Estimating Perceived Video Quality from Objective Parameters in Video over IP Services

Pedro de la Cruz Ramos[1], Joaquín Navarro Salmerón[1], Raquel Pérez Leal[2], Francisco González Vidal[1]

[1]Departamento de Ingeniería de Sistemas Telemáticos
Universidad Politécnica de Madrid
Madrid, Spain
{pcruzr, navarro, vidal}@dit.upm.es

[2]Departamento de Teoría de la Señal y Comunicaciones
Universidad Carlos III de Madrid
Madrid, Spain
rpleal@tsc.uc3m.es

Abstract — **In Video over IP services, perceived video quality heavily depends on parameters such as video coding and network Quality of Service. This paper proposes a model for the estimation of perceived video quality in video streaming and broadcasting services that combines the aforementioned parameters with other that depend mainly on the information contents of the video sequences. These fitting parameters are derived from the Spatial and Temporal Information contents of the sequences. This model does not require reference to the original video sequence so it can be used for online, real-time monitoring of perceived video quality in Video over IP services. Furthermore, this paper proposes a measurement workbench designed to acquire both training data for model fitting and test data for model validation. Preliminary results show good correlation between measured and predicted values.**

Keywords - Video over IP, Perceived Quality, Quality Models, Quality of Experience, Quality of Service.

I. INTRODUCTION

User Quality of Experience (QoE) is a determining factor for successful deployment of emerging multimedia services. QoE is easy to understand, but complex to implement in real systems. This complexity is mainly due to the difficulty of its modeling, evaluation and translation into Quality of Service (QoS) parameters.

A complete QoE management procedure should encompass at least: monitoring the user experience when consuming the service; adapting the provisioning of the content to the varying context conditions; predicting potential QoE level degradation; and recovering from QoE degradation due to system changes. In order to have a complete control of the final user experience, all these tasks must be performed in-service and in a coordinated way.

Among multimedia services, Video over IP applications have reached a remarkable market penetration. Furthermore, Video over IP customers expect a QoE comparable to traditional broadcast systems. So the ability to measure, estimate and monitor user perceived quality in near real time and to relate it to network conditions, is critical for Video over IP service providers.

This paper focuses on the perceived video quality aspects of Video over IP streaming and broadcasting services. A model for estimating the Video Quality Metric (VQM) [1] as defined in ITU-T J.144 [2] is proposed.

Subjective quality measurements, as those defined in ITU-T P.910 [3], are undoubtedly the most precise, and constitute the benchmark for any estimation model. However, they are costly, both in time and resources. Thus, our approach has been to estimate an objective perceptual distortion metric, originally defined as a Full Reference (FR) measure, from coding and Network QoS parameters, using a model similar to those suggested in [4], [5], [6] and [7].

The proposed model takes as input easily measurable video coding and Network QoS (NQoS) parameters, and includes some fitting parameters that depend mainly on the information contents of the video sequences. A method for computing them from Average Spatial and Temporal Information content measures (ASI/ATI), similar to those defined in ITU-T P.910 [3], is proposed. All the values required for the estimation can be obtained without reference to the original video sequence, enabling online, real-time evaluation of perceived video quality in Video over IP services.

In the following sections previous work is reviewed; the estimation model is proposed; the method for computing the fitting parameters is described; a measurement workbench is presented; the main conclusions are summarized; and some future work is outlined.

II. RELATED WORK

In [4], a comprehensive model, based on theoretical considerations, is proposed in order to relate several coding and network parameters to the Perceptual Distortion Metric (PDM) of MPEG-2 sequences. The coefficients of this model mainly depend on the complexity (information contents) of the analyzed sequence.

The dependence of VQM on Video Coding Rate (VCR), display format (resolution), codec type and "motion contents", is analyzed for MPEG-2 and H.264 Advanced Video Coding (AVC) sequences in [5]. Although this model takes into account the effects of codec type and coding parameters, it obviates the dependence of VQM on the transmission network parameters.

In the previous models, the variation of the chosen metric follows a negative power function of VCR. Regarding Packet Loss Ratio (PLR), [4] proposes a linear variation while [5] does not consider its effect at all.

Reference [6] estimates the Perceived Video Quality of H.264 sequences combining coding and network QoS

 ISBN: 978-1-61208-193-9

parameters (namely VCR and the Packet Loss Frequency, PLF) and codec features in a parametric packet level model. This model states that the variation of the Perceived Video Quality with VCR follows a logistic function while its variation with PLF follows a negative exponential.

In [7], a parametric null reference (NR) model, called "Temporal-Visual (T-V) Model" is proposed. The objective of this model is to estimate the Perceived Video Quality of MPEG-2 and H.264 sequences, using network QoS, coding and other parameters. This model states that the Perceived Video Quality is related to VCR by an exponential function while its variation with PLR follows a logistic function.

One of the key aspects when designing a model is the determination of the "fitting" parameters. In the previous proposals different approaches are followed. In [4], the fitting process is performed for each individual sequence. In [5], the sequences are classified in classes according to their 'motion contents' and values are assigned to the parameters for each group of sequences. In [6] and [7], neither the contents nor the spatial or temporal complexity of the sequences are considered.

None of the analyzed models completely fulfill our needs. Some of them are too specific for a particular kind of application or propose forms of variation that do not correspond to our measurements, which rather suggest a (positive or negative) power function. Most of them estimate the subjectively perceived video quality or metrics other than VQM. In [5], VQM is estimated, but it does not take into account the effect of the transmission network. Furthermore, none of the reviewed proposals include the effect of the complexity and/or information contents of the video sequences.

All these reasons lead us to develop a new model for online, real-time estimation of VQM in Video over IP streaming and broadcasting services, using coding and network QoS parameters and the complexity and information contents of the video sequences.

III. MODEL DESCRIPTION

Different measurements, obtained using the Video Quality Experts Group (VQEG) FR-TV1 test sequences [10] and our Measurement Workbench (described later), confirmed the variation of VQM with coding parameters according to the model of [5]. However, these measurements also showed that the variation of VQM with PLR is far from linear in most of the cases. Figure 1 shows the effects of coding and packet loss.

Figure 1a shows the variation of VQM with VCR for all sequences coded using H.264, prior to transmission (i.e., with no transmission losses). VCR is the actual Average VCR (Video Data Size/Duration).

The relation between VQM and PLR is shown in Figure 1b. In this measurement all the sequences have been coded using H.264 with VCR=5 Mbps. The plotted VQM is the average result of several repetitions with the same PLR in order to attenuate random effects. The value for PLR=0 (no losses) is the value of VQM prior to transmission as given in Figure 1a.

VQM can be split into two parts in order to separate the effects of coding and transmission:

$$VQM = VQM_C + VQM_L \qquad (1)$$

where

VQM_C is the contribution of coding to VQM.

VQM_L is the contribution of packet losses to VQM.

By plotting VQM_C and VQM_L in logarithmic scale, it can be noticed that both curves fit very well to a power function, as they are nearly linear in both cases. They can be expressed as:

$$VQM_C = VQM_{REF} \cdot (VCR/VCR_{REF})^{-K_C} \qquad (2)$$

$$VQM_L = (1-VQM_C) \cdot (PLR/PLR_1)^{K_L} \qquad (3)$$

where

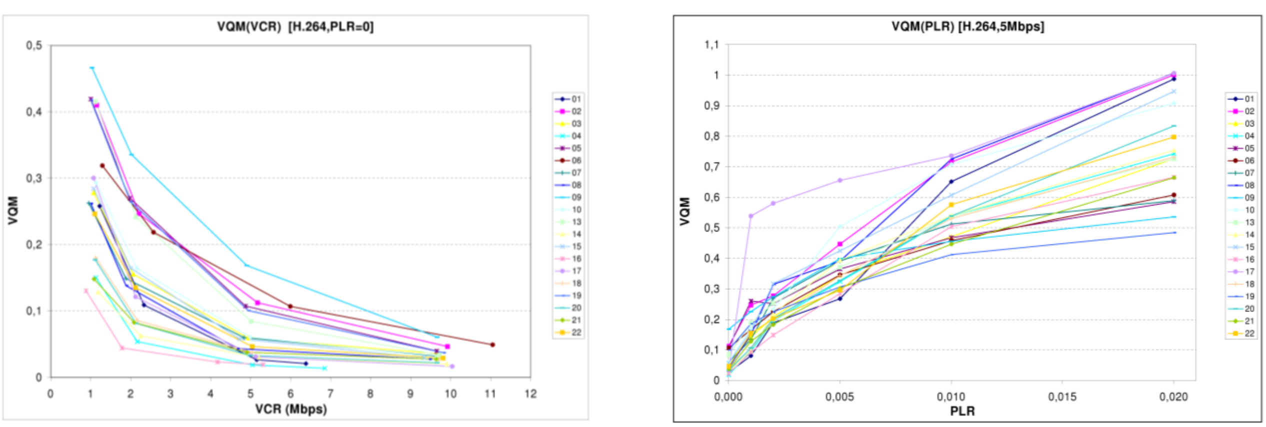
VCR_{REF} is a reference VCR (e.g., 1Mbps).

VQM_{REF} is the value of VQM at the reference VCR.

PLR_1 is the value of PLR for which VQM = 1.

VQM_{REF} and K_C depend on the codec, the coding parameters (except VCR), and the characteristics of the video sequence (e.g., spatial and temporal complexity, information contents, etc.).

PLR_1 and K_L depend on the codec, the coding parameters (including VCR), and the characteristics of the video



(a) (b)

Figure 1. Variation of VQM with VCR (for PLR=0) and PLR.

sequence. Their variation with VCR fits very well to a function of the form:

$$F(VCR) = A + B \cdot VCR \cdot (1+C \cdot e^{-(VCR/D)^{\wedge}K}) \qquad (4)$$

where A, B, C, D, K are fitting parameters that depend on the codec, the coding parameters (except VCR) and the characteristics of the video sequences (type, format and information contents).

For K≠2 this function approximates to a Weibull curve on top of a linear asymptote. For K=2 it corresponds to a Rayleigh curve, also on top of a linear asymptote. Figure 2 shows the fitting of this function to the values of K_L for a group of sequences coded using H.264.

According to this model, for a given PLR, there is a VCR that minimizes VQM, i.e., maximizes the perceived quality. The consequence is that for higher coding rates, and against the common assumption, the perceived quality decreases due to the increment of packet losses. Therefore, in real systems with transmission errors, increasing the coding rate beyond a certain limit is not only useless (as users don't perceive the difference), but may even be counterproductive. This behavior was already noticed in [4].

IV. ESTIMATION OF MODEL PARAMETERS

As seen in the previous section, the characteristics of each sequence, i.e., its type, complexity and information contents, directly influence the perceived video quality. Therefore, a crucial aspect is how to compute the model parameters for each video sequence, without having to fit the model specifically for each of them. This paper proposes the use of two measures similar to the Spatial/Temporal Information (SI/TI) measures, described in [11]. SI/TI measurements evaluate the spatial/temporal information detail in a way similar to the perception of a human viewer. They are standardized in ITU-T Recommendation P.910 [3]. These measurements are rather easy to obtain using well-known techniques such as the Sobel filter (a simple high-pass, edge enhancement digital filter widely used in image processing) and pixel-wise difference.

However, our preliminary results concluded that SI/TI measurements, as originally described, i.e., based on the

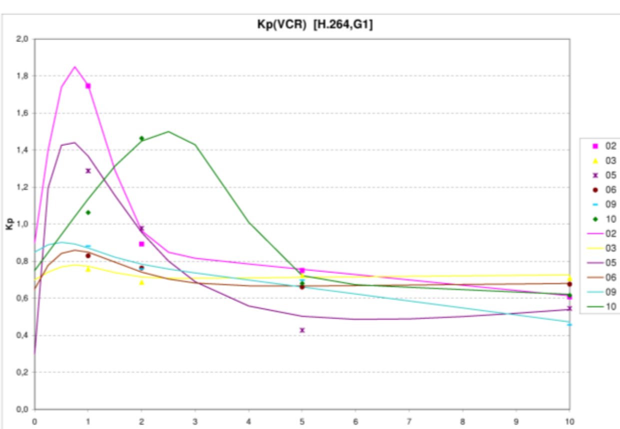


Figure 2. Variation of K_L with VCR for a group of sequences.

maximum SI/TI values of the frames in the sequence, are too sensitive to exceptional SI/TI values of individual frames [8]. Therefore, in order to attenuate this effect, the Average Absolute Spatial/Temporal Information (ASI/ATI) measurements are defined as follows:

1) Use the absolute value of the pixel-wise difference of luminance values of successive frames to compute the Temporal Information values of each frame.

2) Take the average of the SI/TI values of all frames as the ASI/ATI value of the sequence.

ASI/ATI measures will be used as indexes into precomputed "complexity tables". The model parameters for a given sequence will be computed by linear interpolation in these tables. The methods for populating the Complexity Tables and using them to compute the model parameters for arbitrary sequences are described in [8].

The proposed method enables online, real-time monitoring of perceived video quality, because the whole process (ASI/ATI computation, table lookup, interpolation, and model evaluation) takes much less time than the duration of the sequences. In addition, all values required for VQM estimation can be either obtained from the Network Management System (NMS) or measured at the receiving side, so no measurements on the reference sequence are required.

V. MEASUREMENT WORKBENCH

This section describes the measurement workbench that was implemented in order to obtain training data for model fitting and test data for model validation [9]. Figure 3 shows its functional architecture, which was implemented using the following tools:

- Encoder/Decoder: FFmpeg 0.6.1-2/4 + libX264 [for H.264]
- Transmitter/Receiver: Videolan VLC 1.1.5/7
- Network Simulator: NetEm (Linux Kernel 2.6.35)
- Information Measurement: STIX 0.9
- Distortion Measurement: ITS/NTIA BVQM 1.4
- QoS Measurement: WireShark 1.6.0

Specific tools were developed in order to perform ASI/ATI measurements and frame loss concealment. Frame loss concealment is required because the received and reference sequences must have equal length for BVQM to work adequately. The operation of the frame concealer is based on detecting the lost frames and duplicating the previous one (i.e., freezing).

The workbench comprises four physical nodes. The first one is the emitter station that performs encoding and transmission operations. The second one is the receiver station, responsible for reception, decoding and frame loss concealment. The third node is the network simulator, capable of simulating different network parameters and scenarios. The last one is the measurement workstation, used to perform VQM, QoS and ASI/ATI measurements.

All these nodes were physically implemented using DELL Optiplex 755 PCs with Intel Core 2 Duo processors at 2.66GHz with 3GB of RAM. The emitter and receiver stations and the network simulator run under Ubuntu Linux

 ISBN: 978-1-61208-193-9

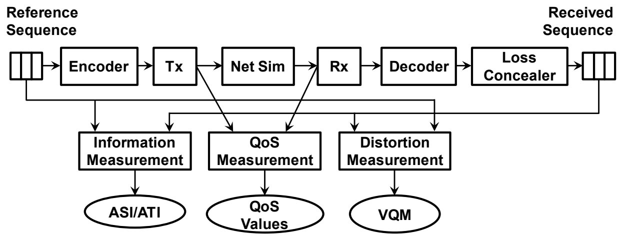


Figure 3. Measurement Workbench Functional Architecture.

10.04.2 LTS, and the measurement workstation under Windows XP Professional SP3. The nodes communicate through a 100Mbps Fast Ethernet LAN.

The test sequence database [10] includes both high and low motion (including static) sequences, spatially simple as well as complex, both natural (filmed) and artificial (animation or computer-generated). All sequences are in ITU-R BT.601 UYVY (Big-YUV) format (either 525lines@60Hz or 625lines@50Hz).

Measurements were made for all sequences coded in MPEG-2, MPEG-4 and H.264 AVC, for several VCR and PLR values. In order to account for random effects of packet losses, these measurements were repeated several times for the same nominal VCR and PLR values. In total more than 6,000 data points were collected. These data will be statistically analised in order to validate the accuracy of the model. Preliminary results show good correlation between measured and predicted values (see Table I).

VI. CONCLUSION AND FUTURE WORKS

This paper proposed a new model for online, real-time estimation and monitoring of perceived video quality in Video over IP streaming and broadcasting services, using the Video Quality Metric (VQM) as objective measure. This model is based on video coding and Network Quality of Service (NQoS) parameters. Our model shows that the contributions to VQM from coding (VQM_C) and packet losses (VQM_L) follow power functions of the Video Coding Rate (VCR) and Packet Loss Ratio (PLR) respectively.

Additionally, the model includes fitting parameters that depend mainly on the complexity (information contents) of the video sequence. These parameters are estimated using the Average Absolute Spatial and Temporal Information (ASI/ ATI) contents of the sequence.

A measurement workbench was implemented. It comprises several nodes, such as emitter and receiver stations, a network simulator and a measurement workstation. This workbench was used with a public test sequence database in order to obtain training data for fitting the model parameters. Preliminary results show good correlation between measured and predicted values.

The following points remain open: sequence classification based on features other than ASI/ATI, and use of different complexity tables for each group of sequences; influence of coding parameters other than VCR; effect of NQoS parameters other than PLR (e.g., Packet Error Ratio (PER) and/or Bit Error Ratio (BER)); influence of error/loss patterns (distribution), in particular the Average Burst Length (ABL); effect of extreme variation of the ASI/ATI values of received (distorted) sequences with respect to that of original sequences, in the computation of fitting parameters from the complexity tables; definition of spatial and temporal information measures based on chrominance values, and inclusion of them in the estimation model.

TABLE I. PRELIMINARY STATISTICAL RESULTS

Codec	Correlation	Avg.Error	RMSE
MPEG-2	0.9519	0.0339	0.0703
MPEG-4	0.9471	0.0551	0.0704
H.264	0.9462	0.0549	0.0749
ALL	0.9511	0.0487	0.0722

ACKNOWLEDGMENT

This research was partially supported by the Spanish Ministry of Science and Innovation grant TEC2008-06539 (ARCO Project).

REFERENCES

[1] M. H. Pinson and S. Wolf, “A New Standardized Method for Objectively Measuring Video Quality,” IEEE Transactions on Broadcasting, Vol. 50, No. 3, pp. 312-322. September 2004.

[2] ITU-T J.144, “Objective Perceptual Video Quality Measurement Techniques for Digital Cable Television in the Presence of a Full Reference,” International Telecommunication Union, March 2004.

[3] ITU-T P.910, “Subjective Video Quality Assessment Methods for Multimedia Applications,” International Telecommunication Union, April 2008.

[4] P. Frossard and O. Verscheure, “Joint Source/FEC Rate Selection for Quality-Optimal MPEG-2 Video Delivery,” IEEE Transactions on Image Processing, Vol. 10, No. 12. December 2001.

[5] J. Joskowicz, J. C. López Ardao, M. A. González Ortega, and C. López García, “A Mathematical Model for Evaluating the Perceptual Quality of Video,” Second International Workshop on Future Multimedia Networking (FMN 2009), Coimbra, Portugal, June 2009. In Lecture Notes on Computer Science (LNCS) 5360, pp. 164-175, Springer Verlag, 2009.

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

[7] M. N. Garcia, R. Schleicher, and A. Raake, “Impairment-Factor-Based Audiovisual Quality Model for IPTV: Influence of Video Resolution, Degradation Type, and Content Type,” EURASIP Journal on Image and Video Processing, Volume 2011, Article ID 629284, pp. 14. 2011.

[8] P. de la Cruz Ramos, F. González Vidal, and R. Pérez Leal, “Perceived Video Quality Estimation from Spatial and Temporal Information Contents and Network Performance Parameters in IPTV,” Proc. of the Fifth IARIA International Conference on Digital Telecommunications (ICDT 2010), pp. 128-131, Athens, Greece, June 2010.

[9] E. Álvarez Villacé, “Design and Implementation of a Measurement Workbench for Estimation of Perceived Video Quality in IPTV,” Master Thesis, Polytechnical University of Madrid, July 2011.

[10] VQEG, “Final Report from the Video Quality Experts Group on the Validation of Objective Models of Video Quality,” ITU-T SG 9,Contribution COM 9-8, June 2000.

[11] A. A. Webster, C. T. Jones, M. H. Pinson, S. D. Voran, and S. Wolf, “An Objective Video Quality Assessment System Based on Human Perception,” SPIE Conference on Human Vision, Visual Processing, and Digital Display, 1993.