Efficient Quality of Multimedia Experience Using Perceptual Quality Metrics

Rahul Gaurav and Hasan F. Ates

Abstract Although there are very few metrics which correlate well with the Human Visual System (HVS), most of them do not. We extensively studied about the Video Quality Assessment (VQA) of 2D svideos for enhancing the Quality of Experience (QoE) using better Quality of Service (QoS). We propose a solution which helps us to find a high correlation between HVS and the objective metrics using Perceptual Quality Metrics (PQM). The motive behind this work is to introduce an objective metric that is adequate to predict the Mean Opinion Score (MOS) of distorted video sequences based on the Full Reference (FR) method.

Keywords Quality assessment · Subjective testing · Objective testing · Structural similarity · QoE · PQM

1 Introduction

Up until now, the most precise and valued way of assessment of the quality of a video is the evaluation using subjects in the form of human participants [1]. As involving human subjects in such applications is laborious hence this leads to a need of a highly robust system which is able to assess the quality effectively without introducing any human observers. Few things can easily be deduced from literature reviews that the focus has been on the Quality of Service (QoS) rather than the Quality of Experience (QoE). The former term tries to objectively quantify the services handed over by the vendor and has nothing to do with the view point of the audience but it is more relevant to the media. While the latter speaks about the

R. Gaurav (✉)
Telecom SudParis, ARTEMIS Department, Evry, France
e-mail: rahulgaurav.kist@gmail.com

H.F. Ates
Department of Electrical-Electronics Engineering, ISIK University,
Sile, Istanbul, Turkey

A. Nagar et al. (eds.), *Proceedings of 3rd International Conference on Advanced Computing, Networking and Informatics*, Smart Innovation, Systems and Technologies 43, DOI 10.1007/978-81-322-2538-6_50

subjective measure of a person's experience. So in order to gauge the performance of the quality assessment, Mean Opinion Score (MOS) comes into play, which is the subjective quality measurement carefully done by using human subjects as observers and helps us to correlate with the obtained objective scores. Clearly, there is a need of a versatile QoS model which complies with the QoE in the best possible way. Our paper proposes one solution to this issue. We worked on such objective metrics which performs better than the state-of-art models and mimics the HVS. To a great extent, our work is inspired by the Perceptual Quality Metric (PQM) for dealing with 3D video datasets [2]. A robust objective algorithm has been proposed namely Perceptual Quality Metric for 2D (PQM2D) using the ideas from the above mentioned work. The aim of this work is to show better results for 2D VQA and outperform the various popular state-of-art metrics like Peak Signal to Noise Ratio (PSNR), Structural Similarity (SSIM) Index and Multi Scale SSIM (MS-SSIM) Index. For the verification phase, series of subjective experiments were performed to demonstrate the level of correlation between objective metrics and the user scores obtained by Subjective Evaluation using human observers, keeping in mind the standards set by the International Telecommunication Union (ITU) [1].

2 Quality of Experience Experimentation

2.1 Introduction

The QoE methods are essentially used to gauge the performance of multimedia or television systems with the help of responses obtained from observers who view the system under test [3]. With the help of this experiment, we will be able to find the MOS of the various video sequences under consideration[4–6].

2.2 General Viewing Conditions

The Table 1 gives us a short overview of the laboratory conditions and some of the details about the display of our system.

2.3 Source Sequences

The videos were obtained from the Laboratory for Image and Video Engineering (LIVE) at The University of Texas at Austin [7]. In our experiments, we used nine reference videos in the test session. Figure 1 shows the histogram of PSNR

Table 1 Laboratory conditions

Parameters	Settings
Peak luminance of the screen	150 cd/m^2
Other room illumination	Low
Height of image on screen (H)	11 cm
Viewing distance	88 cm

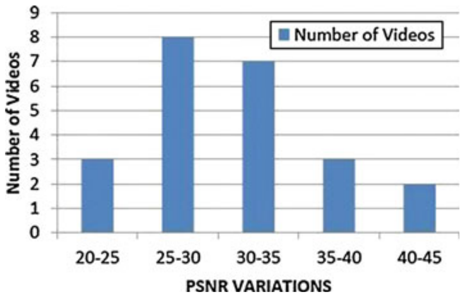

Fig. 1 PSNR range of video sequences

variations for the selected set of source sequences. Test cases were carefully selected so that the maximum range of PSNR is covered to get more reliable results.

2.4 Test Sequences

For the test sequence cases, we used four types of distortions namely wireless distortion, IP network distortion, H.264 compression and MPEG-2 compression.

2.5 Subjective Testing Design

The test methodology used is known as Double Stimulus Impairment Scale (DSIS) [1]. A carefully selected playlist was prepared by the authors, comprising of 24 videos in total, 9 reference samples in total with various kinds of distorted counterparts.

2.6 Observer Selection and Training

Most of the subjects who took part in our research were non-expert undergraduate students from the department of psychology of the ISIK University, Turkey. Each video was rated by 16 subjects in total with the help of a program formulated by the Graphics and Media Lab Video Group in Russia [8].

3 Quality of Service Experimentation

3.1 Introduction

We carried out the FR based objective VQA simulation using MATLAB codes written by the authors for our selected set of videos. The objective algorithms used in our research are the popular state-of-art metrics like PSNR, SSIM and MS-SSIM and proposed metrics PQM2D.

3.2 Peak Signal to Noise Ratio

PSNR is a simple function of the Mean Squared Error (MSE) between the reference and distorted videos and provides a baseline for objective algorithm performance [9].

$$PSNR = 10\log_{10}\frac{255^2}{MSE} \tag{1}$$

3.3 SSIM

We applied the SSIM index frame-by-frame on the luminance component of the video and computed the overall SSIM index for the video as the average of the frame level quality scores. We used two kinds of algorithms for SSIMs namely SSIM-Gaussian (SSIMG) and SSIM Block (SSIMB). The former is the standard SSIM using conventional Gaussian way and in the latter, SSIM is computed on an 8×8 block basis, and the average SSIM for the whole frame is the average of block SSIMs [10].

$$SSIM(x,y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2\mu_y^2 + C_1)(\sigma_x^2\sigma_y^2 + C_2)} \tag{2}$$

3.4 MS-SSIM

The fact which distinguishes MSSIM from SSIM is that this VQA algorithm evaluates multiple SSIM values at multiple resolutions. Although it does not lay stress on the luminance component in general, nonetheless we implemented it frame by frame to the luminance part and finally average value was computed. In defining MS-SSIM, luminance, contrast and structure comparison measures are computed at each scale as follows [11]:

$$l(x,y) = \frac{(2\mu_x\mu_y + C_1)}{(\mu_x^2 + \mu_y^2 C_1)} \quad (3)$$

$$c(x,y) = \frac{\sigma_x\sigma_y + C_2}{(\sigma_x^2 + \sigma_y^2 + C_2)} \quad (4)$$

$$s(x,y) = \frac{\sigma_{xy} + C_3}{(\sigma_x + \sigma_y + C_3)} \quad (5)$$

where C_1, C_2 and C_3 and are small constants given by the following relation are small constants given by the following relation

$$C_1 = (K_1 L)^2, \quad (6)$$

$$C_2 = (K_2 L)^2 \quad (7)$$

and

$$C_3 = \frac{C_2}{2} \quad (8)$$

Furthermore,

$$L = 255, K_1 << 1 and K_2 << 1. \quad (9)$$

Using the above equations, we compute the MS-SSIM values as follows

$$MS - SSIM(x,y) = [l_M(x,y)]^{\alpha_M} \Pi_{j=1}^{M} [c_j(x,y)]^{\beta_j} [s_j(x,y)]^{\gamma_j} \quad (10)$$

Similarly, we used two kinds of MS-SSIMs namely MS-SSIM Gaussian (MS-SSIMG) and MS-SSIM Block (MS-SSIMB) by making slight changes, that is, rather than using the Gaussian window in the former, we computed the SSIM level by level by using 8 × 8 block level at each resolution in the latter.

3.5 Proposed PQM2D Metrics

Using the ideas from [1, 12, 13] and rather than dealing with the 3D video components, our metrics assessed the quality of 2D video sequences extensively. The idea behind the formation of the new metrics is taken from the fact that the luminance value is an essential component that determines the quality of an image. On the contrary, chrominance is basically responsible for colour in the image. Thus, we can say that the luminance provides structure based information about the image rather than the colour of the various objects in the image. This method is based on the idea of finding the difference between luminance values in the test and impaired frames [13]. As it is obvious that there might be variations in the structure as and when the frames become distorted, there should be prominent deviations in the luminance values. Furthermore, these luminance deviations, when considered at a specific pixel coordinate of reference as well as the impaired frames, give us meaningful values. That means, greater the impairment in the structure of the processed frame at a certain pixel coordinate, greater is the luminance deviation from the reference frame, at that very point. The step by step algorithm implementation is given below.

1. Compute the pixel mean, variance and covariance of blocks

$$\mu(b_o), \mu(b_R), \sigma^2(b_o), \sigma^2(b_R), \sigma(b_o b_R). \tag{11}$$

2. Compute weighted distortion coefficient for each pixel in the block

$$\alpha(m,n) = \begin{cases} 0, \mu(b_o) << 1 \quad \text{and} \quad \mu(b_R) << 1 \\ 1, \mu(b_o) << 1 \quad \text{and} \quad \mu(b_R) > 1 \\ \min\left[\frac{(c_o(m,n) - c_R(m,n))^4}{\mu(b_o)^2}\right], \quad \text{else} \end{cases} \tag{12}$$

For contrast distortion in the block, define:

$$K(b_R) = 1 + \frac{\sigma^2(b_o) - (\sigma^2(b_R))^2 + 255}{(\sigma^2(b_o))^2 + (\sigma^2(b_R))^2 - 2(\sigma(b_o b_R))^2 + 255} \tag{13}$$

3. Perceptual distortion Metrics (PDM) in the whole block is defined as:

$$PDM(b_R) = \frac{K(b_R)}{64} \sum_{(m,n)\epsilon(b_R)} \alpha(m,n) \tag{14}$$

After PDM is computed for all blocks, total perceptual distortion in the frame is equal to weighted mean of block distortions:

$$PDM(c_R) = \frac{\sum_{(b_R)\epsilon(c_R)} w(b_R) PDM(b_R)}{\sum_{(b_R)\epsilon(c_R)} w(b_R)} \tag{15}$$

$$w(b_R) = \begin{cases} 1, \mu(b_o) = 0 \\ \frac{255}{\mu(b_o)}, else \end{cases} \tag{16}$$

Finally PQM2D is defined as follows:

$$PQM2D(c_R) = \begin{cases} 0, PQM2D(c_R) < 0 \\ 1 - PDM(c_R), else \end{cases} \tag{17}$$

Frame level PQM2Ds are averaged for the whole video. In order to obtain the overall objective score for a sequence, the scores of the frame level PQM are added up and divided by the number of frames in the sequence and we get the PQM2D score representing the overall quality score judging on a scale of measurement of 0 to 1 where 0 stands for the worst quality and 1 for the best. The main idea for the PQM is based on the fact that the HVS gives the quality by first measuring the errors in the luminance which in fact comprises of the structure in an image and also is quite less sensitive to the chrominance element of an image.

3.6 Simulation Results and Discussions

The various performance criterions applied on the metrics were monotonicity and accuracy, determined on the basis of Pearson Linear Correlation Coefficient (PLCC) and the Spearman Rank Order Correlation Coefficient (SROCC) respectively. The number of selected presentations for each distortion type were 7 for wireless, 5 for H264, 6 for IP and 6 for MPEG-2. In the scatter plots Fig. 2, we used statistical procedures of various regression types like exponential, linear, logarithmic, power for finding the best fitting trend lines in our data values in order to predict the accuracy of our results. The various equations of the best fitting trend lines are also shown in the plot. On the basis of square of correlation, we fitted the best trend lines and after comparing all its values, we found that the linear fit is the best for all our models. Clearly, PQM2D has the highest value and both the MS-SSIM values are lowest. Tables 2 and 3 show us the performance estimation of all the objective models with respect to the statistical measures of coefficients of the PLCC and the SROCC respectively for all selected video scenes and also individually for each of the four distortion types. It is clearly evident from the results of the metric PQM2D, with respect to PLCC and SROCC that it outperforms all the other objective models. Our tactfully organised digital video database taken from the LIVE database also testifies the drawbacks of PSNR and both MS-SSIM as it is substantially lower than most of the objective models. When we study the linear correlations based on distortion types, we see that PQM2D is mostly superior like in IP and MPEG2 distortion cases and close to the superior in case of wireless and H264 ones. Nevertheless both SSIM have shown their fairly efficient performance. For example SSIMG and SSIMB perform the best in wireless and H264 distortions respectively. However SSIMG and SSIMB perform poorly for the IP distortions, causing their overall performance to be lower than the PQM2D. Likewise MS-SSIM has shown inferior performance in most of the distortion types. Therefore it can be said that the PQM2D performs consistently well for all distortion types while other metrics fail for certain types of distortions. When we study the monotonicity of the model using the SROCC results, we still see that PQM2D has the highest overall correlation score. When distortion types are individually considered, correlation values of PQM2D are fairly close to the best one except for the wireless case where it performs sub optimally. Yet, for the full data, PQM2D

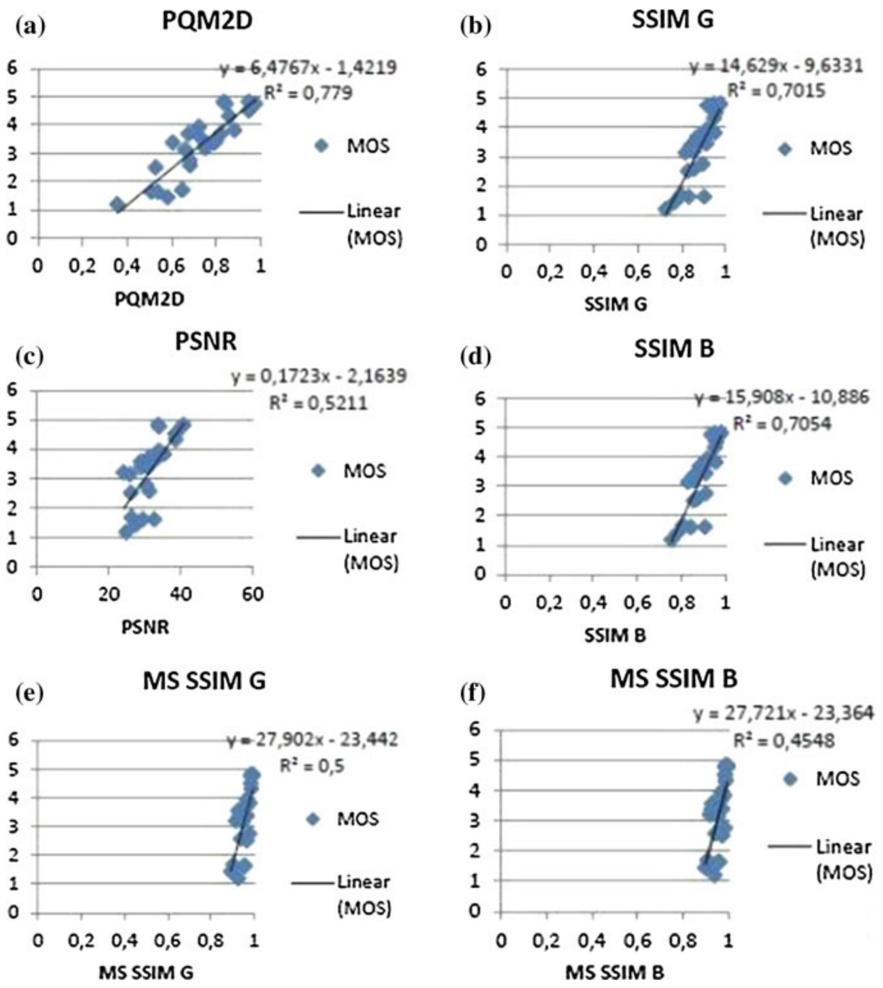

Fig. 2 Scatter plots of objective versus subjective model. **a** PQM2D. **b** SSIM G. **c** PSNR. **d** SSIM B. **e** MS SSIM G. **f** MS SSIM B

Table 2 Comparison of PLCC

Metrics	Wireless	H.264	IP	MPEG2	ALL
PQM2D	0.916	0.942	0.924	0.956	0.883
PSNR	0.537	0.915	0.713	0.918	0.722
SSIMG	0.928	0.940	0.609	0.930	0.838
SSIMB	0.904	0.950	0.622	0.920	0.840
MS-SSIMG	0.812	0.887	0.647	0.830	0.707
MS-SSIMB	0.789	0.868	0.654	0.800	0.674

Table 3 Comparison of SROCC

Metrics	Wireless	H.264	IP	MPEG2	ALL
PQM2D	0.786	0.90	0.771	0.843	0.885
PSNR	0.714	0.900	0.429	0.929	0.780
SSIMG	0.964	1.000	0.600	0.929	0.877
SSIMB	0.955	1.000	0.714	0.929	0.877
MS-SSIMG	0.893	1.000	0.647	0.829	0.746
MS-SSIMB	0.964	0.900	0.829	0.814	0.716

again has the highest SROCC scores among the tested metrics. The higher quality of performance of our metrics PQM2D is elucidated in both the correlation results as it is always slightly larger than SSIMG and SSIMB and also is fairly larger than MS-SSIMG and MS-SSIMB. Nevertheless, the SSIM results are apparently comparable to the best performing algorithm.

3.7 Conclusion and Future Work

The gist of our discussion is that the PQM2D is superior in performance and this gives us the perfect picture of our research theme that a robust objective algorithm, well-correlated with the human perceptual experience can provide us the best method to estimate the digital video quality. In other words, a well formed QoS can only be justified when the QoE has been obtained systematically. Through our objective metric discussions, the sensitiveness of HVS to the luminance component is clearly visible. Evidently, we came across several artefacts in our test videos arising due to different types of distortions in our experiment. Seemingly, it is hard to fathom that a single quality evaluation metrics can deal with all kinds of artefacts. In fact different quality metrics may be required to deal with different artefacts efficiently. The crux of the entire paper is that complexity of the HVS is still not much known and as we solve the complexity day by day, we can have more reliable and precise results for quality assessment. For future work and in order to enhance the MOS prediction models, other features of HVS can be stressed upon. Another possible enhancement could be made while dealing with the temporal features which are not employed in most of the QoS models. Presumably, incorporating both spatial as well as the temporal component into the QoE model could lead to a rather effective prediction of the QoE.

References

1. International Telecommunication Union-Radiocommunication Sector: Methodology for the subjective assessment of the quality of television pictures, Recommendation ITU-R BT, pp. 500–513 (2012)
2. Joveluro, P.: Perceptual quality estimation techniques for 2D and 3D videos. PhD Thesis, University of Surrey, UK (2011)

3. Seshadrinathan, K., Soundararajan, R., Bovik, A.C., Cormack, L.K.: A subjective study to evaluate video quality assessment algorithms. In: SPIE Proceedings Human Vision and Electronic Imaging (2010)
4. Video Quality Experts Group (VQEG): http://www.its.bldrdoc.gov/vqeg/vqeg-home.aspx
5. International Telecommunication Union-Radiocommunication Sector: Subjective video quality assessment methods for multimedia applications, Recommendation ITU-T P. 910 (1999)
6. Roque, L.: Quality evaluation of coded video. Masters Thesis, Universidade Tecnica de Lisboa, Portugal (2009)
7. LIVE, Image and Video Quality Assessment Database: http://live.ece.utexas.edu/research/quality/subjective.htm
8. Vatolin, D., Petrov, O.: MSU Perceptual Video Quality Tool 1.0 (Software for Subjective Video Quality Evaluation). Moscow State University, Russia (2006)
9. Winkler, S.: Digital Video Quality. Wiley, Chichester (2010)
10. Wang, Z., Bovik, A., Sheikh, H., Simoncelli, E.: Image quality assessment: from error visibility to structural similarity. IEEE Trans. Image Process. **13**(4), 600–612 (2004)
11. Seshadrinathan, K., et al.: Study of subjective and objective quality assessment of video, IEEE Trans. Image Process. (2010)
12. Amirshahi, S.A.: Towards a perceptual metric for video quality assessment. Masters Thesis report, Gjovik University College, Norway (2010)
13. Gaurav, R.: Video quality assessment using subjective and objective metrics. Masters Thesis, ISIK University, Turkey (2013)