DOI: 10.2352/ISSN.2470-1173.2016.15.IPAS-197

Video Quality of Experience metric for streaming services

Pradip Paudyal and Yiwei Liu and Federica Battisti and Marco Carli; Department of Engineering, Roma TRE University, Rome, Italy

Abstract

The knowledge of the user Quality of Experience (QoE) of the accessed services is of crucial importance for the robust design and adaptation of multimedia streaming networks. In this article, a video QoE estimation metric for video streaming services is presented. The proposed metric does not require information on the original video or the impairments affecting the communication channels. Results show that the proposed method can efficiently estimate the video QoE.

Introduction

Advancements in multimedia infrastructures and access devices result in an large demand for multimedia services, in particular video streaming and video conference. In this context the users Quality of Experience (QoE), which is the degree of delight or annoyance of the user of an application or service [1] , is becoming a key factor for network and content service providers. For this reason, there is an increasing interest in shifting the focus of quality assessment from compliance with system design goals to fulfillment of user needs or expectations [2, 3] . In this scenario, the interests of the academic and industry researchers are pooled toward the efficient design and adaptation of multimedia communication networks, particularly for video. For this purpose effective and efficient design of a real time video quality metric is of crucial importance.

In state-of-the-art, many efforts have been devoted to estimate the video quality or the QoE. Most of them are investigating the effects of common network parameters on the received video. Authors in [4] studied the impacts of temporal jerkiness on video quality, which is caused by packet loss or late arrival of packets and, based on their findings, propose a neural network based Video Quality Metric (VQM). Authors in [5] present an analytical model for no-reference video quality metric by taking into account both video play-out rate and network throughput. A VQM based on the spatio-temporal natural scene statistics and motion coherency in the video scenes has been proposed in [29] . A video quality metric based on the statistical estimation of the Peak Signal-to-Noise Ratio (PSNR) of the coded transform coefficients for H.264/AVC encoded sequences is presented in [7] . By considering the features of H.264/AVC encoding, such as blocking, blurring, and spatial activity, a VQM is presented in [8] . In [9] a machine learning approach is recommended to estimate the QoE. The QoE is represented as an engagement and expressed as a function of the quality metrics. The engagement could be the video play time, number of visit and the quality metric represents observed indicates such as buffering ratio, average bit rate, etc. In [10] a video quality estimation method is proposed by considering the policies applied for packet processing by routers and the level of total network utilization. In [11] the authors propose a novel rate adaptation algorithm called QoE enhanced adaptation algorithm over Dynamic Adaptive Streaming over HTTP (DASH). The adaptation algorithm preserves the minimum buffer length to avoid interruption and minimizes the video quality changes during the playback. In [12] a multi-factor QoE evaluation model based on the content classification by spatial and temporal information for H.264/AVC encoded video has been proposed. Authors in [13] propose an acceptability-based QoE estimation model by considering encoding parameters, bit rate, video content characteristics, and mobile device display resolution. A QoE metric for HDTV by using PSNR metric and PLR (Packet Loss Rate) artifact is presented in [14] . Article [15] presents a methodology to analyze the impact of different MAC-level parameters on video QoE over IEEE 802.11n wireless networks by using Random Neural Network (RNN). The authors in [20] present user perceived QoE prediction/mapping model (natural and generic exponential model) from network related QoS parameters. Finally, [17] presents the performance comparison of QoS to QoE mapping models for wired and wireless communications.

The above mentioned state-of-the-art metrics have the following limitations: i) they are not specifically designed for QoE estimation for video streaming services, ii) they are complex and time consuming, iii) their performances are far from the subjective scores in a real scenario, and iv) they generally assume the presence of one main network artifact. In practical streaming services this approach can lead to a wrong QoE estimation, since only part of data can be affected by one particular impairment.

The goal of this article is to propose a generic blind video QoE metric, called V_{QoE}, for streaming services. Since the metric is designed for real time services, the frame by frame QoE estimation approach has been used. The frame quality is expressed by considering the portion of frames that, due to transmission errors, can not be correctly decoded. These will be in the following addressed to as *broken blocks*. The overall video QoE is expressed as the average of the QoE estimated for all frames. The results show that the proposed method is computationally less complex and faster than the other considered blind video quality metrics and its performances are superior.

The rest of the paper in organized as follows: Section presents the proposed approach, Section describes the performed tests for evaluating the effectiveness of the proposed metric and in Section the conclusions are drawn.

Proposed Approach

The video QoE is computed in three steps: i) the total number of broken blocks for each frame is computed, ii) the QoE for each frame is estimated from the total number of broken blocks by using a mapping function between broken blocks and QoE, iii) the overall video quality is estimated as the average of frame quality.

©2016 Society for Imaging Science and Technology
DOI: 10.2352/ISSN.2470-1173.2016.15.IPAS-197

Broken blocks estimation

The basic steps for the estimation of the number of broken blocks are detection and verification.

In detail [18] , let F_k be the k^{th} generic frame of the video sequence. It can be partitioned in $N_r \times N_c$ blocks $\mathbf{B}_k^{(i,j)}$ of $r \times c$ pixels with top-left corner located in (i,j). Moreover, $\Delta\mathbf{B}_k^{(i,j)} = \mathbf{B}_k^{(i,j)} - \bar{B}_k^{(i,j)}$ denotes the deviation of the luminance in k^{th} frame of the block $\mathbf{B}_k^{(i,j)}$ from the corresponding mean values.

The inter-block correlation $\rho_k^{B(i,j)}$ can be computed as:

$$\rho_k^{\mathbf{B}(i,j)} = \frac{\left\langle \Delta\mathbf{B}_k^{(i,j)}, \Delta\mathbf{B}_{k-1}^{(i,j)} \right\rangle}{\left\| \Delta\mathbf{B}_k^{(i,j)} \right\|_{L_2} \left\| \Delta\mathbf{B}_{k-1}^{(i,j)} \right\|_{L_2}}. \tag{1}$$

where $< \bullet, \bullet >$ denotes the inner product and $\|\bullet\|_{L_2}$ the L_2-norm.

To identify the distorted block, the blocks have been classified into three groups; low, medium and high content variation. The content variations are evaluated based on their temporal inter-block correlation $\rho_k^{B^{(i,j)}}$. Moreover, based on the content variation groups, the corresponding variability map $\blacksquare_k^V = \{\Gamma_k^{VB^{(i,j)}}\}$ has been defined by comparing the inter-frame correlation of each block as:

$$\Gamma_k^{VB^{(i,j)}} = \begin{cases} 1, & if\ \rho_k^{B^{(i,j)}} < \theta_l \\ 0, & if\ \theta_l \le \rho_k^{B^{(i,j)}} \le \theta_h \\ 2, & if\ \rho_k^{B^{(i,j)}} > \theta_h \end{cases} \tag{2}$$

where the two thresholds, θ_l and θ_h have been selected in order to grant $|P_{fa} - P_{md}| < \varepsilon_1$. Here P_{fa} is the probability of false alarm, P_{md} is the probability of missed detection, and ε_1 is a significantly small value. Value of the ε_1 is experimentally determined from the training session. In this study the values of θ_l and θ_h are set to 0.2 and 0.9.

The blocks with medium content variation, $(\Gamma_k^{VB^{(i,j)}} = 0)$ are less likely to be broken. The blocks with low and high content variation should be further analysed, to find out whether they should be considered as broken blocks or not. For each block, if $\Gamma_k^{VB^{(i,j)}} = 2$, it has to be further checked. If at least v blocks among the surrounding blocks present a strong temporal correlation, where the parameter v has been identified by experimental tests, then the block is classified as belonging to a static region. Then its potential distortion index $\Gamma_k^{CB^{(i,j)}}$ is set to zero, in other words the block is not considered as a broken block. That is:

$$\Gamma_k^{CB^{(i,j)}} = \begin{cases} 0 & if\ |\varsigma| > v \\ \Gamma_k^{VB^{(i,j)}} & otherwise \end{cases} \tag{3}$$

where ς equals to how many surrounding blocks of $B_k^{(i,j)}$ with $\Gamma_k^{VB^{(p,q)}} = 2$.

Finally, if $\Gamma_k^{CB^{(i,j)}} \neq 0$, then further test is necessary. Let E_l and E_r be the L_1 norms of the vertical edges respectively on the left and on the right boundary of the block, and with A_c, A_l and A_r the average values of the L_1 norms of the vertical edges inside the current block and of the left and right adjacent blocks. A block with $\Gamma_k^{CB^{(i,j)}} \neq 0$ is classified as affected by visible distortion if:

$$\left| E_l - \frac{(A_c + A_l)}{2} \right| > \theta \quad or \quad \left| E_r - \frac{(A_c + A_r)}{2} \right| > \theta \tag{4}$$

where the threshold θ has been defined on the basis of experimental trials (in this article $\theta = 100$). In particular it corresponds to JND (Just Noticeable Difference) collected and evaluated for 90% of subjects.

The same procedure is applied to the horizontal direction. If the block edges are consistent (i.e. no visible distortion has been detected along horizontal and vertical directions) $\Gamma_k^{CB^{(i,j)}}$ is reset to 0. As a result, total number of broken block (TOT_k) is computed as the total amount of blocks where $\Gamma_k^{CB^{(i,j)}} \neq 0$.

QoE estimation

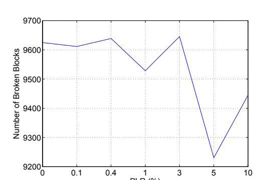
After the total number of broken blocks in a frame is computed, the QoE has to be determined from the total number of broken blocks by using a mapping function. In the state-of-art many QoS to QoE mapping models are presented [19] . In [17] the performances of widely used Quality of Service (QoS) to QoE mapping models are presented. From the result it is shown that some of the models including IQX Hypothesis [20] have a very good mapping capability for key QoS parameters: jitter, PLR and bandwidth limitation and their performances are very close. Therefore in this article, based on its mapping capability and simplicity, IQX Hypothesis has been selected to estimate the QoE from the total number of broken blocks. Hence, the QoE for k^{th} frame is estimated as:

$$QoE_k = a.exp(-b.(TOT_k)) + c \tag{5}$$

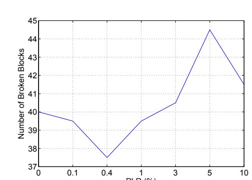
where a, b, and c are the regression parameters which can be achieved from the training session and TOT_k is the total number of broken blocks for k^{th} frame. The overall video QoE is considered as the average of the frame quality.

Results and discussion

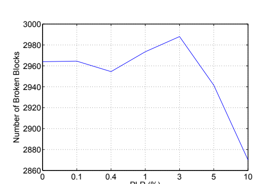
To evaluate the performances of the proposed algorithm, the availability of a video quality database is very important. In state-of-art, many video quality database have been recommended [21] , among them for this study the EPFL-PoliMI video quality assessment database [24] and LIVE Video Quality Database [23] are considered.



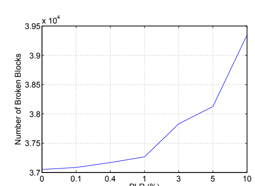
(a) *Crowdrun* Cb component



(b) *Ice* Cb component



(c) *Foreman* Cb component

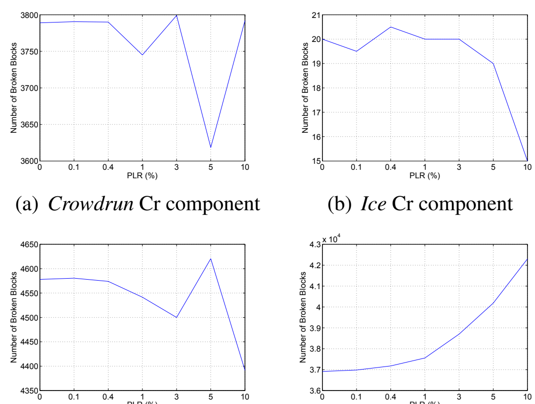


(d) *Mobile* Cb component

Figure 1. *Total number of broken blocks of the Cb component for the 4CIF (*Crowdrun *and* Ice*) and CIF (*Foreman*, and* Mobile*) videos.*

DOI: 10.2352/ISSN.2470-1173.2016.15.IPAS-197

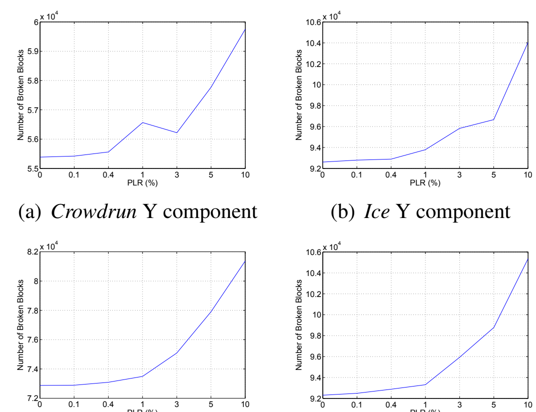
For the analysis from the EPFL-PoliMI database, 156 video streams (78 video sequences at CIF and 78 sequences at 4CIF spatial resolution), encoded with H.264/AVC and corrupted by simulating the packet loss due to transmission over an error-prone network and their corresponding subjective scores, the Mean Opinion Score(MOS), are considered. In this database, in order to simulate burst errors, six different PLR (0.1%, 0.4%, 1%, 3%, 5%, 10%) patterns were used and every reference sequence has a two impaired sequences for same level of PLR. From the LIVE Video Quality Database, 80 videos (with a resolution of 768X432 pixels and H.264/AVC encoded) and their corresponding subjective scores, the Difference Mean Opinion Scores (DMOS), are considered. Among all videos, 10 are the reference sequences, 40 were affected by the wireless distortions (four test videos per reference) and the remaining 30 were affected by IP distortions (three test videos per reference). The details about the databases are presented in [24] and [23] .



(a) *Crowdrun* Cr component (b) *Ice* Cr component

(c) *Foreman* Cr component (d) *Mobile* Cr component

Figure 2. *Total number of broken blocks of the Cr component for the 4CIF (*Crowdrun *and* Ice*) and CIF (*Foreman *and* Mobile*) videos.*



(a) *Crowdrun* Y component (b) *Ice* Y component

(c) *Foreman* Y component (d) *Mobile* Y component

Figure 3. *Total number of broken blocks of the Y component for the 4CIF (*Crowdrun *and* Ice*) and CIF (*Foreman *and* Mobile*) videos.*

From the analysis of the achieved results it can be notices that mainly the luminance component of the video has a direct and significant impact on the video quality. For sake of compactness in the following only the results of a subset of all considered videos will be included: the 4CIF video sequences *Crowdrun* and *Ice* and the CIF sequences *Foreman* and *Mobile* extracted from the EPFL-PoliMI database. The analysis has been performed on 52 impaired video sequences: the original *Crowdrun*, *Ice*, *Foreman* and *Mobile* sequences and 12 impaired ones obtained by considering six PLR values. For each video sequence, the relation between the PLR and total number of broken blocks for each video chrominance component (Cb and Cr) is shown in Figures 1 and 2. It can be noticed that the total number of broken blocks does not increase for higher values of PLR. However, for the luminance component Y, from Figure 3 it can be noticed that to high values of PLR correspond to higher number of broken block, and this trend is confirmed for all the considered video sequences. From these results we can conclude that mainly the luminance component shows a closer and synclastic relationship between PLR and broken blocks. As demonstrated in [17] [20] this behavior has an impact on QoE. Base on these considerations, it is possible to reduce the computational complexity by considering only the luminance component for quality metric assessment.

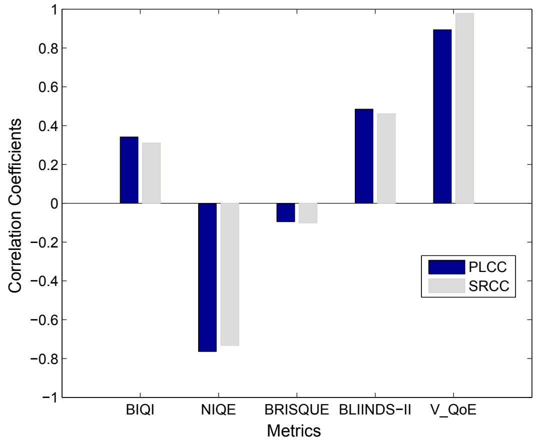
In the following, the QoE computed through the proposed metric, V_{QoE} is compared with other widely discussed No-Reference image/video quality metrics: Naturalness Image Quality Evaluator (NIQE) [25] , Blind Image Quality Index (BIQI) [26] , Blind/Referenceless Image Spatial QUality Evaluator (BRISQUE) [27] , BLind Image Integrity Notator using DCT Statistics-II (BLIINDS-II) Index [28]) and Blind prediction of natural video quality [29] . Among them, the first four metrics have been initially designed for image quality estimation and lately used for video quality estimation by averaging the single frame pair quality score for all the video. More recently, these tools have also been used for the performance comparison of newly proposed video quality metrics [29] . To collect the subjective assessment a subjective experiment has been designed and performed, following [30, 31] recommendations. To compare the performance of the metrics, the Spearman Rank Order Correlation Coefficient (SROCC) and Pearson Linear Correlation Coefficient (PLCC) between the collected subjective socres and predicted scores have been computed. During the experiment 80% of the videos were selected randomly for the training and rest of the videos were used for test.

The performance of the metrics for CIF and 4CIF videos from EPFL-PoliMI database are presented in Figure 4. The results show that the proposed method V_{QoE} has highest values of PLCC and SROCC compared to the considered metrics for both CIF and 4CIF videos. Besides that, the proposed method also shows higher PLCC and SRCC values than Video BLIND ($PLCC = 0.75$ and $SRCC = 0.807$ as mentioned in [29]) for the same database.

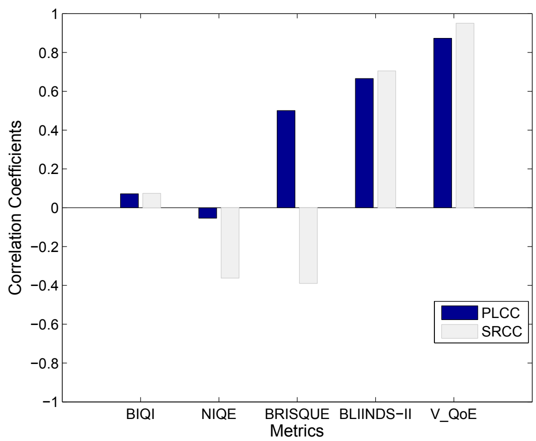
Moreover, the performances of the proposed method are also evaluated for the LIVE Video Quality Database. Among all videos available in the LIVE database we selected only a subset that is more related to the purpose of the proposed metric V_{QoE}. We considered the distorted videos in the categories wireless and IP together with their corresponding subjective scores. The results showthat the metric V_{QoE} has $PLCC = 0.7909$ and $SRCC = 0.8571$. As presented in [29] for the same database but for all the video categories (MPEG-2, H.264, wireless, and IP) the performance of Video BLIND is $PLCC = 0.881$ and $SRCC = 0.759$.

For real time video streaming services the computational

DOI: 10.2352/ISSN.2470-1173.2016.15.IPAS-197



(a) For CIF videos



(b) For 4CIF videos

Figure 4. *Performance comparison of the proposed metric (V_{QoE}) with considered state-of-art metrics.*

complexity and processing time of the QoE metric is as important as its QoE prediction capability. The considered state of art metrics including Video BLIND are more complex, and need more processing power and computational time. This is mainly due to the fact that the metrics have been designed by considering computationally heavy and complex techniques [32] , as generalized Gaussian density parameter estimation techniques and motion coherency computation. In the proposed metric V_{QoE} only the luminance component has been considered for QoE estimation and the overall processing time depends mainly on the block-wise correlation operation.

Conclusion

In this paper, a blind and realtime video Quality of Experience (QoE) metric V_{QoE} for video streaming network, has been proposed. The video quality is computed based on the video distortion, which is measured as number of broken blocks. The frame QoE has been derived from number of broken blocks by means of IQX Hypothesis. The overall video quality is expressed as an average of frame QoE. The performance of the proposed method is compared with widely discussed blind image/video metrics. The result shows that the proposed metric outperforms the considered state of art metrics and is also faster.
As a future work, the performance of the algorithm will be tested on other databases and compared with other metrics.

References

[1] Le Callet, Patrick, Sebastian Mller, and Andrew Perkis, Qualinet white paper on definitions of quality of experience, European Network on Quality of Experience in Multimedia Systems and Services (COST Action IC 1003) (2012).

[2] Timmerer Christian , Ebrahimi Touradj, and Pereira Fernando, Toward a New Assessment of Quality, IEEE Computer Society, Vol. 48, pg. 108-110. (2015).

[3] Battisti, Federica, Marco Carli, Elena Mammi, and Alessandro Neri, A study on the impact of AL-FEC techniques on TV over IP Quality of Experience, EURASIP J. Adv. Sig. Proc. 2011 (2011): 86.

[4] Xue, Yuanyi, Beril Erkin, and Yao Wang, A Novel No-Reference Video Quality Metric for Evaluating Temporal Jerkiness due to Frame Freezing, Multimedia, IEEE Transactions on 17, no. 1 (2015): 134-139.

[5] Seyedebrahimi, Mirghiasaldin, Christopher Bailey, and Xiao-Hong Peng, Model and performance of a no-reference quality assessment metric for video streaming, Circuits and Systems for Video Technology, IEEE Transactions on 23, no. 12 (2013): 2034-2043.

[6] Saad, Michele A., Alan C. Bovik, and Christophe Charrier, Blind prediction of natural video quality, Image Processing, IEEE Transactions on 23, no. 3 (2014): 1352-1365.

[7] Brando, Toms, and Maria Paula Queluz, No-reference quality assessment of H. 264/AVC encoded video. Circuits and Systems for Video Technology, IEEE Transactions on 20, no. 11 (2010): 1437-1447.

[8] Oelbaum, Tobias, Christian Keimel, and Klaus Diepold, Rule-based no-reference video quality evaluation using additionally coded videos, Selected Topics in Signal Processing, IEEE Journal of 3, no. 2 (2009): 294-303.

[9] Balachandran, Athula, Vyas Sekar, Aditya Akella, Srinivasan Seshan, Ion Stoica, and Hui Zhang, A quest for an internet video quality-of-experience metric, In Proceedings of the 11th ACM workshop on hot topics in networks, pp. 97-102. ACM, 2012.

[10] Sevcik, Lukas, Miroslav Voznak, and Jaroslav Frnda, QoE prediction model for multimedia services in IP network applying queuing policy, In Performance Evaluation of Computer and Telecommunication Systems (SPECTS 2014), International Symposium on, pp. 593-598. IEEE, 2014.

[11] Suh, Dongeun, Insun Jang, and Sangheon Pack, QoE-enhanced adaptation algorithm over DASH for multimedia streaming, In Information Networking (ICOIN), 2014 International Conference on, pp. 497-501. IEEE, 2014.

[12] Liu, Jichun, Yang Geng, Deyuan Wang, Wenjing Li, and Xuesong Qiu, An objective multi-factor QoE evaluation based on content classification for H. 264/AVC encoded video, In Computers and Communications (ISCC), 2013 IEEE Symposium on, pp. 000137-000142. IEEE, 2013.

[13] Song, Wei, and Dian W. Tjondronegoro, Acceptability-based QoE models for mobile video, Multimedia, IEEE Transactions on 16, no. 3 (2014): 738-750.

[14] Issa, Omneya, Filippo Speranza, Wei Li, and Hong Liu, Estimation of time varying QoE for high definition IPTV distribution, In Con-

DOI: 10.2352/ISSN.2470-1173.2016.15.IPAS-197

sumer Communications and Networking Conference (CCNC), 2012 IEEE, pp. 326-330. IEEE, 2012.

[15] Paudel, Indira, Jeevan Pokhrel, Bachar Wehbi, Ana Cavalli, and Badii Jouaber, Estimation of video QoE from MAC parameters in wireless network: A Random Neural Network approach, In Communications and Information Technologies (ISCIT), 2014 14th International Symposium on, pp. 51-55. IEEE, 2014.

[16] Fiedler, Markus, Tobias Hossfeld, and Phuoc Tran-Gia, A generic quantitative relationship between quality of experience and quality of service, Network, IEEE 24, no. 2 (2010): 36-41.

[17] Battisti, Federica, Marco Carli, and Pradip Paudyal, QoS to QoE mapping model for wired/wireless video communication, In Euro Med Telco Conference (EMTC), 2014, pp. 1-6. IEEE, 2014.

[18] Battisti, Federica, Marco Carli, and Alessandro Neri, No reference quality assessment for MPEG video delivery over IP, EURASIP Journal on Image and Video Processing 2014, no. 1 (2014): 1-19.

[19] Alreshoodi, Mohammed, and John Woods, Survey on QoE QoS correlation models for multimedia services, arXiv preprint arXiv:1306.0221 (2013).

[20] Fiedler, Markus, Tobias Hossfeld, and Phuoc Tran-Gia, A generic quantitative relationship between quality of experience and quality of service, Network, IEEE 24, no. 2 (2010): 36-41.

[21] Fliegel, Karel, QUALINET Multimedia Databases v5. 5.” (2014).

[22] De Simone, Francesca, Matteo Naccari, Marco Tagliasacchi, Frederic Dufaux, Stefano Tubaro, and Touradj Ebrahimi, Subjective assessment of H. 264/AVC video sequences transmitted over a noisy channel, In Quality of Multimedia Experience, 2009. QoMEx 2009. International Workshop on, pp. 204-209. IEEE, 2009.

[23] Seshadrinathan, Kalpana, Rajiv Soundararajan, Alan Conrad Bovik, and Lawrence K. Cormack, Study of subjective and objective quality assessment of video, Image Processing, IEEE transactions on 19, no. 6 (2010): 1427-1441.

[24] De Simone, Francesca, Matteo Naccari, Marco Tagliasacchi, Frederic Dufaux, Stefano Tubaro, and Touradj Ebrahimi, Subjective assessment of H. 264/AVC video sequences transmitted over a noisy channel, In Quality of Multimedia Experience, 2009. QoMEx 2009. International Workshop on, pp. 204-209. IEEE, 2009.

[25] Mittal, Anish, Ravi Soundararajan, and Alan C. Bovik, Making a completely blind image quality analyzer, Signal Processing Letters, IEEE 20, no. 3 (2013): 209-212.

[26] Moorthy, Anush Krishna, and Alan Conrad Bovik, A two-step framework for constructing blind image quality indices, Signal Processing Letters, IEEE 17, no. 5 (2010): 513-516.

[27] Mittal, Anish, Anush Krishna Moorthy, and Alan Conrad Bovik, No-reference image quality assessment in the spatial domain, Image Processing, IEEE Transactions on 21, no. 12 (2012): 4695-4708.

[28] Saad, Michele A., Alan C. Bovik, and Christophe Charrier, Blind image quality assessment: A natural scene statistics approach in the DCT domain, Image Processing, IEEE Transactions on 21, no. 8 (2012): 3339-3352.

[29] Saad, Michele A., Alan C. Bovik, and Christophe Charrier, Blind prediction of natural video quality, Image Processing, IEEE Transactions on 23, no. 3 (2014): 1352-1365.

[30] Assembly, ITU Radiocommunication, Methodology for the subjective assessment of the quality of television pictures, International Telecommunication Union, 2003.

[31] Video Quality Experts Group, Report on the validation of video quality models for high definition video content, tech. re p., http://www. vqeg. org (2010).

[32] Wen-Hsiung, Chen, C. Smith, and S. Fralick, A Fast Computational Algorithm for the Discrete Cosine Tranfsorm, IEEE Transactions on Communications 25, no. 9 (1977): 1004-1009.