Rule-Based Combination of Video Quality Metrics

Jesús Gutiérrez , Alexander Raake , Savvas Argyropoulos , and Narciso García

Abstract—Lately, several algorithms have been proposed to automatically estimate the quality of video sequences, even some have been included in international standards. However, the majority only provide high performance under particular conditions and with certain types of degradations. Therefore, some proposals have been presented setting out the combination of various quality metrics to improve the performance and the range of application. In this paper, a rule-based combination of standardized metrics is presented, in contrast to most of these type of approaches based on combinational models. The proposed system consists of a first stage in which the type of degradation affecting the video quality is identified to be caused by coding impairments or transmission errors. Then, the most appropriate metric for that distortion is applied. Specifically, VQM and VQuad have been considered for coding and transmission distortions, respectively. The results show that the overall performance is better than using the quality metrics individually.

I. INTRODUCTION

Nowadays, knowing the quality of experience (QoE) of the end users of video applications is one of the major issues for various sectors involved in the development of multimedia services. This has led to an intensive research effort in the last years, due to the importance of having that knowledge to offer video services that satisfy the user expectation, especially, taking into account the technological restrictions of each particular case of use (e.g., coding of the content, bandwidth limitations for transmission, etc.).

Although the best way to know the quality perceived by the users is by means of subjective experiments, in which the observers themselves provide their opinions about the application under study [1], several methods have been proposed to estimate automatically the perceived quality, commonly known as objective quality metrics. However, the performance of these techniques is highly dependent on the application scenario, since it conditions the available data used to estimate the quality [2]. For example, when the video to be evaluated could be compared to a reference video (e.g., evaluation of video encoders) the performance of the metrics, known as full-reference (FR), is better than in the case when there is no other data to carry out the evaluation than the test video, called no-reference (NR) metrics, as in systems for quality monitoring of delivery video services. There is also an intermediate solution based on the comparison of certain features of the original video with the corresponding obtained from the test sequence, which are called reduced-reference

(RR) metrics and can expand the range of application of the FR metrics providing better performance than NR approaches [2].

Furthermore, the quality perceived by the users of multimedia applications is influenced by several factors [3], which makes its estimation a highly complex task. Thus, to simplify the problem in a certain degree, the majority of the metrics are based on the analysis of common degradations of the video content (e.g., coding artifacts, packet losses, audio distortions, etc.). This way it is possible to increase the estimation performance of the algorithms [4].

Therefore, the idea of combining various metrics focused on different degradations arises to obtain a better estimation of the overall quality [2]. In fact, some approaches have been proposed fusing diverse metrics to obtain a single quality value, for instance, by means of linear and non-linear combinations [5], or neural networks [6]. However, these methods to merge metrics are not always as robust as desired, since the parameters and weights of the models could be dependent of particular factors of the specific case of study (e.g., the degradations or the databases used) losing some generality.

Unlike those techniques, a rule-based approach is proposed here to combine different reliable quality metrics to achieve robust evaluations of the video quality in scenarios with distinct degradations. In essence, the scheme is based on an initial step in which the type of degradation that affects the video is detected and identified, and a latter stage in which the most suitable metric for that sort of distortion is applied. Furthermore, this paper deals with the efficient use of standardized video quality metrics [7][8], since this type of metrics has been considered in the current implementation of the proposed framework.

The rest of the paper is structured as follows. In section II, a review of the existing approaches that combine different video quality metrics to increase the performance is presented. Section III explains the details of the proposed scheme. In section IV, the experimental setup used validate the approach is described, while section V shows the main results. Finally, in section VI the conclusions of the work are exposed.

II. RELATED WORK

Several of the proposed video quality metrics are developed to deal with certain degradations or to operate in certain scenarios under limited conditions. Therefore, the combination of different specific metrics is an instinctive idea that turns up

in order to obtain robust and reliable systems for video quality evaluation.

For instance, an intensive research activity is being carried out to find robust techniques to estimate the overall audiovisual quality perceived by the users, which means considering both the quality of the video and the audio content. For that purpose, various models have been proposed combining (e.g., linearly) metrics specifically developed for each type of content [9]. The same idea has been also considered in the first attempts to estimate the QoE related to 3D images, since new factors and degradations appear, for example, fusing metrics originally developed for conventional content with depth information [10]. Another approach of the potential of the combination of metrics is the development of hybrid metrics, led by the Video Quality Experts Group (VQEG). These methods are based on the use of NR metrics that analyze the bitstream of encoded video with others that operate at pixel level (i.e., over the decoded frames) [11][12].

Extending this concept, some proposals have been presented combining various metrics that were developed for specific image degradations. For example, Liu *et al.* considered ten well-known metrics for image quality to apply a multi-metric fusion using a non-linear combination obtained from a training process by regression [5]. In addition, it is also worth noting the no-reference framework proposed by Moorthy and Bovik, consisting of a first stage in which the algorithm estimates the probability of the presence of a set of distortions in the image, and the second stage assesses the quality of the image from the evaluation of each distortion weighted with their probability [13].

Although the complexity is increased in the case of video quality metrics, some proposals have been presented. For example, Leontaris *et al.* designed a simple linear combination of a metric for blocking artifacts with an algorithm to detect edge artifacts, obtaining better performance than using each metric separately [14]. Another approach was proposed by Pahalawatta and Tourapis [15] combining temporal metrics with the well-known *structural similarity index* (SSIM) metric [16]. Additionally, some NR techniques have been proposed for video quality assessment, such as the combination of blockiness, blurriness and noisiness by means of a weighted Minkowski summation developed by Farias and Mitra [17], or the combination of three NR audio-visual metrics by Martinez and Farias [18]. Also, recently, interesting approaches have been proposed using machine learning to compose a framework to fuse various image quality metrics [6].

In contrast to these approaches, the following section presents a rule-based method that allows identifying the type of impairment that degrades the video quality, and then, the most appropriate metric could be used. Thus, the idea is similar to the proposal by Moorthy and Bovik [13] in the sense that a two-step framework is used, where the first stage identifies the type of degradation, while the second step provides the quality estimation. However, it is worth noting that the approach presented in this paper carries out a hard classification of the degradations, so only the metric that best deals with that distortion operates to provide the quality assessment. In addition, the present proposal, works over video content and
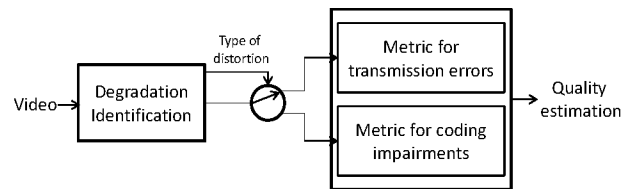

Fig. 1. Diagram of the proposed scheme

aims to estimate the quality in distribution systems. Thus, the current approach, based on reliable FR metrics, is the first step towards a monitoring scheme using NR video metrics to deal not only with coding impairments (as in the work by Farias and Mitra [17]), but also with transmission errors (e.g., jerkiness, tiling, etc.).

III. PROPOSED SCHEME

Taking advantage of robust metrics of the state of the art that provide satisfactory results dealing with certain degradations, it is possible to develop a tool, combining them, to obtain better results, especially with coding artifacts and transmission errors. In fact, the efficiency of the existing metrics depends mainly on the degradations that affect the video quality and the data that can be used to estimate it. Although, as a first approach, the work in this paper has been focused on the efficient use of standardized FR metrics, thanks to their verified high-performance, the concept could be also useful in the case of NR metrics, especially for developing efficient architectures for video quality monitoring [19].

The basis of the proposed approach is depicted in Fig. 1. Firstly, a detection stage is applied over the video sequence to find out if some degradation is affecting the video quality, and to identify whether the impairment is caused by the processes of encoding or transmission. Then, the most appropriate metric for that type of impairment is applied, to obtain the final quality estimation. This procedure makes possible the improvement of the performance of each metric separately, and therefore, the obtainment of accurate quality assessments in a wider range of scenarios. In the following subsections, the details of the implementation of each phase are described.

Although in the current work two metrics have been considered to deal with coding and transmission errors, the scalability of the proposed framework allows the use of more metrics to evaluate other impairments (e.g., audio impairments, service outage, etc. [20]) achieving a more general quality evaluation system. Moreover, as a first step towards a robust and reliable framework, well established standardized metrics have been considered in this work, which are FR pixel-based algorithms, both for the detection of impairments and evaluation phases of the framework. However, bitstream-based metrics could be easily used to detect transmission impairments, although the use of pixel-based metrics is also interesting for certain applications and scenarios, such as when the bitstream is encrypted.

A. Detection and Identification of Degradations

The objective of this stage of the system is to detect if the quality of the video is being affected by some degradation, and
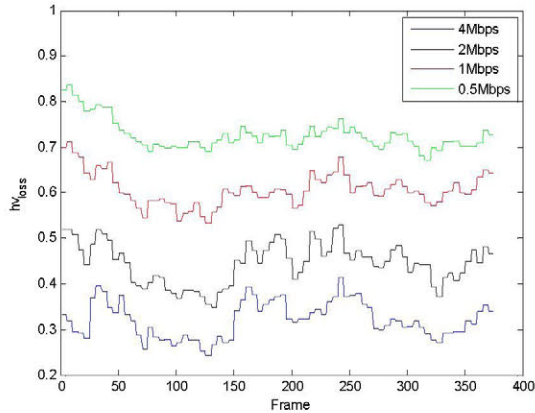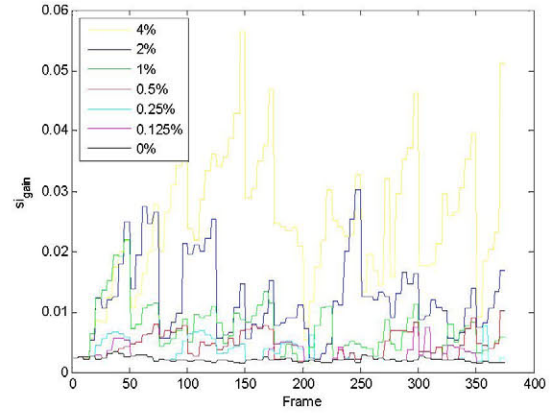
Fig. 2. VQM feature *hv_loss*

Fig. 3. VQM feature *si_gain*

Fig. 4. VQM feture *chroma_extreme*

to identify whether the impairment is caused by coding artifacts or transmission errors.

In the current work, the Video Quality Metric (VQM) proposed by Pinson and Wolf [21] has been used in this detection stage. The main reasons to consider this metric for this purpose are, firstly, that it is an already standardized method by the International Telecommunication Union (ITU) [7] and, secondly, that it is possible to access some features that could be used to identify the degradation that is impairing the quality of the video. Specifically, this RR model is based on a linear combination of seven independent parameters to provide a quality estimation of the video. These factors are related to the spatial information of the images, color, contrast, and motion activity. The normal functioning of VQM provides a single quality score for a sequence of short duration (e.g., 10 seconds). However, working with the source code, it is possible to access at a frame level the independent features that are combined, thus obtaining further information about the properties of the test video.

After the study of the parameters extracted by VQM from the video, it was found that some factors related to the spatial information and the color of the frames could be used to identify the source of the degradation impairing the video quality. In particular, these features are [21]:

- *hv_loss*: Detects a shift of edges from horizontal and vertical directions to diagonal orientations, which may be caused also by blurring effects on the edges.

- *si_gain*: Measures the edge sharpening or enhancement effects.

- *chroma_extreme*: Detects severe localized color impairments, as those caused by tiling effects resulting from transmission errors.

A descriptive example of the behavior of these features could be seen in Fig. 2, Fig. 3 and Fig. 4, respectively. On one side, *hv_loss* is represented for a sequence encoded with four different coding qualities from 4 Mbps to 0.5 Mbps. It can be observed that the four curves could be clearly differentiated allowing the identification of the effects of the encoding process. On the other side, in the figures of *si_gain* and
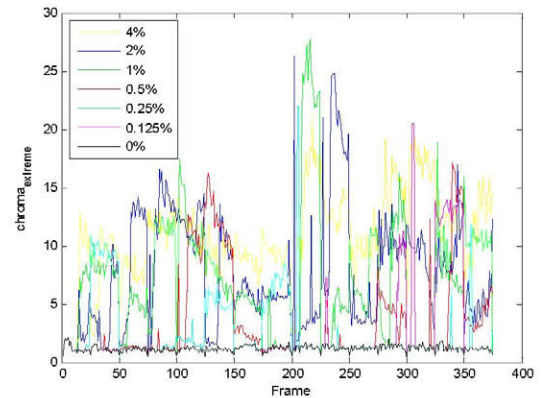
chroma_extreme, the black line corresponds to a segment without transmission degradations and a coding bitrate of 4 Mbps, and the rest are related to different patterns of packet losses. As shown, the values of these features when there are transmission errors are greater and fluctuate much more. Therefore, the detection of transmission errors could be done analyzing the parameters *chroma_extreme* and *si_gain*, since they cause severe color impairments and the appearance of sharp edges due to tiling (some macroblocks or slices are substituted with the corresponding ones in previous frames that could be incoherent). In addition, quality degradations caused by coding impairments could be identified by means of *hv_loss*, which mainly deals with properties affected by these distortions.

It is worth noting that some of the VQM features are not computed for each frame, but one value each 0.2 seconds is obtained. Thus, in the practical implementation of the detection stage, in these cases each value was held for the following frames until a new value was acquired, in order to compare all the features with the same amount of samples as frames.

B. Quality metric application

Again, although there are several video quality metrics that could be used in the proposed scheme, in this particular case,

two metrics were considered, because they have shown high performance and have been standardized by the ITU. On one side, the general model of the VQM metric has been considered for coding artifacts [7], and on the other side, VQuad [8] has been used for transmission errors.

One of the main advantages of the use of VQM is that it is also used in the first stage of the framework, which makes the system more efficient. In addition, its high performance, especially with the estimation of the impact of coding artifacts, has been proved in an exhaustive study of the VQEG [7]. Apart from the details of VQM that have been already described in the previous subsection, it is worth noting that the algorithm provides an overall quality score in the range [0, 1] from a linear combination the different spatial, temporal and chromatic features (0 corresponding to the best quality).

VQuad is a standardized FR metric that analyses various perceptual degradations, such as blockiness, tiling, blurring, and jerkiness. Then, an estimation of mean opinion score (MOS) is provided by combining these factors using a multiplicative aggregation between the effects of the temporal degradations and the spatial degradations. The quality assessment is given on a scale from 1 to 4.5, being this latter value the best possible quality. This metric could be especially useful for the evaluation of the effects of transmission errors due to the analysis of impairments such as tiling (visible macroblock and slice edges) and jerkiness (temporal artifacts like freezing).

The output of the overall system could be adjusted depending on the final application of the architecture, however, some alternatives could be just to get the output of each metric in its own scale (indicating the applied metric), or to scale the values provided by VQM and VQuad to the range [1, 5] (being 5 the worst quality), which is a common MOS scale.

IV. EXPERIMENTAL SETUP

To evaluate the performance of the proposed approach, a video database consisting of five video sources was used. The content of the sequences is representative of typical TV programs, covering various levels of motion and spatial complexity. Also, the sequences have SD resolution and no audio. The rest of the main features of the sequences are described in Table I.

These sequences were encoded using the standard H.264/AVC, and various bitrates to test the behavior of the approach with coding degradations. Besides, concerning transmission errors, six different severities of uniform packet losses were used, considering a *slicing* error concealment strategy, based on the substitution of the lost macroblocks of the frames with those correctly decoded in previous frames. Table II shows the combinations between coding bitrate and packet loss rates that were used in the tests of the proposed system for video quality estimation. As it can be seen, the effects of transmission errors, coding degradations and both jointly were considered. In total 125 videos were used to test the approach (since two different cases of degradation were considered for the case of a packet loss rate of 4% in videos encoded with 4 Mbps).

TABLE I. PROPERTIES OF THE TEST VIDEO SEQUENCES

ID	Sequence	Duration	Framerate
A	Movie trailer	15 s	24 fps
B	Interview	15 s	25 fps
C	Soccer	16 s	25 fps
D	Movie	16 s	24 fps
E	Music video	16 s	25 fps

TABLE II. DISTORTION PATTERNS CONSIDERED IN THE TESTS

Coding Bitrate	Packet Loss Rate (%)						
	0	0.125	0.25	0.5	1	2	4
Raw	X						
4 Mbps	X	X	X	X	X	X	X
2 Mbps	X		X	X		X	
1 Mbps	X	X	X	X	X	X	X
0.5 Mbps	X						

The performance of the objective metrics is given according to the standard procedures [22], thus, taking into account the correlation of the estimated quality values with the scores provided by the observers in a subjective test, described in detail in [23]. In particular, the quality assessment experiment was carried out in a viewing environment compliant to the recommendation ITU-R BT.500 [1]. The Absolute Category Rating (ACR) methodology was used with a continuous 11-point quality scale [24]. The MOS values used to test the performance of the proposed framework were obtained from the scores of 21 observers and mapping to a five-grade quality scale.

V. EVALUATION RESULTS

In the following subsections the experimental results of both steps of the proposed architecture are presented, as well as the performance of the whole system in estimating subjective scores.

A. Results for Impairment Detection

As aforementioned, the detection of the type of impairment has been done analyzing some features of the VQM metric temporally. Specifically, the values of these features over the whole test video segment have been considered (i.e., 15-16 seconds). Nevertheless, there is an interesting possibility based on the continuous monitoring of these values frame by frame, detecting some events (e.g., abrupt changes) that could indicate the beginning of a certain type of distortion. This way it is possible to set an interval that could be analyzed by the corresponding quality metric. This approach would be considered in future work.

Taking this into account, to identify transmission errors, the standard deviations of the features *si_gain* and *chroma_extreme* for a complete video sequence have been considered. In fact, the values of these statistics have been compared with the standard deviation of those features for a reference video. In particular, the sequences encoded at 4Mbps (highest coding quality apart from the uncompressed versions,

and a realistic quality to be broadcasted) for each video content have been considered as references to compare with the segments affected by transmission errors and with those encoded with lower qualities. In essence, when one of the following conditions is satisfied, a transmission error is considered, and the segment is analyzed by VQuad in the following step of the system. In the following expressions, the sub-index T corresponds to the test sequence, while R is for the reference.

$$std(si\_gain_T) > std(si\_gain_R) \cdot T_1 \qquad (1)$$

$$std(chroma\_extreme_T) > std(chroma\_extreme_R) \cdot T_1 \qquad (2)$$

In addition, to improve the performance of this stage reducing the number of coding impairments classified as transmission errors, a condition over the mean of the feature hv_loss has been considered. In particular, if the following condition is satisfied, a coding impairment is identified, so contrasting this condition with the previous ones, false detections of transmission impairments could be reduced.

$$mean(hv\_loss_T) > mean(hv\_loss_R) \cdot T_2 \qquad (3)$$

Having all this in mind, the thresholds were set to values that optimized the correct detection of the impairments under consideration. Thus, being $T_1=4$ and $T_2=1.5$, the 94.74% of the transmission errors were correctly identified, without any false detection, and the rest of the impairments were classified as coding artifacts. Further work will be also focused on the possibility of applying non-absolute factors, but adaptive factors related to the content under analysis.

B. Performance of the Proposed Framework

The performance of the metrics in the estimation of the quality of the video affected by the considered coding and transmission impairments, has been done according to typical procedures, i.e. comparing the estimated values with the MOS scores obtained from subjective tests with the same test material [11]. In particular, the correlation between the estimations and the subjective ratings has been analyzed computing the Pearson coefficient (PC) and the Spearman coefficient (SC). In addition, the root mean squared error (RMSE) and the outlier ratio (OR) have been calculated. Outliers have been considered those estimated values that are different from the corresponding subjective rating in more than twice the standard deviation obtained from the subjective tests. In the cases under study no outliers were obtained.

The performance of the whole system is evaluated taking into account that the transmission errors correctly identified in the detection stage are analyzed by the VQuad metric, while the rest are evaluated with VQM, since they are considered coding errors. The results obtained are shown in Table III, where the performances of other popular metrics of the state of the art are provided to compare the reliability of the proposed framework. In particular, these metrics are: Peak Signal-to-Noise ratio (PSNR) computed according to the implementation in [25], Visual Signal-to-Noise Ratio (VSNR) [26], and SSIM [16]. PSNR is the most popular objective quality metric due to its simplicity, since just the intensity differences between the reference and the test frames are computed. However, it has been shown that the performance of PSNR may be poor in terms of correlation with subjective scores in many cases [27]. VSNR is a wavelet-based metric that measures the visual fidelity of images based on human vision properties. Finally, SSIM is one of the most popular FR metrics based on the analysis of the degradation of the structural information of the images. In the tests, to obtain a global score for a video sequences from SSIM and VSNR, the average of the results for each frame was computed. Furthermore, the performance of VQM and VQuad on the whole dataset are included in the table. Previously to correlate the scores of the metrics with the subjective results, they were mapped using polynomial functions for VSNR, SSIM, VQM, a logistic function for PSNR, and a linear transformation for VQuad, since these were the best performing mappings in each case [22]. In addition, the results obtained using VQM over all the sequences degraded with coding artifacts, and VQuad over all those impaired by transmission errors are shown (VQM&VQuad). This would be the ideal case of operation of the framework, since this means that all the impairments are correctly identified. In fact, as shown in the table, the best performance is reported in this case. Also, the results provided by the global framework are considered, that is, taking into account the performance of the detection stage, so VQM was applied over all the coding impairments and 5.26% of the transmission errors, and VQuad over the rest (see subsection V.A).

To compare the performances of the metrics, statistical analysis should be considered to prove whether the differences are statistically significant [28]. Thus, to study the statistical significance between two correlation coefficients the method for overlapping correlations was used [29], since the compared correlations share a variable in common (the MOS obtained from the subjective tests). The results reflect the best performance of the ideal framework (VQM&VQuad) over the rest of the metrics. A slight reduction of the performance is obtained with the global framework (i.e., including the detection step). However, on one side, the difference between the correlation coefficients of these two cases are not statistically significant at level $p<0.05$; and on the other side, it could be seen that the good performance on detecting and identifying the impairments makes that the estimation results of the overall framework are still better than using VQM or VQUAD independently, and provides better quality estimations than the rest of the metrics considered in the tests (i.e., SSIM, PSNR, VSNR). The differences between the correlation coefficients of the global framework against the rest of the metrics are statistically significant at level $p<0.05$.

TABLE III. PERFORMANCE OF THE WHOLE SYSTEM COMPARED WITH OTHER METRICS

Metric	PC	SC	RMSE
PSNR	0.771	0.777	0.335
VSNR	0.852	0.874	0.276
SSIM	0.757	0.731	0.344
VQM	0.707	0.646	0.372
VQuad	0.883	0.867	0.246
VQM&VQuad	0.916	0.902	0.208
Global framework	0.911	0.902	0.219

In Fig. 5 the performance of the whole system is depicted, representing the video quality estimations over the MOS values obtained from the subjective tests. A slight deviation could be observed between both sets of points, due to the mapping of the estimations. In the figure, the red points correspond to coding impairments and the blue ones to degradations caused by transmission errors.

VI. CONCLUSION

In this paper a rule-based system combining video quality metrics has been proposed, with the aim of increasing the performance and the range of application in comparison to use one metric individually.

The system consists of a first step where the degradation is detected and classified as a distortion caused by the encoding process or by transmission. Then, the impaired video segment is analyzed with one metric specialized for that degradation. The efficient use of standardized quality metrics has been one of the objectives of this particular implementation of the framework, thus VQM was considered for coding artifacts and VQuad for transmission errors. The obtained results have shown that this way it is possible to increase the performance of each metric separately considering both types of distortions.

Future work will be focused on the improvement of the stage for detecting and identifying impairments using adaptive thresholding, and the study of using no-reference video quality metrics for quality monitoring in video distribution networks.

Fig. 5. Results of the combination between VQM and VQuad

REFERENCES

[1] ITU, "Methodology for the subjective assessment of the quality of television pictures", *Recommendation ITU-R BT. 500-13*, Jan. 2012.

[2] J. Apostolopoulos and A. Reibman, "The challenge of estimating video quality in video communication applications", *IEEE Signal Processing Magazine*, vol. 29, no. 2, pp. 156–160, Mar. 2012.

[3] ITU-T SG 12, "Definition of Quality of Experience (QoE) ", COM12-LS 62 – E, TD 109rev2 (PLEN/12), Geneva, Switzerland, Jan. 2007.

[4] S. Winkler and P. Mohandas, "The Evolution of Video Quality Measurement: From PSNR to Hybrid Metrics", *IEEE Transactions on Broadcasting*, vol. 54, no. 3, pp. 660–668, Sep. 2008.

[5] T. Liu, W. Lin, and C.-C. J. Kuo, "A multi-metric fusion approach to visual quality assessment", *International Workshop on Quality of Multimedia Experience*, pp. 72–77, Mechelen, Belgium, Sep. 2011.

[6] A. Barri, A. Dooms, B. Jansen, and P. Schelkens, "A locally adaptive system for the fusion of objective quality measures", *IEEE Trans. Image Process.*, vol. 23, no. 6, p. 1, Jun. 2014.

[7] ITU, "Objective perceptual video quality measurement techniques for digital cable television in the presence of a full reference, *Recommendation ITU-T J.144*, Mar. 2004.

[8] ITU, "Objective perceptual multimedia video quality measurement of HDTV for digital cable television in the presence of a full reference", *Recommendation ITU-T J.341*, Jan. 2011.

[9] 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 Video Processing*, vol. 11, no. 629284, pp. 1–14, Mar. 2011.

[10] A. Benoit, P. Le Callet, P. Campisi, and R. Cousseau, "Quality assessment of stereoscopic images", *EURASIP Journal on Image and Video Processing*, vol. 2008, pp. 1–13, Jan. 2009.

[11] VQEG, "Hybrid Perceptual/Bitstream validation test final report", Jul. 2014.
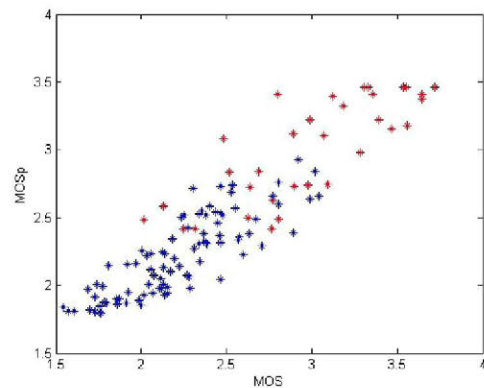
[12] VQEG, "JEG Hybrid evaluation project," Mar. 2013.

[13] A. Moorthy and A. Bovik, "Blind image quality assessment: from natural scene statistics to perceptual quality", *IEEE transactions on Image Processing*, vol. 20, no. 12, pp. 3350–64, Dec. 2011.

[14] A. Leontaris, P. C. Cosman, and A. R. Reibman, "Quality evaluation of motion-compensated edge artifacts in compressed video", *IEEE transactions on image processing*, vol. 16, no. 4, pp. 943–56, Apr. 2007.

[15] P. V Pahalawatta and A. M. Tourapis, "Motion estimated temporal consistency metrics for objective video quality assessment", *International Workshop on Quality of Multimedia Experience*, pp. 174–179, San Diego, USA, Jul. 2009.

[16] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, "Image quality assessment: From error visibility to structural similarity", *IEEE Trans. on Image Processing*, vol. 13, no. 4, pp. 600–612, Apr. 2004.

[17] M. Farias and S. Mitra, "No-reference video quality metric based on artifact measurements," in *IEEE International Conference on Image Processing*, vol. 3, pp. 141–144., Genoa, Italy, Sep. 2005.

[18] H. B. Martinez and M. C. Q. Farias, "A no-reference audio-visual video quality metric," *Eur. Signal Process. Conf.*, pp. 2125–2129, Sep.

[19] P. Pérez, J. Gutiérrez, J. J. Ruiz, and N. García, "Qualitative monitoring of video quality of experience", *Proc. IEEE International Symposium on Multimedia*, pp. 470–475, Dana Point, USA, Dec. 2011.

[20] G.W. Cermak, "Consumer opinions about frequency of artifacts in digital video", *IEEE Journal on Selected Topics in Signal Processing*, vol. 3, no. 2, pp. 336-343, Apr. 2009.

[21] 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, Sep. 2004. (Code available: http://www.its.bldrdoc.gov/resources/video-quality-research/software.aspx)

[22] ITU, "Methods, metrics and procedures for statistical evaluation, qualification and comparison of objective quality prediction models", *Recommendation ITU-T P.1401*, Jul. 2012.

[23] M.-N. Garcia and A. Raake, "Frame-layer packet-based parametric video quality model for encrypted video in IPTV services," *International Workshop on Quality of Multimedia Experience*, pp. 102–106, Mechelen, Belgium, Sep. 2011.

[24] ITU, "Subjective video quality assessment methods for multimedia applications", *Recommendation ITU-T P.910*, Mar. 1999.

[25] Institute for Telecommunication Sciences, Video Quality Metric Software available at: http://www.its.bldrdoc.gov/vqm/

[26] D. M. Chandler, S. S. Hemami, and S. Member, "VSNR: A wavelet-based visual signal-to-noise ratio for natural images", *IEEE Transactions on Image Processing*, vol. 16, no. 9, pp. 2284–2298, 2007.

[27] Q. Huynh-Thu and M. Ghanbari, "Scope of validity of PSNR in image/video quality assessment", *Electronics Letters*, vol. 44, no. 13. pp. 800-801, Jun. 2008.

[28] ITU, "Method for specifying accuracy and cross-calibration of Video Quality Metrics (VQM)", *Recommendation ITU-T J.149*, Mar. 2004.

[29] G. Y. Zou, "Toward using confidence intervals to compare correlations", *Psychological methods*, vol. 12, no. 4, pp. 399–413, Dec. 2007.