Porto Alegre, v.14, n.2, jul./dez. 2011.
ISSN impresso 1516-084X ISSN digital 1982-1654

Quantitative Modeling of the Impact of Video Content in the ITU-T G.1070 Video Quality Estimation Function

Modelagem Quantitativa do Impacto de Conteúdo de Vídeo na Função de Estimação da Qualidade do Vídeo do ITU-T G.1070

José Joskowicz
Universidad de La República

J. Carlos López Ardao
Universidad de Vigo

Rafael Sotelo
Universidad de Montevideo

Abstract
In this paper we present an enhancement to the video quality estimation model described in ITU-T Recommendation G.1070 "Opinion model for video-telephony applications", in order to include the impact of video content, for different display sizes and codecs. This enhancement provides a much better approximation of the model results with respect to the perceptual MOS values for a wide range of video contents. SAD (Sum of Absolute Differences) is used as an estimation of the video spatial-temporal activity, and is included as a new parameter in the model. The results are based on more than 1500 processed video clips, coded in MPEG-2 and H.264/AVC, in bit rate ranges from 50 kb/s to 12 Mb/s, in SD, VGA, CIF and QCIF display formats.
Keywords: Video Quality Estimation. Video Codecs. Video signal processing.

Resumo
Neste artigo apresentamos um aprimoramento do modelo de estimação de qualidade de vídeo descrita no "modelo de parecer para aplicações de vídeo-telefonia", descrito pela Recomendação ITU-T G.1070, de modo a incluir o impacto do conteúdo de vídeo, para diferentes tamanhos de visualização e codecs. Este aprimoramento fornece uma melhor aproximação dos resultados do modelo com respeito aos valores perceptual do MOS para uma ampla gama de conteúdos de vídeo. O SAD (soma das diferenças absolutas) é usado como uma estimativa da atividade espaço-temporal do vídeo, e é incluído como um novo parâmetro no modelo. Os resultados são baseados em mais de 1500 videoclipes processados, codificados em MPEG-2 e H.264/AVC, em taxas de bits de 50 kb/s para 12 Mb/s, em formatos de exibição SD, VGA, CIF e QCIF.
Palavras-chave: Estimação de qualidade de vídeo. Codecs de vídeo. Processamento de sinal de vídeo.

JOSKOWICZ, José; LOPEZ ARDAO, J. Carlos; SOTELO, Rafael. Quantitative Modeling of the Impact of Video Content in the ITU-T G.1070 Video Quality Estimation Function. *Informática na Educação*: teoria & prática, Porto Alegre, v. 14, n. 2, p. 191-32, jul./dez. 2011.

1 Introduction

Video-telephony applications are growing quickly in the market, using IP (Internet Protocol) as the underlying protocol. In these emerging applications it is critical to provide an appropriate QoE (Quality of Experience) for the end user, in accordance to the offered service. QoE can be defined as the overall performance of a system, especially from the user perspective. Many factors can affect the QoE in video-telephony, but the audio and video qualities are the most important aspects to consider.

Different evaluations and standardized efforts have been made, and are currently ongoing, in order to derive objective models and algorithms to predict the perceived video quality in different scenarios.

Picture metrics, or media-layer models, are based on the analysis of the video content.

Porto Alegre, v.14, n.2, jul./dez. 2011.
ISSN impresso 1516-084X ISSN digital 1982-1654

These metrics can be classified into FR (Full Reference), RR (Reduced Reference) and NR (No Reference) models. In the first one, FR models, the original and the degraded video sequences are directly compared, so full access to both, source and degraded video, is needed. In the RR models, some reduced information about the original video is needed, and it is used along with the degraded video in order to estimate the perceived video quality. NR models are based only in the degraded video in order to make an estimation of the perceived video quality.

Data metrics, or packet-layer models, are based on network information (i.e. IP packets). These metrics can be classified into packet-header models, bit-stream-layer model and hybrid models. The packet-header models use only general information about the network (i.e. packet loss rates), and does not take into account packet contents. Bit-stream-layer models can access IP packets payload, and extract some media related information. Hybrid models use a combination of the other methods.

Parametric models predict the perceived video quality metrics based on some reduced set of parameters, related to the encoding process, video content and/or network information. These models typically present a mathematical formula, representing the estimation of the perceived video quality as a function of different parameters. Parametric models can be applied to packet-layer models, media-layer models or a combination of both.

ITU-T Recommendation G.1070 "Opinion model for video-telephony applications" (ITU-T, 2007) describes a computational model for videophone applications over IP networks that is useful as a QoE planning tool. This model takes into account the bit rate, frame rate and packet loss rate for video quality estimation, but does not take into account video content. Video perceived quality can have great variations for different video contents for the same video codec, bit rate and frame rate.

In previous works (JOSKOWICZ et al., 2009a), we showed that the video content must be taken into account in order to provide an appropriate estimation of the perceived video quality. In this paper, we propose how to include the video content in the G.1070 model, taking into account the contributions of our previous works.

The new parameters and relation proposed are calculated for MPEG-2 (ISO/IEC 2000) and H.264/AVC (ITU-T, 2005) in bit ranges from 50 kb/s to 12 Mb/s, and in display formats SD (Standard Definition, 720 × 576 pixels), VGA (Video Graphics Array, 640 x 480 pixels), CIF (Common Intermediate Format, 352 × 288 pixels) and QCIF (Quarter Common Intermediate Format, 176 × 144 pixels)

The RMSE (Root Mean Square Error) and Pearson correlation for the original and the proposed model are calculated and compared.

The paper is organized as follows: Section 2 briefly describes the video quality function proposed in G.1070. Section 3 describes previous enhancements proposed for the G.1070 model. Section 4 shows the effects of video content in the perceived quality. Section 5 describes how to include the effect of video content in the G.1070 model and a comparison is made between the original model and the proposed enhancements. Section 6 summarizes the main contributions.

2 Video Quality Function in ITU-T Recommendation G.1070

The ITU-T Recommendation G.1070 describes a parametric model for point-to-point interactive videophone applications over IP networks that is useful as a QoE (Quality of

Experience) and QoS (Quality of Service) planning tool for assessing the combined effects of variations in several video and speech parameters that affect the perceived quality. The model takes into account the speech and the video perceived quality (YAMAGISHI et al., 2006), and combines both in an integration function for overall multimedia quality (HAYASHI et al., 2007). Speech quality estimation is based on the ITU-T Recommendation G.107 (ITU-T, 2005), known as the E-Model. Video quality estimation V_q is calculated as shown in Equation (1).

$$V_q = 1 + I_c e^{-\frac{P_{plv}}{D_{Pplv}}} \quad (1)$$

where I_c represents the basic video quality (as defined in Equation (2)), determined by the codec distortion and is a function of the bit rate and frame rate, P_{plv} is the packet loss rate and D_{Pplv} expresses the degree of video quality robustness due to packet loss, and can also depend on bit rate and frame rate.

The basic video quality I_c for each bit rate b and frame rate f (the video quality due to encoding degradation only, without packet loss) is expressed in Equations (2) and (3).

$$I_c = v_3 \left(1 - \frac{1}{1 + \left(\frac{b}{v_4} \right)^{v_5}} \right) I_f \quad (2)$$

$$I_f = e^{-\frac{(\ln(f) - \ln(v_1 + v_2 b))^2}{2(v_6 + v_7 b)^2}} \quad (3)$$

where b is the bit rate, f is the frame rate, I_f is the degradation factor due to the effect of the frame rate and $v_1..v_6$ are six coefficients of the model.

Combining Equation (1) and (2), the video quality V_q without packet loss is expressed in Equation (4).

$$V_q = 1 + v_3 \left(1 - \frac{1}{1 + \left(\frac{b}{v_4} \right)^{v_5}} \right) I_f \quad (4)$$

According to the ITU-T Recommendation G.1070, coefficients $v_1..v_6$ are dependent on codec type, video display format, key frame interval, and video display size, and must be calculated using subjective video quality tests. Provisional values are provided only for MPEG-4 in QVGA (Quarter VGA, 320 × 240 pixels) and QQVGA (Quarter QVGA, 160 × 120 pixels) video formats. Belmudez et al (2010) proposed a new set of parameters for the MPEG-2 codec.

Yamagishi et al. (2008) presented a similar model to equations (2) and (3) applicable for IPTV services in HD (High Definition, 1440 x 1080 pixels), and coefficient values are provided for the H.264 codec.

In the Recommendation G.1070, the model coefficients are not dependant on video content. This means that, for a given codec, video display format, key frame interval, and video display size, according to the Recommendation G.1070, there is only one set of coefficient for all video contents.

3 Enhancements to ITU-T Recommendation G.1070

In a previous work (JOSKOWICZ et al., 2009b), we studied the performance of the G.1070 model for clips coded at 25 fps, and we proposed some enhancements to the G.1070 model:

Porto Alegre, v.14, n.2, jul./dez. 2011.
ISSN impresso 1516-084X ISSN digital 1982-1654

a) The v_3 and I_f coefficients can be set to a fixed value (v_3=4, I_f=1) and performance is not affected, leaving only two coefficients from the original model (v_4 and v_5).

b) One new parameter is added (*a*), that takes into account the display size. With this new parameter, the model can be applied with the same set of coefficients to different display formats. The product *a.b* was defined as the "scaled bit rate".

c) We showed that the two remaining coefficients (v_4 and v_5) are highly correlated to subjective video movement content, and we defined three sets of these two parameters, one for "low movement content" applications, one for "medium movement content" and one for "high movement content" applications.

The Enhanced G.1070 model presented in the referred paper is expressed in Equation (5)

$$V_q = 1 + 4\left(1 - \frac{1}{1 + \left(\frac{ab}{v_4}\right)^{v_5}}\right) \qquad (5)$$

where V_q represents the video quality determined by the codec distortion, *b* is the bit rate (in Mb/s), *a* is a constant parameter that depends on the display format and v_4 and v_5 are other model coefficients with a strong dependence on video content. Video clips are classified in three categories, according to the subjective movement content, and the model coefficients v_4 and v_5 are calculated for each category (High, Medium and Low movement content).

The best values for each coefficient were calculated in our previous work (JOSOWICZ et al., 2009), and are presented here: The parameter *a* depends on the display format, according to Table 1. The coefficients v_4 and v_5 depends on the "subjective movement content", and are presented in Table 2. In that paper the effects of the frame rate were not analyzed (i.e., frame rate was set to 25 fps for all the clips) and the movement content was derived only based on qualitative analysis, no quantitative analysis was proposed to the video classification into each category.

Table 1 – Best values for *a*

Display Format	*a*
SD	1
VGA	1.4
CIF	3.2
QCIF	10.8

SOURCE: Joskowicz et al. 2009a

Table 2 – v_4 and v_5 values for each movement content

Movement	v_4	v_5
Low Movement	0.366	1.32
Medium Movement	0.67	1.36
High Movement	1.088	1.56

SOURCE: Prepared by the authors

4 Effects of Video Content in Perceived Quality

In this section we used the NTIA (National Telecommunications and Information Administration) "General Full Reference Model", available on line (NTIA), and standardized in ITU-T Recommendation J.144 (ITU-T, 2004) for per-

ceived MOS (Mean Opinion Score) estimation. This model performs a quality comparison between the two video clips, one defined as the original video and the other as the degraded video. The model can classified as a FR (Full Reference) model. For each video clips pair (original and degraded), the NTIA algorithm provides a VQM (Video Quality Metric), with values between 0 and 1 (0 when there are no perceived differences and 1 for maximum degradation) that can be directly associated with the DMOS (Differential Mean Opinion Score). The DMOS values returned from the NTIA model can be related to the MOS using Equation (6). The interpretation of the MOS values is presented in Table 3.

$$MOS = 5 - 4DMOS \quad (6)$$

Table 3 – MOS to perceived quality relation

MOS	Quality
1	Bad
2	Poor
3	Fair
4	Good
5	Excellent

SOURCE: Prepared by the authors

The dispersion between the ITU-T J.144 NTIA model with respect to subjective tests was presented by Pinson et al. (2004) and is graphically showed in Figure 1 (extracted from the work performed by Pinson et al.). The model performance was independently tested in the VQEG (Video Quality Experts Group) FR-TV project (VQEG, 2003) and outperforms the classic PSNR (Peak Signal to Noise Ratio) video metric.

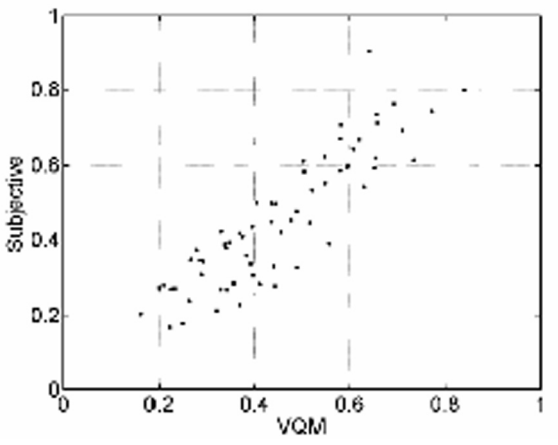

Figure 1 – ITU-T J.144 NTIA model performance (compared to subjective tests)
SOURCE: Pinson et al. 2004

The error obtained using the ITU-T J.144 NTIA model with respect to subjective tests can be estimated in 10% (+/- 0.1 in the 0-1 scale). This means that the order of magnitude of the standardized algorithm error is 0.1 in a DMOS scale from 0 to 1 (PINSON;WOLF, 2004). MOS errors, using this model, can be estimated in +/-0.4 in the 1-5 scale (4 times the DMOS error).

Table 4 – Source video clips used

Source	Name	Source	Name
src 2	Barcelona	src 14	New York
src 3	Harp	src 16	Betes pas bêtes
src 4	Moving graphic	src 17	Le point
src 5	Canoa Valsesia	src 18	Autums leaves
src 7	Fries	src 19	Football
src 9	Rugby	src 20	Sailboat
Src 10	Mobile & Calendar	src 21	Susie
Src 13	Baloon-pops	src 22	Tempete

SOURCE: Prepared by the authors

Porto Alegre, v.14, n.2, jul./dez. 2011.
ISSN impresso 1516-084X ISSN digital 1982-1654

Figure 2 shows the relation between MOS and bit rate, for the sixteen clips detailed in table 4, coded in MPEG-2 (using the coding parameters detailed in Table 5), in SD display format. Similar curves are obtained for other codecs (i.e. H.264/AVC) and display formats (i.e. VGA, CIF, QCIF). The clips were coded at different bit rates, obtaining from one original clip many "degraded" clips (each one at a different bit rate).

The MOS values were derived from DMOS, using Equation (6). The DMOS values were calculated using the ITU-T J.144 NTIA Model, comparing the original video clip with the corresponding degraded video clip. The video clips used are available at the VQEG web page (VQEG). All the clips have durations between 8 and 10 seconds, and spans over a wide range of contents, including sports, landscapes, "head and shoulders" and so.

Table 5 – MPEG-2 and H.264 coding parameters used

MPEG-2	H.264
Profile/Level: MP@ML Max GOP size: 15 GOP Structure: Automatic Picture Structure: Always Frame Intra DC Precision: 9 Bit rate type: Constant Bit Rate Interlacing: Non-Interlaced Frame Rate: 25 fps	Profile/Level: High/3.2 Max GOP size: 33 Number of B Pict between I and P: 2 Entropy Coding: CABAC Subpixel mode: Quarter Pixel Bit rate type: Constant Bit Rate Interlacing: Non-Interlaced Frame Rate: 25 fps

SOURCE: Prepared by the authors

As can be seen in Figure 2, all the clips have better perceived quality for higher bit rates, as expected. In MPEG-2, in SD, for bit rates higher than 6 Mb/s all the clips have an almost "perfect" perceived quality (MOS higher than 4.5). At 3 Mb/s all the clips are in the range between "Good" and "Excellent". However for less than 3 Mb/s the perceived quality strongly depends upon the clip content. For example at 2 Mb/s, MOS varies between 3.6 and 4.8, and at 0.9 Mb/s MOS varies between 1.9 (between "Bad" and "Poor") and 4.2 (between "Good" and "Excellent"). Similar results are obtained for different display formats and codecs. At low bit rates, the perceived video quality strongly depends upon the video content.

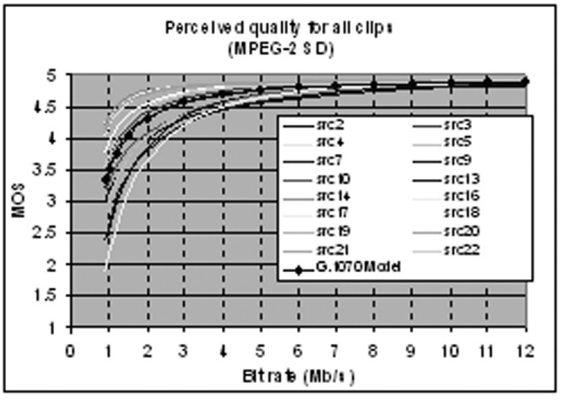

Figure 2 – Perceived Quality as a function of the Bit Rate and video content
SOURCE: Prepared by the authors

ITU-T Recommendation G.1070 does not take into account the video content. The best that the model can estimate is the average video quality for different contents, at each bit rate, as shown in Figure 2 in a bold black line. However, when we analyze the graph in Figure 2, we can see that video quality strongly depends on video content, especially for low bit rates. In our previous works (JOSKOWICZ et al., 2009) we proposed to classify the video clips into three categories, according to the subjective movement content, and different set of values for the model were calculated for each category. In the following sections, we will describe how to derive the Enhanced G.1070 model coefficients presented in Equation (5) from objective video quality spatial-temporal estimations.

5 Including Video Content in the Enhanced G.1070 Model

Each curve in Figure 2 can be modeled with Equation (5), and the best values for v_4 and v_5 can be obtained for each clip. Table 6 shows the values of v_4 and v_5 that best fits Equation (5) to each curve in Figure 2, as well as the RMSE (Root Mean Square Error) and the subjective movement content ("Mov" column, classified into "Low", "Medium" and "High"). The RMSE is related to the "distance" between the estimated and the actual values. Low values of RMSE means low "distance" and therefore better estimations.

Table 6 – v4 and v5 values that best fits to the actual NTIA curves and Avg SAD/pixel

Src	V_4	v_5	RMSE	Avg SAD/ pixel	Mov
4	0.252	1.2	0.17	0.684	Low
21	0.29	1.2	0.22	1.251	Low
20	0.29	1.24	0.18	1.303	Low
14	0.252	1.2	0.21	1.386	Low
18	0.442	1.28	0.20	1.599	Low
16	0.328	1.28	0.23	1.804	Low
22	0.594	1.48	0.19	3.315	Med
3	0.594	1.4	0.18	3.457	Med
10	0.784	1.32	0.13	3.600	Med
7	0.708	1.44	0.25	3.632	Med
2	0.86	1.24	0.15	4.243	High
19	1.05	1.68	0.17	4.520	High
5	1.012	1.6	0.22	5.148	High
13	1.24	1.84	0.19	5.656	High
9	1.24	1.6	0.25	6.164	High
17	1.506	2.04	0.26	8.256	High

SOURCE: Prepared by the authors

Subjective movement content is related to the video spatial-temporal activity. Different estimations for the video spatial-temporal activity were evaluated (including average, maximum and minimum values for moving vectors, MSE between frames, mean absolute deviations), and a strong correlation between the v_4 and v_5 coefficients with the average SAD (Sum of Absolute Differences) of the original clip has been found. SAD is a simple video metric used for block comparison and for moving vectors calculations, and can be efficiently calculated in real time (VANNE et al., 2006) (OLIVARES et al., 2004). Each frame is divided into small blocks (i.e. 8x8 pixels) and for every block in one frame the most similar (minimum SAD) block in next frame is found. This minimum sum of absolutes differences is assigned as the SAD for each block in each frame (up to the n-1 frame). Then all the SAD values are averaged for each frame and for all the frames in the clip, and divided by the block area, for normalization. We call this value "average SAD/pixel", providing an overall estimation about the spatial-temporal activity of the video clip in time windows of 8 – 10 seconds.

The relations between the v_4 and v_5 parameters for MPEG-2 encoding with the average SAD per pixel are presented in Table 6. The table also shows the RMSE obtained with these values for each clip using Equation (5), for SD, VGA, CIF and QCIF display format, and bit rates from 50 kb/s to 12 Mb/s. The table is ordered by v_4 and the subjective movement content is also presented. A similar order is obtained when using H.264/AVC. The previous categories, based on subjective movement content, can be mapped to the SAD values, according to Table 7.

Porto Alegre, v.14, n.2, jul./dez. 2011.
ISSN impresso 1516-084X ISSN digital 1982-1654

Table 7 – Movement content vs SAD

Movement	Avg SAD/pixel
Low Movement	$s < 2$
Medium Movement	$2 \leq s < 4$
High Movement	$4 \leq s$

SOURCE: Prepared by the authors

Figure 3 shows the relation between v_4 and the average SAD per pixel, for MPEG-2 and H.264/AVC codecs. Similarly, Figure 4 shows the relation between v_5 and the average SAD per pixel, for MPEG-2 and H.264/AVC codecs. In these figures, the subjective movement content is graphically showed with different colors, confirming that low values for SAD/pixel are related to low subjective movement content and high values are related to high movement content.

An estimation of v_4 and v_5 for MPEG-2 and H.264/AVC, as a function of the average SAD/pixel can be performed as presented in Equation (7).

$$v_4 = c_1 s^{c_2} + c_3$$
$$v_5 = c_4 s^{c_5} + c_6 \quad (7)$$

where s is the original video average SAD per pixel. The best values for c_1.. c_6 were calculated (using least squares method) and are presented in Table 8.

Using Equation (7), the video quality estimation presented in Equation (5) only depends on the encoded bit rate and the spatial-temporal activity of the original video clip, measured as the average SAD/pixel.

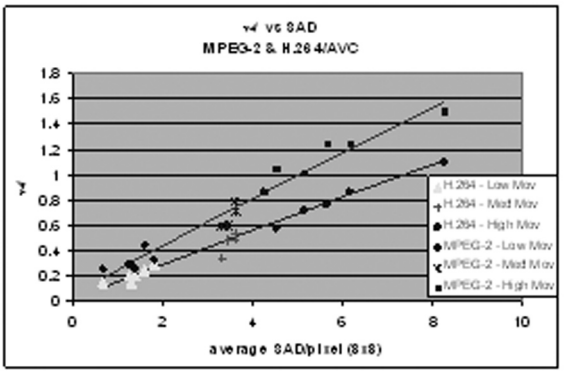

Figure 3 – Relation between v4 with respect to SAD
SOURCE: Prepared by the authors

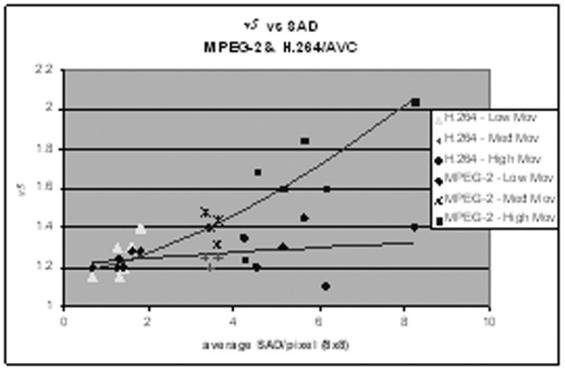

Figure 4 – Relation between v5 with respect to SAD
SOURCE: Prepared by the authors

Table 8 – ci coefficients for MPEG-2 and H.264

Coefficients	MPEG-2	H.264/AVC
c_1	0.208	0.150
c_2	0.95	0.95
c_3	0.036	0
c_4	0.036	0.030
c_5	1.52	0.68
c_6	1.17	1.20

SOURCE: Prepared by the authors

The dispersion between the MOS values derived using Equations (5) and (7) and the perceived MOS values (using the ITU-T J.144 NTIA VQM), for the sixteen video clips used, coded in MPEG-2 and H.264/AVC (using the coding pa-

rameters detailed in Table 5), in SD, VGA, CIF and QCIF display format, with bit rates from 25 kb/s to 12 Mb/s are plotted in Figure 5. In this figure, each point represents a video clip coded in a specific combination of codec, bit rate and display format. It is worth noting that subjective rating scales have ranges of 1 unit, in the 1-5 MOS scale. On the other hand, the ITU-T J.144 NTIA algorithm has errors in the order of +/- 0.4 regarding to MOS measures of subjective quality. In Figure 5, the dotted lines represent the estimated 10% (+/- 0.4) error margin of the NTIA model. Only 27 from the 2064 (about 1%) points are outside the dotted lines, meaning that the predicted MOS values have the same degree of precision than the VQM standard that has been used. The Pearson correlation between the values derived using Equations (5) and (7) and the perceived MOS values (using the NTIA VQM) is 0.985. The Pearson correlation metric evaluates the precision of the prediction. It varies from 0 to 1, where 1 indicates a direct relationship and 0 indicates no relationship at all. In this case, 0.985 indicates a very high correlation between the values derived using Equations (5) and (7) and the perceived MOS values.

The RMSE between the values derived using Equations (5) and (7) and the perceived MOS values is 0.12, meaning a very low "distance" between the model and the standard VQM.

Figure 6 represents the same data and shows the dispersion using the original G.1070 model, where only one value of v_4 and v_5 is used for all the clips of each codec. In this case, 411 from the 2064 (about 20%) points are outside the dotted lines. The Pearson correlation between the values derived from the original G.1070 model and the perceived MOS values (using the NTIA VQM) is 0.863. The RMSE is 0.36.

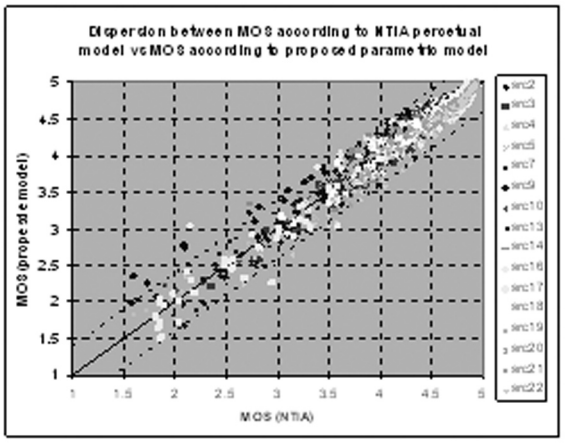

Figure 5 – MOS dispersion in proposed model
SOURCE: Prepared by the authors

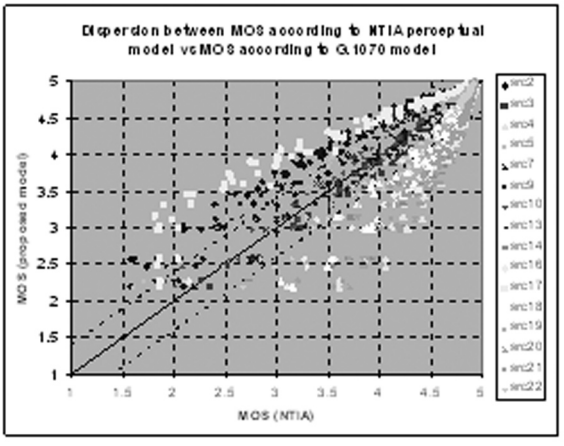

Figure 6 – MOS dispersion in G.1070 original model
SOURCE: Prepared by the authors

In many cases, at the receiver side, there is no access to the original signal, so the Average SAD/pixel of the original video can not be calculated. In these cases, at the receiver side, it will be no possible to apply the model in order to make the video quality estimation. But the Average SAD/pixel can also be calculated at the receiver, based on the degraded clip. Typically, in the decoder (degraded clip), the average SAD/pixel has lower values than the original clip, due to the loss of definition introduced in the coding

Porto Alegre, v.14, n.2, jul./dez. 2011.
ISSN impresso 1516-084X ISSN digital 1982-1654

process. Figure 7 shows how the average SAD/pixel of the same clip varies according to the coded bit rate, for the clips coded in H.264/AVC in VGA display format.

All the curves in Figure 7 can be modeled according to Equation (8)

$$s = s_0 \left(1 - \frac{1}{1 + \left(\frac{b}{s_1} \right)^{s_2}} \right) \quad (8)$$

where s is the average SAD/pixel for each clip coded at a given bit rate b, s_o is the average SAD/pixel of the original clip and s_1 and s_2 are two constants. Measuring the average SAD/pixel $\boldsymbol{s}$ at the receiver side, the original average SAD/pixel (s_o) can be derived using Equation (8), knowing only the coded bit rate and the average SAD/pixel (s) measured at the decoder. Using this estimation, there is no need to access the original signal in order to apply the model. All the model parameters can be evaluated at the decoder: the used codec, the bit rate, the display size, and the spatial and temporal video activity, derived from the average SAD/pixel measured in the decoder.

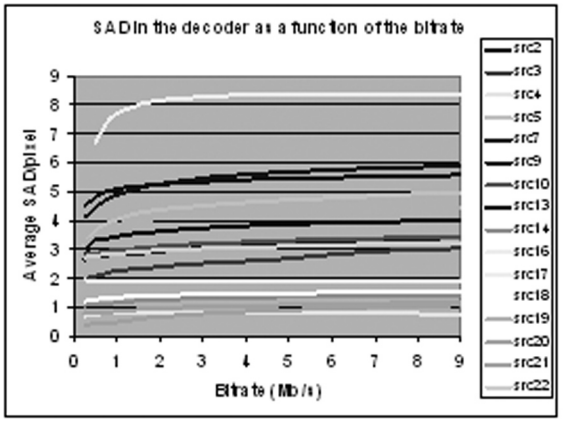

Figure 7 – Average SAD/pixel at the decoder side, as a function of the bit rate
SOURCE: Prepared by the authors

The best values for $\boldsymbol{s_1}$ and $\boldsymbol{s_2}$ were calculated for H.264/AVC, and are presented in Table 9.

Table 9 – si coefficients for H.264

Coefficients	Value
s_1	0.0757
s_2	0.667

SOURCE: Prepared by the authors

The dispersion between the MOS values derived with the proposed model (measuring the average SAD/pixel at the decoder, for clips coded in H.264/AVC in VGA display format) and the perceived MOS values according to the ITU-T J.144 NTIA model are plotted in Figure 8. The Pearson correlation between the values derived using Equations (5), (7) and (8) (measuring the average SAD/pixel at the decoder) and the perceived MOS values is 0.975, and the RMSE is 0.19. Only 7% of the points are outside the +/- 10% error margin of the ITU-T J.144 NTIA model.

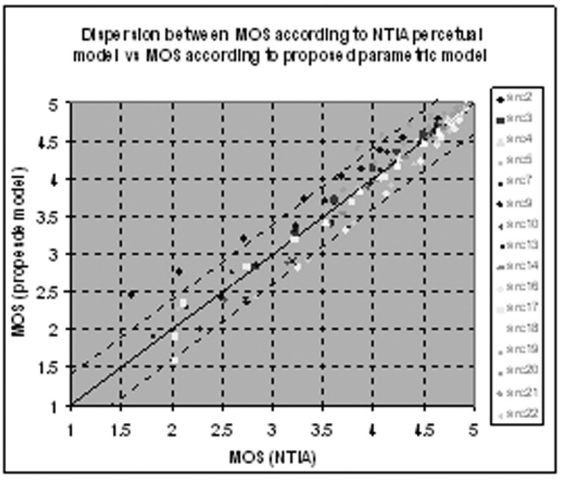

Figure 8 – MOS dispersion between the proposed model (with access only to the degraded video) with respect to ITU-T J.144 NTIA model, for VGA in H.264
SOURCE: Prepared by the authors

In Table 10 a summary of the models performance is presented. The proposed model has higher Pearson correlation, lower RMSE and lower percentage of points outside the +/-10% than the original G.1070 model, in both cases, measuring the average SAD/pixel at the original clip (encoder) or at the degraded clip (decoder).

Table 10 – Models performance comparison
(PC=Pearson correlation, RMSE = Root Mean Square Error, Outliers = points outside +/- 10%)

Model	PC	RMSE	Outliers
Proposed model with access to original SAD	0.985	0.12	1%
Proposed model with access to received SAD	0.975	0.19	7%
ITU-T G.1070	0.863	0.36	20%

SOURCE: Prepared by the authors

7 Conclusion

The video quality estimation proposed in ITU-T Recommendation G.1070 "Opinion model for video-telephony applications" does not take into account video content. We have shown that, specially in the low bit rate range, the video quality have high variations depending on video content for same bit rate and other codec parameters. In this work, an enhancement has been proposed to the model, including the characteristics of video content based on the average SAD/pixel of the clip as a spatial-temporal estimation. It was found a strong relation between two model parameters (v_4 and v_5) and the average SAD/pixel of the original video, and a mathematical equation was proposed for modeling this relation.

The model was evaluated measuring the average SAD/pixel at the original video and at the decoded (degraded) video. A relation has been proposed in order to derive the original average SAD/pixel as a function of the bit rate and the average SAD/pixel measured at the receiver side.

The proposed model enhancement has been evaluated, using sixteen different video sources, spanning over a wide range of contents, including sports, landscapes, animated pictures, "head and shoulders", and so on. These video clips were coded in MPEG-2 and H.264/AVC, in SD, VGA, CIF and QCIF display, and with bit rates from 50 kb/s to 12 Mb/s. In total 96 different formats were used, and more than 1500 processed video sequences were analyzed.

The Pearson correlation of the original ITU-T G.1070 model is 0.863 for all these processed video clips and the RMSE is 0.36. Measuring the average SAD/pixel at the original signal (i.e. in the encoder), the Pearson correlation of the proposed enhancement is 0.985 and the RMSE is 0.12. Measuring the average SAD/pixel at the degraded signal (i.e. in the decoder), the Pearson correlation of the proposed enhancement is 0.975 and the RMSE is 0.19, slightly worst than in the first case, but still much better then the original G.1070 model.

The result shows that the proposed enhancements perform much better than the original model and fits very well with respect to the perceptual video quality estimations derived from the standardized ITU-T J.144 VQM.

The proposed model can be used as a basis for the developing of new Reduced References or No References parametric models, and for on-line video quality estimations algorithms.

Porto Alegre, v.14, n.2, jul./dez. 2011.
ISSN impresso 1516-084X ISSN digital 1982-1654

References

BELMUDEZ, B.; MÖLLER, S: Extension of the G.1070 video quality function for the MPEG2 video codec, 2010 Second International Workshop on Quality of Multimedia Experience (QoMEX), 2010

HAYASHI Takanori; YAMAGISHI, Kazuhisa; TOMINAGA, Toshiko; TAKAHASHI, Akira: Multimedia Quality Integration Function for Videophone Services, IEEE Global Telecommunications Conference, 26-30 Nov. 2007
ISO/IEC 13818-2:2000. Information technology – generic coding of moving pictures and associated audio information: Video

ITU-T Recommendation J.144 Objective perceptual video quality measurement techniques for digital cable television in the presence of a full reference, February 2004

ITU-T Recommendation G.107: The E-model, a computational model for use in transmission planning, March 2005

ITU-T Recommendation G.1070 Opinion model for video-telephony applications, April 2007

ITU-T Recommendation H.264 Advanced Video Coding for Generic Audiovisual Services, March 2005

JOSKOWICZ, Jose; LOPEZ ARDAO, José-Carlos: Enhancements to the Opinion Model for Video-Telephony Applications, LANC'09, September 24–25, 2009, Pelotas, Brazil , pp. 87-94

JOSKOWICZ, Jose; LOPEZ ARDAO, José-Carlos; GONZALEZ ORTEGA, Miguel A.; GARCIA, Cándido López: A Mathematical Model for Evaluating the Perceptual Quality of Video, FMN 2009, LNCS 5630, pp. 164–175, 2009

NTIA - Video Quality Metric (VQM) Software [Online]. Available at: www.its.bldrdoc.gov/n3/video/vqmsoftware.htm

OLIVARES, Joaquín; HORMIGO, Javier; VILLALBA, Julio; BENAVIDES, Ignacio: Minimum Sum of Absolute Differences Implementation in a Single FPGA Device, Lecture Notes in Computer Science (LNCS), Volume 3203/2004, pp 986-990

PINSON, Margaret H; WOLF, Stephen: A New Standardized Method for Objectively Measuring Video Quality, IEEE Transactions on Broadcasting, Volume 50, Issue 3, September 2004, pp. 312-322

VANNE, J.;AHO, E.;HAMALAINEN, T.D.;KUUSILINNA, K.: A High-Performance Sum of Absolute Difference Implementation for Motion Estimation, IEEE Transactions on Circuits and Systems for Video Technology, Volume 16, Issue 7, July 2006 Page(s):876 – 883

VQEG Phase I Test Sequences. [Online]. Available at: ftp://vqeg.its.bldrdoc.gov/SDTV/VQEG_PhaseI/TestSequences/Reference/

VQEG Final Reprt from the Video Quality Experts Group on the Validation of Objective Models of Video Quality Assessment: Phase II. 2003. Available at <http:www.its.blbrdoc.gov/vqeg/projects/frtv_phaseII/downloads/VQEGIIII_Final_Report.pdf>. Acessed: 2010, Jan. 10.

WOLF, Stephen; PINSON, Margaret H.: Low Bandwidth Reduced Reference Video Quality Monitoring System, First Int'l Workshop on Video Proc. and Quality Metrics, Jan 2005

YAMAGISHI, K.; HAYASHI, T: Parametric Packet-Layer Model for Monitoring Video Quality of IPTV Services, IEEE International Conference on Communications 2008 (ICC 08), 19 May 2008

YAMAGISHI, Kazuhisa; HAYASHI Takanori: Opinion Model for Estimating Video Quality of Videophone Services, IEEE Global Telecommunications Conference, Nov. 27 2006

Recebido em maio de 2011
Aprovado para publicação em julho de 2011

José Joskowicz
Instituto de Ingeniería Eléctrica – Facultad de Ingeniería – Universidad de la República – Montevideo, Uruguay. E-mail: josej@fing.edu.uy

J. Carlos Lopez Ardao
ETSE Telecomunicacion –Universidad de Vigo – Vigo, España. E-mail: jardao@det.uvigo.es

Rafael Sotelo
Facultad de Ingeniería – Universidad de Montevideo – Montevideo, Uruguay. E-mail: rsotelo@um.edu.uy