Perceived Video Quality Estimation from Spatial and Temporal Information Contents and Network Performance Parameters in IPTV

Pedro de la Cruz Ramos, Francisco González Vidal, Raquel Pérez Leal

pcruzr@dit.upm.es   vidal@dit.upm.es   rperez@dit.upm.es

Departamento de Ingeniería Telemática

Universidad Politécnica de Madrid

Madrid, Spain

Abstract—The paper proposes a model for estimation of perceived video quality in IPTV, taking as input both video coding and network Quality of Service parameters. It includes some fitting parameters that depend mainly on the information contents of the video sequences. A method to derive them from the Spatial and Temporal Information contents of the sequences is proposed. The model may be used for near real-time monitoring of IPTV video quality.

Keywords - Quality of Experience; Video Quality; Spatial Information; Temporal Information; Video Quality Metric.

I. INTRODUCTION

IPTV customers expect a Quality of Experience (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 IPTV service providers.

This paper focuses on the perceived video quality aspects of IPTV QoE. A new model for estimating the Video Quality Metric (VQM) [1] as defined in ITU-T J.144 [2] is proposed.

The model is based on an innovative combination of previous models [3][4][5], and takes as input easily measurable video coding and Network Quality of Service (NQoS) parameters. It includes some fitting parameters that depend mainly on the information contents of the video sequences. A new method for computing them from the standard Spatial Information (SI) and Temporal Information (TI) contents of the sequence, as defined in ITU-T P.910 [6], is proposed. All the values required for the estimation can be obtained without reference to the original video sequence.

In the following sections previous work is reviewed, the estimation model is proposed, the method for computing the fitting parameters is described, results showing the feasibility of the approach are presented, the main conclusions are summarized, and some future work is outlined.

II. RELATED WORK

In [3], the effect of Video Coding Rate (VCR) and Video Transport Stream (VTS) Packet Loss Ratio (PLR) on the Motion Picture Quality Metric (MPQM) [7] is investigated.

A fitting model is proposed, whose coefficients depend on video coding parameters and sequence complexity.

In [4], the effect of several coding and network parameters on the Perceptual Distorsion Metric (PDM) [8] is studied. Instead of just providing a fitting model, a more comprehensive model, based on theoretical considerations, is derived, whose coefficients mainly depend on the complexity (information contents) of the sequence.

In [5], the dependence of VQM [1] on VCR, codec type, display format (resolution) and 'motion contents', is thoroughly investigated. This model considers in detail the effects of codec type and coding parameters, but ignores the effect of the transmission network.

An important aspect is the determination of the 'fitting' parameters. In [3][4], the parameters are determined by fitting the model for each individual sequence. In [5], the sequences are classified in groups according to their 'motion contents' and values are assigned to the parameters for each class. However, nothing is said about the criteria used to classify the sequences. Thus, none of these models may be directly used for real-time monitoring of IPTV video quality.

III. VIDEO QUALITY ESTIMATION MODEL

As objective measure of video quality we have selected VQM. It was evaluated by the Video Quality Experts Group (VQEG) [9], obtaining an excellent score (it was the only model statistically better than all other for all tests). It has been standardised by ITU-T in recommendation J.144 [2], and the software for computing it is publicly available. Furthermore, there are publicly available results [5] for VQM as a function of coding parameters for a subset of the test sequences we have used.

Our hypothesis is that, as both PDM and VQM are perceptual distortion metrics, their dependency on coding and NQoS parameters should be very similar. The results in [5] support this hypothesis. Our proposal is then to use a combination of the models described in [4][5] in order to estimate VQM from the same parameters:

$$VQM =$$
$$X_S/k \cdot (a \cdot VCR/X_R)^{-\varepsilon_S} + X_L \cdot PLR \cdot (1 + VCR/(2 \cdot P \cdot FR \cdot SF \cdot ABL)) \quad (1)$$

where

VQM	is the Video Quality Metric
VCR	is the Video Coding Rate
PLR	is the Packet Loss Ratio in the Video Transport Stream (VTS)
P	is the VTS packet size
FR	is the Frame Rate (number of frames per second)
SF	is the Slicing Factor (average number of slices per frame)
ABL	is the Average Burst Length (average number of consecutive lost packets)
a	is a fitting parameter that depends on the display format (resolution)
k	is an enhancement factor depending on the codec (and coding parameters)
X_R	is a Reference Coding Rate (RCR)
X_S, X_L, ε_S	are model parameters depending mainly on the complexity (information contents) of the sequence

The key point is how to determine the values of the 'fitting' parameters (X_S, X_L and ε_S) for each video sequence. Our proposal is to use the SI/TI measures [10] for this purpose. These measures are rather easy to compute, being based on simple operations such as the Sobel filter (a simple high-pass, edge enhancement digital filter widely used in image processing) and pixel-wise difference. They measure the spatial/temporal information detail as perceived by a human viewer, and are standardised in ITU-T Recommendation P.910 [6].

SI/TI are used as indexes into precomputed 'complexity tables', so that the model parameters for a given sequence are computed by interpolation between the values for the indexes nearest to the SI/TI values of the sequence.

Figure 1 schematically depicts the process for populating the complexity tables from a set of reference (training) sequences: First, the SI/TI values of each sequence are computed. Then each sequence is encoded, transmitted over a simulated network, and decoded in order to obtain a distorted sequence. The reference and distorted sequences are processed in order to obtain their VQM.

This process is repeated for different values of the coding and NQoS parameters. Then the VQM, coding and NQoS parameter values are fed to a fitting model in order to obtain the optimal values of the fitting parameters. The parameter values are stored in the complexity tables at the indexes corresponding to the SI/TI values of the original sequence. The whole process is repeated for as many reference sequences as possible in order to populate as many 'cells' of the complexity tables as possible.

Once the complexity tables have been populated, they can be used to estimate the value of VQM for a distorted sequence, given the coding and NQoS parameters. The process is depicted in Figure 2. First NQoS, video coding and VTS parameters are obtained from the Network Management System (NMS), the decoder/Set-Top Box (STB), or measured at the network/STB interface. Then the SI/TI values of the received sequence are computed and used
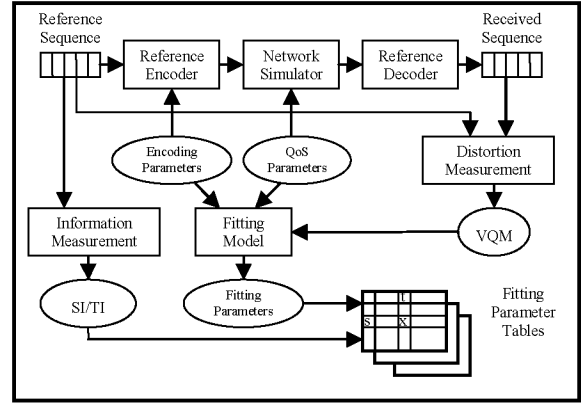

Figure 1.   Complexity table population process.

for interpolation in the complexity tables in order to obtain the fitting parameter values. Finally the coding, QoS and fitting parameter values are fed to the estimation model in order to obtain an (estimated) value for VQM. It must be noticed that no measurements on the reference sequence are required.

IV.   CURRENT RESULTS

The following results are based on the VQEG FR-TV1 test sequences [11].

Table I lists the SI/TI values of the reference sequences. It also includes the 'motion contents' values in [5]. It can be seen that the test sequence set covers a wide SI/TI range.

SRC10 and SRC15 are the same sequence, just coded in different formats. However, their SI/TI values are different. This seems to indicate that for the same scene, the lower the resolution, the higher the SI/TI values.

The correspondence between 'motion contents' and TI would roughly be: low motion for TI<30, medium motion for 30≤TI≤38, high motion for TI>38. However, there are some sequences whose 'motion' and TI values do not match.

Our opinion is that part of the subjective 'motion' impression (and coding difficulty) is caused by spatial, rather than temporal, information contents. For instance: SRC02 (low TI, medium-high SI, high motion), SRC07 (high TI, medium SI, medium motion), SRC22 (low TI, medium-high SI, medium motion). This supports our hypothesis that it is the **combination** of both SI and TI, which should be used as complexity index.
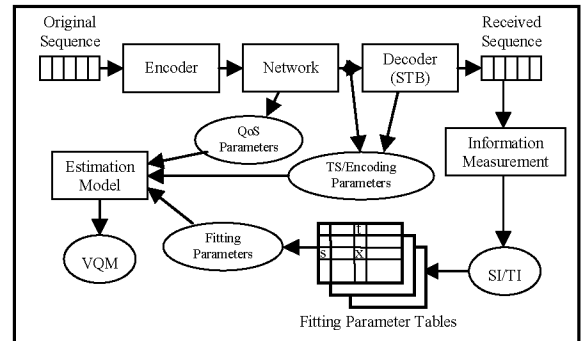

Figure 2.   VQM estimation process.

TABLE I.   SPATIAL/TEMPORAL INFORMATION OF REFERENCE SEQUENCES

Sequence	Name	Format	SI	TI	Motion
src01	Tree	SD TV 625	130	3	
src02	Barcelona	SD TV 625	116	23	High
src03	Harp, music	SD TV 625	145	33	Medium
src04	EBU test	SD TV 625	107	16	Low
src05	Canoe, Valsesia	SD TV 625	82	33	High
src06	Race, F1 car	SD TV 625	107	47	
src07	Fries	SD TV 625	103	64	Medium
src08	Scrolling message	SD TV 625	74	33	
src09	Rugby	SD TV 625	104	43	High
src10	Mobile, calendar	SD TV 625	150	38	Medium
src13	Balloon, Pops	SD TV 525	74	53	High
src14	New York	SD TV 525	77	28	Low
src15	Mobile, calendar	SD TV 525	175	40	Medium
src16	Betes_pas_betes	SD TV 525	110	79	Low
src17	Letters, Le_point	SD TV 525	115	58	High
src18	Autumn, Leaves	SD TV 525	55	9	Low
src19	Football	SD TV 525	90	39	High
src20	Sailboat, Galleon	SD TV 525	179	4	Low
src21	Susie	SD TV 525	53	18	Low
src22	Tempest	SD TV 525	120	24	Medium

SRC16 (high TI, medium SI) is considered as of 'low motion'. The high TI is caused by three 'scene cuts'. If they are ignored, TI decreases to 33 (medium). This may indicate that the current SI/TI definitions (the maximum of the SI/TI values for individual frames) are too sensitive to exceptional values. Alternative definitions based on the average or the Rooted Mean of Squares (RMS) are being investigated.

Part of the motion impression (and coding difficulty) may be caused by (spatial and/or temporal) changes in chromaticity (colour), which are not captured by the SI/TI measures, as they are based on luminance values only. Some (spatial and temporal) Colour Information (CI) measures will be investigated in the future.
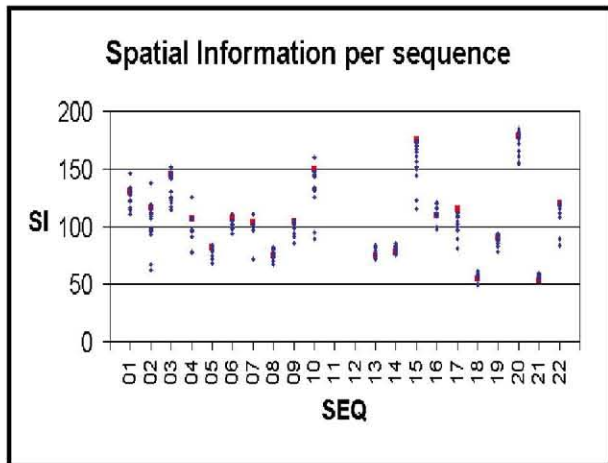
Figures 3 and 4 show respectively the SI/TI values of both the original (reference) and distorted versions of all sequences. It can be seen that the SI/TI values of the distorted sequences may differ substantially from those of the reference sequences. A redefinition of the SI/TI value of a sequence (currently defined as the maximum of the values for all frames) in order to reduce these differences is being investigated.

Figure 5 shows the VQM values for all the distorted versions of each sequence when compared with the corresponding reference sequence, computed using the VQM General Model with Full Reference calibration, as described in ITU-T Recommendation J.144 [2]. It can be seen that the range of distortions is very wide, from almost unnoticeable (VQM<0.2) to quite annoying (VQM≈0.8).

Preliminary Complexity Tables have been computed from the parameter values published in [4][5], assuming MPEG-2 encoding and MPEG-TS transport with P=184 bytes, SF=36 slices/frame (one slice for every 16-line stripe in 625-line SD TV), for X_R=1 Mbps. They are shown in Table II (the extreme points (0,0) and (200,100) are fictitious).

Table III gives the values of the parameters computed by the interpolation procedure for some sequences not included in the tables. The whole process (SI/TI computation, table lookup, interpolation, and model evaluation) takes much less than the duration of the sequences, thus the method can be used for real-time monitoring of perceived video quality.

V.   CONCLUSION AND FUTURE WORK

A new model for estimation of perceived video quality from both video coding parameters and VTS QoS parameters has been proposed. It includes some fitting parameters that depend on the complexity of the video sequence: our proposal is to estimate them from the standard SI/TI contents of the sequence.

Methods have been described for a) computing the fitting parameter values for different SI/TI combinations and storing them in Complexity Tables, and b) deriving the fitting parameter values for a given sequence from the Complexity Tables and using them, combined with the video

Figure 3.   Spatial Information of all sequences.
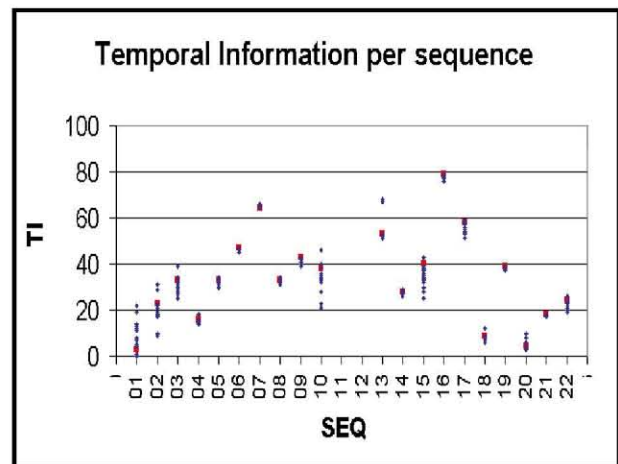

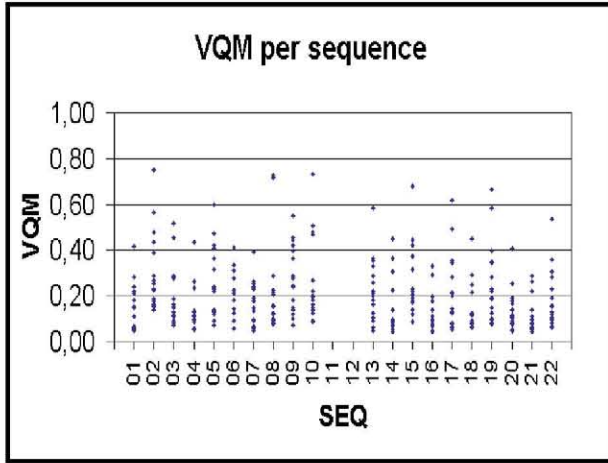Figure 4.   Temporal Information of all sequences.

Figure 5.   VQM values of distorted sequences.

coding and NQoS parameters, to estimate the VQM for the received sequence. All the values required for the estimation can be measured or computed at the receiving side, without any reference to the original video sequence. The estimation process takes much less than the duration of the sequence, thus the method can be used for real-time monitoring of perceived video quality.

Preliminary complexity tables have been computed and used to compute fitting parameter values for sequences not included in the tables, showing the feasibility of the approach.

Some issues have been identified: Influence of display format on SI/TI values; effect of exceptional frame SI/TI values on the overall sequence SI/TI; influence of colour intensity, variability and change on perceived information and/or motion contents.

TABLE II.   COMPLEXITY TABLE (X_R=1MBPS)

SI	TI	X_S	ε_S	X_L
0	0	0.000	0.000	0
53	18	0.210	1.320	94
55	9	0.270	1.140	91
74	53	0.570	1.140	196
77	28	0.210	1.155	137
82	33	0.495	1.065	149
90	39	0.510	1.185	166
103	64	0.390	1.200	211
104	43	0.555	0.990	191
107	16	0.150	0.885	127
110	79	0.225	1.065	235
115	58	0.690	1.110	219
116	23	0.435	0.885	123
120	24	0.210	1.320	118
145	33	0.330	1.095	89
150	38	0.420	0.960	85
179	4	0.195	1.155	30
200	100	1.000	1.500	378

TABLE III.   COMPUTED PARAMETERS FOR SOME SEQUENCES

Sequence	Name	SI	TI	X_S	ε_S	X_L
src01	Tree	130	3	0.158	0.991	89
src06	Race, F1 car	107	47	0.591	1.022	198
src08	Scrolling Message	74	33	0.291	1.152	151
src15	Mobile, calendar (525)	175	40	0.581	1.110	167

Some open points are under investigation: effect of extreme variation of the SI/TI values of distorted sequences with respect to original sequences; suitability of the current SI/TI definitions, based on the maximum of individual frame values; use of spatial/temporal Colour Information (CI) measures, based on chrominance values.

Furthermore, this model will be combined with models for audio quality, lip sync, channel change time, and other aspects of IPTV quality in order to obtain a comprehensive Perceived Quality Estimation Model for the IPTV service.

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, September 2004.

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

[3]   O. Verscheure, P. Frossard, and M. Hamdi, "MPEG-2 Video Services over Packet Networks: Joint Effect of Encoding Rate and Data Loss on User-Oriented QoS," ACM NOSSDAV'98, Cambridge, UK, July 1998.

[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.

[6]   ITU-T P.910, "Subjective Video Quality Assessment Methods for Multimedia Applications," September 1999.

[7]   C. J. van den Branden Lambrecht and O. Verscheure, "Perceptual Quality Measure using a Spatio-Temporal Model of the Human Visual System," SPIE Conference on Digital Video and Compression: Algorithms and Technologies, Vol. 2668, pp. 450-461, San José, California, USA, January/February 1996.

[8]   S. Winkler, "A Perceptual Distortion Metric for Digital Color Video," SPIE Conference on Human Vision and Electronic Imaging, Vol. 3644, pp.175-184, San José, California, USA, January 1999.

[9]   VQEG, "Final Report from the Video Quality Experts Group on the Validation of Objective Models of Video Quality, Phase II (FR-TV2).," ITU-T SG 9, Contribution COM 9-60, September 2003.

[10]   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.

[11]   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.