_h + B_v}{2}$$

Where B stands for the evaluation of blockiness of the whole image/frame. B_h and B_v will be small assuming that on the border of two joint 8×8 blocks the differencing signal is small. If B increases greatly, it means errors have occurred in the transmission process. This conclusion will be further elaborated in our following discussion of experimental results (see Figure 5).

In the compression process of MPEG-2 video sequences, the 8×8 pixel coding blocks is the unit for DCT-based image transformation. After the transformation, the DCT-based coefficients in each 8×8 pixel coding block will delegate the information of the coding block, which is used by the decoder.

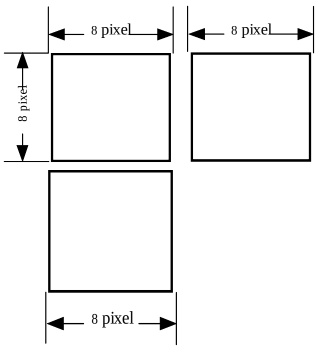

Figure 3. The calculation of the 8×8 pixels coding block

The coefficients will be stored inside RTP/UDP packets and transmitted to the destination. If a packet is lost, some of the coefficients are lost. Decoders can not restore the exact source video sequences in the decompression process. Then Blockiness effect occurs.

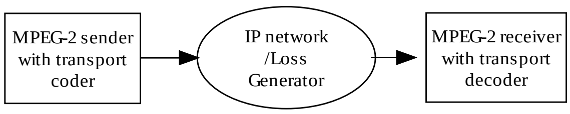

Figure 4. Experiment testbed

The quality of video sequences depends on how many blocks can not been restored. The more the influenced blocks are, the inferior the quality of images is. We devise a lot of experiments and the experimental testbed is composed of three components as indicated in Figure 4.

The transport stream recoded by a JVC HDV Camera was coded with Bit-rate=20Mbps, Frame Rate=30fps, and Frame Size=1280 × 720. In our experiments, we use the software developed in our lab as a loss generator simulating various loss patterns for simulation. The receiver decodes the received stream and stores the extracted frames. The decoder in our experiments is Elecard MPEG-2 Video Decoder, which is used in Elecard MPEQ Player.

In order to quantify QoE, we use NR quality evaluation value vector. Our experiments focus on a series of 20 seconds HDTV (High Definition TV) video sequences. After the video sequences are decoded and stored as a list of frames, we use NR quality evaluation function to generate the quality metric of images. If the transmission channel has been influenced by some impulse which brings packet loss into our video sequences, there will be great distortions on end-users QoE. As shown in Figure 5, the quality value vector of packet loss impacted video sequences can be greatly different from those of the original value vector.

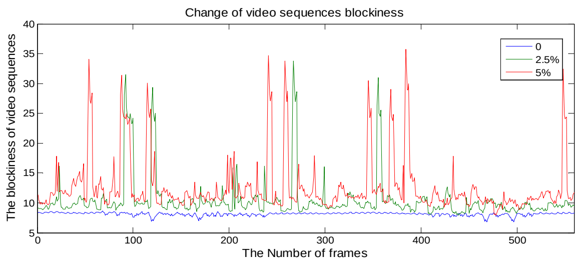

Figure 5. Video sequences blockiness within 20s' video sequences

Figure 5 is the evaluation of blockiness of a series of video sequences. The curves of different colors stand for the different packet loss rate. The blue curve stands for no loss, the green one for 2.5%, and the red one for 5%. It is obvious that the no-loss curve is the smoothest and maintains constant compared to the other two. The red curve with larger packet loss rate proves the lowest quality of video sequences, where the quality evaluation value vector varies quickly and acutely. This indicates that the packet loss rate has some correlation with the quality of video sequences. We will give a analytical framework of this relationship to uncover that the network parameters have strong relationship with the perceptual quality, and the extent to which the users QoE correlates with the network parameters.

It can be validated that the distribution of the quality value vector fits the Gauss distribution under large samples when there is no packet loss in the video sequences transmission. This is the premises of our following reasoning. This property of the quality value vector is important statistically to estimate the precision and guarantee the reliability of our correlation model between the quality value vector and the network condition measures.

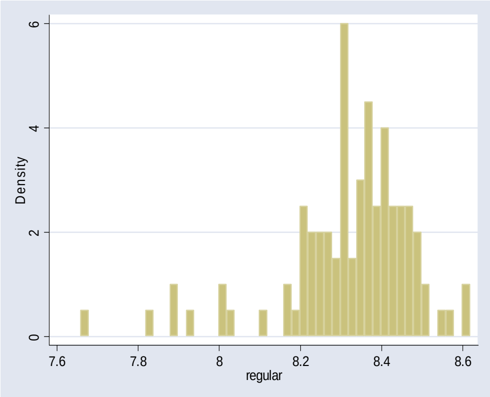

Figure 6. Distribution density of quality value vector (size: 100)

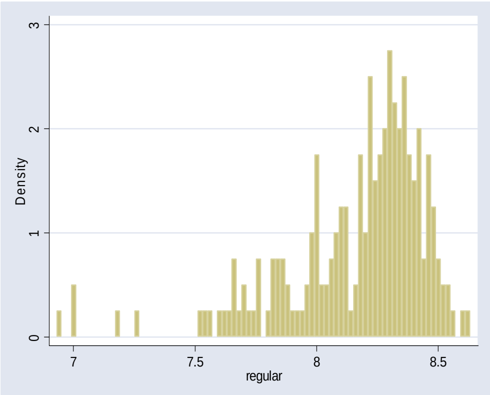

Figure 7. Distribution density of quality value vector (size: 200)

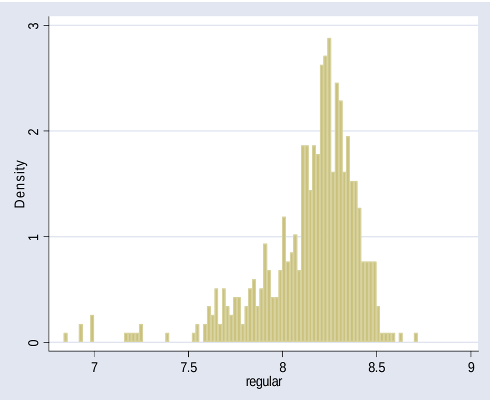

Figure 8. Distribution density of quality value vector (size: 591)

Figure 6 ~ Figure 8 are results from our tests, and the value can be found to be in the range of [6.8 to 8.7], where more than 80% of the values are in the range of [8, 8.5] when the distribution can be treated as a normal one. Considering that the metric value is related with different scenes, when the number of the frames/images we select is enough, the quality value will be of normal distribution.

We define two parameters which can depict the situation of users QoE

$$meanquality = E(frameseq) = \frac{1}{Num}\sum_{i=1}^{Num} framq_i \qquad (1)$$

$framq_i$ stands for the quality value from our NR metric. If the mean quality is larger, we can infer that the quality of the frame is more inferior.

When studying network condition, we focus on packet loss. Several practically measurable parameters are selected, as introduced in RFC 3357 [10].

The first parameter is the average distance of the lost packets,

$$meandistance = \frac{1}{N}\sum_{i=0}^{N-1}(losspos_{i+1} - losspos_i) \qquad (2)$$

Video sequences streams are composed of RTP/UDP packets. We can take the streams as a vector of values. If one packet is lost, we can record the position. For example, suppose the packets sent at position 2, 5, 7, 9, 10, … , are lost, we will get a list of vector

$$1\ \underbrace{0}_{2}\ 1\ 1\ \underbrace{0}_{5}\ 1\ \underbrace{0}_{7}\ 1\ \underbrace{0}_{9}\ \underbrace{0}_{10}\ \ldots$$

The distance vector can be obtained from the original loss vector. In this example, the distance vector of the lost vector above is 3, 2, 2, 0, … , and the average length of the loss vector defined as “meandistance” can be easily calculated from equation (2).

To account for the change of packet loss and better represent the packet loss pattern, “vardistance” is also considered. For example, the two distance vector 1,1,1,1,1 and 0,0,0,0,5 have the same meandistance, but obviously their respective users QoEs are quite different, according to the characteristics of MPEG-2 stream. Actually, with the first loss pattern, we can see a series of consecutive video sequences with a lot of blocks inside every frame. While with the second one, sudden blank frame occurs in the video sequences rather than any blockiness in the individual frames because five consecutive packets are lost and the decoder cannot reconstruct the last frame. Clearly, the two loss pattern examples cause quite different users QoE.

Therefore, the second parameter “vardistance” is defined to reflect the degree of decentralization, as

$$vardistance = \frac{1}{N-1}\sum_{i=0}^{N-1}(losspos_{i+1} - losspos_i - meandistance)^2 \qquad (3)$$

3. The regression analysis and results

With the two selected parameters as a measure of network condition, the influence of different network parameters can be compared using regression analysis to study the extent to which they can individually and together characterize the perceptual quality.

To build up the relationship between the parameters of the video quality evaluation value vector and those of the network situation, many methods, such as ANOVA, CANOVA, ARMA, and ARIMA, can be used. In this paper, linear regression analysis is employed in our research for the cost and speed consideration of implementation in practical usage.

1. MQ vs. MD

In linear regression, the correlation of meanquality (denoted as MQ) vs. meandistance (denoted as MD) is

$$MQ = \alpha_0 + \alpha_1 MD + u_1 \qquad (4)$$

In this paper, MQ stands for users QoE. From the samples in our experiment, the regression results in equation (5)

$$MQ = 9.285411 - 0.0045895MD + u \qquad (5)$$

We define $\hat{MQ}$ as the fitted value from the regression equation. In order to evaluate the effects of the regression result, we use Good-of-Fit as the metrics. For each *i*, let

$$MQ_i = \hat{MQ}_i + \hat{u}_i$$

The covariance between $\hat{MQ}_i$ and $\hat{u}_i$ is zero because $\hat{MQ}_i = \beta_0 + \beta_1 MD$ and $\sum_{i=1}^{N} MD * \hat{u}_i = 0$. So we define the total sum of squares (SST), the explained sum of squares (SSE), and the residual sum of squares (SSR).

$$SST = \sum_{i=1}^{N} (MQ_i - \overline{MQ})^2$$

$$SSE = \sum_{i=1}^{N} (\hat{MQ}_i - \overline{MQ})^2$$

$$SSR = \sum_{i=1}^{N} \hat{u}_i^2$$

SST is a measure of the sample variation in the MQ_i. That is to say, it measures how spread out MQ_i are in the sample. Similarly, SSE measures the sample variation in the $\hat{MQ}_i$ (where $\overline{\hat{MQ}} = \overline{MQ}$), and SSR measures the sample variation in $\hat{u}_i$. The total variation in MQ can be always expressed as the sum of the explained variation and the unexplained variation SSR. Thus

$$SST = SSE + SSR \qquad (6)$$

$$R^2 = SSE/SST = 1 - SSR/SST = 1 - (SSR/n)/(SST/n) = 1 - \sigma^2_u/\sigma^2_{MQ} \qquad (7)$$

Detailed proof of (6) and (7) can be found in [10].

R^2 is the ratio of the explained variation to the total variation, as the fraction of the sample variation in MQ that is explained by MD. It always lies between zero and one. In practice, we often use Adjusted R-Square (denoted as $\overline{R^2}$) to evaluate the effects of regression.

In regression equation (7), we can get R^2 = 0.7101 and $\overline{R^2}$ = 0.7074, which means that about 71% of the variation in MQ can be explained by MD.

2. MQ vs. VD

Similarly, the linear regression of the correlation of MQ vs. vardistance (denoted as VD) results in equation (8), and also from equation (7), we can get R^2 = 0.4380 and $\overline{R^2}$ = 0.4328, which means that about 43% variations in MQ can be explained by VD.

3. MQ vs. (MD + VD)

Form the above results, we can conclude that both MD and VD can reflect the change of our evaluation metric MQ, while MD does it better.

If we take both MD and VD into account, a two-variable regression equation can be drawn as equation (9) with R^2 = 0.8367 and $\overline{R^2}$ = 0.8611, meaning that about 86% variations in MQ can be explained by MD and VD.

$$MQ = 9.695456 - 0.0105648MD + 0.0000141VD + u \quad (9)$$

Clearly, VD can help MD to improve the precision of estimating the change of MQ.

With advanced measurement of more network parameters, we can expect more precise estimation. In general, if we get *m* network parameters which may influence the video parameters (e.g. perceptual quality), these network parameters can be ranked according to their impact to the video parameters and the top *n* network parameters can chosen to best estimate the perceptual quality. The basic process is like how we compare the MD vs. VD. And we should use some iterative algorithms to select *n* parameters out of the *m* ones. The selection method may follow the procesure similar to those used in Bioinformatics [13].

The MPEG-2 video syntax defines three different types of pictures: I-frame, P-frame (Predicted frame) and B-frame (Bidirectional frame). Because of the characters of MPEG-2 video, frames of different types have different sizes. So that errors in different frames have different influence on the perceptual quality. These phenomena can be observed clearly in [11]. So the distribution of lost packet and the different importance of the packets influence the quality evaluation value vector, which needs more research.

4. Conclusion and further discussion

This paper put forward a new analytical framework between evaluation of network parameters and MPEG-2 video sequences quality parameters. It is useful in IPTV industry by helping IPTV services providers to figure out the users QoE with minimal cost. The emergence of NR quality assessment of videos/images provides the possibility to measure the video sequences remotely.

The analytical framework in this paper set up the bridge between the network situation and NR quality assessment of videos/images. We focus on the packet loss and draw two parameters from the network: MD and VD. We choose an evaluation algorithm and draw the mean of quality evaluation value vector as the quality parameters related to the video sequences. The parameter, MQ, can reflect the overall quality of a series of video sequences.

From the analytical framework, we can see when the packet loss is the unique cause for the quality degradation in video sequences, we can come to a conclusion that MD can estimate the change of video sequences evaluation metrics better than VD. While if taking both MD and VD into account, we can improve the precision of estimation greatly. When more detailed information about the video sequences is available, we expect one can better reflect the change of perceptual quality with the network parameters, and contribute greatly to IPTV industry.

5. References

[1] S. Wolf and M.H. Pinson, "In-service performance metrics of MPEG-2 video systems", in Proc. Measurement Techniques of the Digital Age Technical Seminar, International Academy of Broadcasting (IAB), ITU, and Technical University of Braunschweig, Montreux, Switzerland, Nov, 1998.

[2] M. Yuen and H.R. Wu, "A survey of hybrid MC/DPCM/DCT video coding distortions", Signal Processing, Vol. 4, No. 11, pp.317-320, 1998.

[3] X.X. Lu, R.O. Morando, and M.E. Zarki, "Understanding video quality and its use in feedback control", in Proc. International Packet Video Workshop, 2002.

[4] N. Feamster and H. Balakrishnan, "Packet loss recovery for streaming video", in Proc. International Packet Video Workshop, 2002.

[5] S. Tao and R. Guerin, "Application-specific path switching: a case study for streaming video", in Proc. of ACM Multimedia, Oct, 2004.

[6] C.T. Chen, "Error detection and concealment with an unsupervised MPEG2 video decoder", Journal of Visual Communication and Image Representation, Vol. 6, No. 3, pp.265-279, 1995.

[7] Y.Q. Zhang and X.B. Lee, "Performance of MPEG codecs in the presence of errors", in Proc. SPIE, Vol. 2094, pp.1739-1749, 1993.

[8] J.M. Boyce and R.D. Gaglianello, "Packet loss effects on MPEG video sent over the public Internet", in Proc. ACM International Multimedia Conference, 1998.

[9] Z. Wang, H.R. Sheikh, and A.C. Bovik, "No-reference perceptual quality assessment of JPEG compressed images", in Proc. ICIP, Vol. 1, 2002.

[10] J.M. Wooldridge, "Introductory econometrics: a modern approach", 2e, by South-Western, a division of Thomson Learning, 2003.

[11] H.X. Rui, C.R. Li, and S.K. Qiu, "Evaluation of packet loss impairment on streaming video", in Proc. International Packet Video Workshop, 2006.

[12] IneoQuest Article, "MDI/QoE for IPTV and VoIP quality of experience for media over IP", IneoQuest Video over IP Quality Specialists, 2006.

[13] L.J. Wu and S. Li, "Combined literature mining and gene expression analysis for modeling neuro-endocrine-immune interactions", in Proc. International Conference in Intelligent Computing, 2005.