Study of No-Reference Video Quality Metrics for HEVC Compression

Kais Rouis¹, Mikolaj Leszczuk², Lucjan Janowski², Zdzisław Papir², and Jamal Bel Hadj Tahar¹

¹ Innov’COM Lab, Sup’Com, University of Carthage, École Nationale d’Ingénieurs de Tunis, Université de Tunis El Manar, Tunisia
² AGH University of Science and Technology, Department of Telecommunications, Krakow, Poland

Abstract—The paper proposes a No-Reference (NR) quality assessment measurement originally developed for H.264, used for High Efficiency Video Coding (HEVC). In particular, authors present an investigation of NR metrics to objectively estimate the perceptual quality of a set of processed video sequences. The authors take into account typical distortions introduced by the block-based coding approaches like HEVC codec. The underlying processing used for the quality assessment considers the blockiness caused by the boundaries of each coded block and the blurring as a lack of spatial details. The correlation between the NR quality metrics and the well-known and most widely used objective metric, the Video Quality Model (VQM), is performed to validate the quality prediction accuracy based on the provided scores. The Pearson correlation coefficients obtained stand for promising results for different types of videos.

Keywords—High Efficiency Video Coding, No-Reference metrics, Quality of Experience, Video Quality Assessment.

1. Introduction

In addition to traditional Quality of Service (QoS), Quality of Experience (QoE) poses a real challenge for Internet service providers, audiovisual services, broadcasters, and new Over-The-Top (OTT) services. The leading operators have to solve the problem of accurate QoE prediction since the end-user satisfaction is a real added value in the market competition. QoE tools should be proactive and provide innovative solutions that are well adapted for new audiovisual technologies. Therefore, objective audiovisual metrics are frequently dedicated to monitoring, troubleshooting, investigating, and setting benchmarks of content applications working in real-time or off-line.

To advance the field of video quality assessment, Video Quality Experts Group (VQEG) performs subjective video quality experiments, validates objective video quality models, and collaboratively develops new techniques. VQEG proposed to monitor audio visual quality by Key Performance Indicators (KPI), which are able to isolate and focus investigation, set-up algorithms, increase the monitoring period, and guarantee good prediction of video quality. It is known that, depending on the technologies used in audiovisual services, the impact of QoE can change completely. So, based on that proposed concept, it is possible to select the best algorithms and activate or switch off features in a default audiovisual perceived list. The scores are separated for each algorithm and preselected before the testing phase. Then, each artifact KPI can be analyzed by working on the spatially and/or temporally perceived axes [1].

The proposed concept is an interesting approach because it can detect the artifacts present in videos, as well as predict the quality as described by consumers. In realistic situations, when video quality decreases in audiovisual services, customers can call a helpline to describe the annoyance and visibility of the defects or degradations in order to describe the outage. In general, they are not required to provide a Mean Opinion Score (MOS). As such, the concept is completely in phase with user experience. There are many possible reasons for video disturbance, and they can arise at any point along the video chain transmission (filming stage to end-user stage).

VQEG experiments were carried out over several steps with experimental set-ups for concept verification. The impairments included in the experiments were limited to MPEG-2 and H.264. Nevertheless, in year 2013, the first version of the High Efficiency Video Coding (HEVC) standard was completed, approved, and published. HEVC is a video compression standard, a successor to H.264/MPEG-4 AVC (Advanced Video Coding), which was jointly developed by the ISO/IEC JTC 1/SC 29/WG 11 Moving Picture Experts Group (MPEG) and ITU-T SG16/Q.6 Video Coding Experts Group (VCEG) as ISO/IEC 23008-2 MPEG-H Part 2 and ITU-T H.265 [2], [3].

In this paper, the experiments carried out over several steps with an HEVC experimental set-up for the proposed concept verification are presented.

The remainder of this paper is structured as follows. Section 2 is devoted to the state-of-the-art background. Section 3 discusses NR video quality assessment. Section 4 presents objective video quality methods. Section 5 analyses results on KPI. Section 6 discusses further work and summarizes the paper.

2. Related Works

This section presents brief survey of current NR approaches for standardized models together with their limitations. Most of the models in ITU-T recommendations were val-
idated on video databases that used one of the following hypotheses:

- frame freezes lasting up to 2 s,
- no degradation at the beginning or at the end of the video sequence; no skipped frames,
- clean video reference (no spatial or temporal distortions),
- minimum delay supported between video reference and video (sometimes with constant delay),
- up or down-scaling operations not always taken into account [4].

As mentioned earlier, most quality models are based on measuring common artifacts/KPI, such as blur, blocking, and jerkiness, for producing a prediction of the MOS. Consequently, the majority of the algorithms generating a predicted MOS show a mix of blur, blocking, and jerkiness metrics. The weighting between each KPI could be a simple mathematical function. If one of the KPIs is not correct, the global predictive score is completely wrong. Other KPIs mentioned by VQEG are usually not taken into account (exposure time distortion, noise, block loss, freezing, slicing, etc.) in predicting MOS [4]. ITU-T has been working on similar distortions for many years [5]. However, only for Full-Reference (FR) and Reduced-Reference (RR) approaches. The history of the ITU-T Recommendations for video quality metrics is shown in Table 1. Table 2 shows a synthesis of the set of standardized metrics that are based on video signals [4]. As can be noticed from both tables, there is a lack of developments for the NR approach.

Table 1

<table>
<thead>
<tr>
<th>Model type</th>
<th>Format</th>
<th>Recommendation</th>
<th>Year</th>
</tr>
</thead>
<tbody>
<tr>
<td>RR</td>
<td>SD</td>
<td>J.249 [9]</td>
<td>2010</td>
</tr>
</tbody>
</table>

In a related research, Gustafsson et al. [14] addressed the problem of measuring multimedia quality in mobile networks with an objective parametric model [4]. Closely related work are ongoing standardization activities at ITU-T SG12 on models for multimedia and Internet Protocol Television (IPTV) based on bit-stream information. SG12 is currently working on models for IPTV. Q.14/12 is responsible for these projects, provisionally known as non-intrusive parametric model for assessment of performance of multimedia streaming (P.NAMS) and non-intrusive bit-stream model for assessment of performance of multimedia streaming (P.NBAMS) [4]. P.NAMS uses packet-header information (e.g., from IP through MPEG2-TS), while P.NBAMS also uses payload information, i.e., coded bit-stream [15]. However, this work focuses on the overall quality (in MOS units), while the proposed concept is focused on KPIs [4].

Most of the recommended models are based on global quality evaluation of video sequences, as in the P.NAMS and P.NBAMS projects. The predictive score is correlated to subjective scores obtained with global evaluation methodologies (SAMVIQ, DSCQS, ACR, etc.). Generally, the duration of video sequences is limited to 10 or 15 s in order to avoid a forgiveness effect (the observer is unable to score the video properly after 30 s and may give more weight to artifacts occurring at the end of the sequence). When one model is deployed for monitoring video services, the global scores are provided for fixed temporal windows and without any acknowledgement of the previous scores [4].

Recently, the interest is oriented toward the HEVC standard, which has proved high efficiency compared to its predecessors. Several tools are introduced in the coding process, such as the increasing number of intra prediction modes and the frequent use of inter coded pictures within a closed Group Of Pictures (GOP). These characteristics ensure an important coding gain relative to the encoding parameters but in the other hand, the complex structure of picture division and the new configurations’ models can be the source of certain artifacts. However, very limited works concern the quality assessment approaches for HEVC compression. In particular, the coding parameters and the impact of network losses on the decoder side were investigated [16]. The distortions of HEVC videos are more significant than H.264 videos. The proposed NR distortion measure exploits the spectral densities between the frames and precisely, the energy variation in the temporal domain for each coding unit.

One can bear in mind that FR measures are in general not applicable as the reference content might be not available. In the same vein, the bitstream features were selected to estimate the perceptual quality, including the different prediction modes and statistics of the motion vector [17]. In this method the measures are predicted in a NR manner.

Table 2

<table>
<thead>
<tr>
<th>Resolution</th>
<th>Type of ITU-T model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>FR</td>
</tr>
<tr>
<td>HDTV</td>
<td>J.341 [10]</td>
</tr>
<tr>
<td>SDTV</td>
<td>J.144 [6]</td>
</tr>
</tbody>
</table>
The quality monitoring becomes primordial in communication and broadcasting environments for improving the end user’s QoE [18]. A NR Peak Signal-to-Noise Ratio (PSNR) estimation was proposed for such a model [19]. Distributions of transform coefficients are considered based on the quad-tree coding structure and the distortion model was derived according to the coding unit depth level. The concept of QoE in [20] is used for a practical recognition problem for video transmitted over a network link, where subjective satisfaction of the user is imperative. This latter requires achieving specific functionalities such as even detection and object recognition. The proposed methods measure the usefulness of degraded quality video and the solutions have been proposed to optimize the network QoS parameters.

Designing algorithms for video quality assessment requires a consistent dataset of coded video sequences. For the case of HEVC it is a key factor for an effective performance evaluation of developed metrics, to take advantage of a publicly available database, which includes several compressed versions of different sequences. In [21] Full-Reference measurements are provided with a large database of FULL-HD HEVC encoded videos based on a variety of HEVC compression characteristics.

A variety of NR quality estimation methods exist for the AVC videos but on the other side, widely used examples such pixel-based approaches are still not applied or tested for the HEVC compressed videos.

3. No-Reference Video Quality Assessment

In this section, NR measurement techniques in the spatial domain for two KPI are proposed: blur and blockiness. Assuming that we do not own a knowledge and assumptions of the original content or the distortion process of the HEVC compression. In fact, the NR pixel-based approach for measuring artifacts of the visual quality is proposed by considering a given model of degradation to investigate the performance of the mentioned metrics.

3.1. No-Reference Blockiness Metric

The same approach is used for calculating the blockiness artifact published in [22]. It is calculated locally for each coding block. Absolute differences in pixel luminance were calculated separately for intra-pairs, represented by neighboring pixels from a single coding block, and inter-pairs, represented by pixels from neighboring blocks. A ratio between the total values of intra- and inter-differences is calculated over the entire video frame. For a real time application the metric should be calculated over a time window (the number of video frames). Mean value for the window represents a blockiness level. For the purposes of the experiment the window size was equal to the sequence length (10 s). It was verified that the level of the blockiness artifact does not change significantly over time within the same video scene. Thus, any other window size or different method for temporal pooling would yield similar results.

3.2. No-Reference Blur Metric

The blurred image in compression techniques appears when high spatial frequency components of the image spectrum are truncated. For instance, possible reasons of blurring can be out-of-focus capturing or relative motion between the camera and the captured object. Besides, high compression performance can introduce blur when processing the data of images’ sequence. Perceptually, the blur artifact appears along edges and textured regions. In this work, the width of the edges is measured in order to characterize smoothing blur effect [23]. First, the Sobel filter as an edge detector is applied to find the gradient of the image. It is obvious that below a certain threshold, blur remains as just noticeable and visually unperceived. According to that threshold, the pixels being the part of the edges are differentiated. Then the width of an edge is measured, depending on its growth direction (left or right). Finally, the global blur value is obtained by averaging over all edges of the whole image.

4. Objective Video Quality Methods

Huge variety of proposed works concerning the video quality measurement use the objective metrics such as the simplest and commonly used ones: the PSNR and Mean-Squared Error (MSE). But in general, it is not ensured that error visibility would always the appearance of quality artifacts for most of distortions. Assuming that the structural information is highly captured from the viewing field by the human visual system, extracting this kind of information provides a good estimation of the perceived distortion. Therefore, the Structural Similarity (SSIM) has been used recently to characterize complex structured signals [24]. However, the different types of video coding and transmission systems require a more general model that covers a wide range of quality degradations. In fact an extensive objective and subjective tests should be performed to provide an effective perceptual measurement. The Video Quality Model (VQM) was indeed proposed by the Institute for Telecommunication Science (ITS) [25] and standardized by the American National Standards Institute (ANSI). It was further included in Draft Recommendations from ITU-T SG9. The VQM has proved a good performance for measuring perceptual effects of different types of video impairments such as blurring, jerkiness and block distortion. The calculation of VQM taking as input the original and processed videos follows these main steps:

- calibrate the processed video with respect to the original sequence by estimating and correcting the spatial-temporal shifts, as well as adjusting the contrast and brightness,
• extract a set of quality features to characterize perceptual changes from particular spatial-temporal regions in the video stream; for instance, in the chrominance, temporal and spatial properties,
• compare the extracted features form the processed video with those of the original sources,
• conclude the VQM value using a linear combination of the obtained parameters.

From the described functions, it makes sense that the VQM has a high correlation with subjective scores, which makes us believe that using it as a reference metric would provide accurate testing results.

5. Experiments and Results

In order to effectively evaluate the video quality based on HEVC compressions, the dataset of the project developed by the Joint Effort Group (JEG) of Video Quality Experts Group (VQEG) is used [21]. It presents a large-scale database of HEVC coded videos for researchers involved in designing hybrid quality metrics. Different encoding parameters were performed on ten sequences representing different characteristics. Among interesting benefits of the mentioned dataset, objective quality measurements are provided at frame-level granularity. This database is exploited by applying the NR metrics of blur and blockiness. It is primordial to investigate the accuracy of these metrics for the HEVC distortions and make useful interpretations about the specificities of the target approach.

5.1. Selected Compression Parameters

The performance of the quality metrics is investigated based on a diverse set of encoding parameters. Table 3 presents the retained HEVC configurations in order to carry out tests over an increasing data compression. The distortion is intrinsically related to the following values selected from the adopted database [21].

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>WIDTH</td>
<td>1280</td>
</tr>
<tr>
<td>GOPTYPESIZE</td>
<td>GOP8</td>
</tr>
<tr>
<td>RATECONTROL_QP</td>
<td>26, 32, 38, 46</td>
</tr>
<tr>
<td>RATECONTROL_FRAME_mbit/s</td>
<td>1, 2, 4, 8, 16</td>
</tr>
<tr>
<td>REFRESH</td>
<td>1</td>
</tr>
<tr>
<td>INTRAPERIOD</td>
<td>16</td>
</tr>
<tr>
<td>SLICEARGUMENT</td>
<td>0</td>
</tr>
</tbody>
</table>

The resolution of the ten original sources is 1280 × 720 pixels. The authors take into account all available fixed QP values as it represents a basic distortion source along with the frame rate control. The refresh number corresponds to the decoding refresh type, to apply a non-IDR clean random access point. This encoder option allows the use of an open GOP. The slicing value signifies one slice per frame. As a result, 90 processed video sequences are generated based on the above parameters. The prior-knowledge of these settings is not considered in developing the NR metrics and authors just provide it for a precise description of the compression rate and consequently the distortion strength.

5.2. Results and Analysis

Table 4 displays the results of the applied metrics on the ten processed videos. For each sequence, the Pearson correlation coefficient is used to validate the performance of the blur and blockiness measurements relative to the VQM values, offered by the JEG project for each encoded sequence according to the given parameters. From the shown results, the efficiency of the blur metric is confirmed for each source which means that the distorted edges are well predicted, providing high correlation values. It is further clear that the blockiness metric works well for the majority of the sources.

<table>
<thead>
<tr>
<th>Source</th>
<th>Blockiness</th>
<th>Blur</th>
</tr>
</thead>
<tbody>
<tr>
<td>src01</td>
<td>-0.67</td>
<td>1.00</td>
</tr>
<tr>
<td>src02</td>
<td>-0.97</td>
<td>0.91</td>
</tr>
<tr>
<td>src03</td>
<td>-0.97</td>
<td>0.96</td>
</tr>
<tr>
<td>src04</td>
<td>-0.77</td>
<td>0.99</td>
</tr>
<tr>
<td>src05</td>
<td>-0.87</td>
<td>0.99</td>
</tr>
<tr>
<td>src06</td>
<td>-0.57</td>
<td>0.91</td>
</tr>
<tr>
<td>src07</td>
<td>-0.96</td>
<td>0.96</td>
</tr>
<tr>
<td>src08</td>
<td>-0.95</td>
<td>0.92</td>
</tr>
<tr>
<td>src09</td>
<td>-0.36</td>
<td>0.95</td>
</tr>
<tr>
<td>src10</td>
<td>0.69</td>
<td>0.99</td>
</tr>
</tbody>
</table>

The authors mention here that the origin of the negative scores is caused by the metric’s construction, as increasing the compression rate corresponds to lower values of blockiness and vice versa. However, the correlation tends to drop for the case of src09 due to the complex nature of the motion and spatial activity in the video. Src09 consists of several combined shots separated with a black-pixels frame. Besides, the positive correlation of the src10 means that the trends of values are opposite to the expected ones.

The scatter plots in Figs. 1 and 2, representing the blockiness and blur metrics for all sequences, respectively, reveal a partial success even the measures are convincing for each source separately. The global correlation coefficient of blockiness is 0.55 whilst 0.23 for blur, which gives rise to useful interpretations for a more complete evaluation ap-
The temporal and spatial inconsistency could be incorporated to overcome such problems.

6. Conclusion and Further Work

In case of pixel-based NR methods, an accurate model which combines different kind of artifacts would generate an estimation of the perceptual quality using weighting factors. The strength of weights, which could be determined by a regression analysis, is computed with respect to a particular single metric. This latter is combined to another distortion measure, based on a linear or non-linear model. Furthermore, even the VQM measurement combines several features and represents with a certain precision perceptual characteristics, implicating subjective scores in the assessment process still more effective. For instance, correctness functions such as sigmoid model, can be applied on the predicted measures according to the subjective evaluation as it requires parameters’ estimation.

The HEVC specificities as the highly flexible quad-tree structure and effective prediction tools allow an accurate exploitation of the video content in addition to the high compression performance. Assessing quality of HEVC processed videos for different types of distortions require sophisticated techniques for a successful NR approach. In this work, the proposed metrics as a basic step to establish a completing framework of quality assessment are analyzed, taking into consideration particular aspects introduced in this new codec.

Acknowledgments

The work was co-financed by The Polish National Centre for Research and Development (NCBR), as a part of the EUREKA Project no. C 2012/1-5 MITSU and contract number 11.11.230.018. This research was further supported by the Bureau For Academic Recognition and International Exchange.

References


Kais Rouis received his B.Sc. in Computer Sciences in 2009 and his M.Sc. in Intelligent and Communicating Systems in 2011. He is currently a Ph.D. student at the National Engineering School of Tunis at the University of Tunis El Manar. His work revolves around image and video coding and wireless communication techniques.

E-mail: kais.rouis.phd@ieee.org
Innov'COM Lab, Sup’Com University of Carthage
Ecole Nationale d’Ingénieurs de Tunis Université de Tunis El Manar
Tunisia

Mikolaj Leszczuk started his professional career in 1996 at Comarch SA as manager of the Multimedia Technology Department, and then at Comarch Multimedia as the CEO. Since 1999 has been employed at the AGH Department of Telecommunications. In 2000 he moved to Spain for a four-month scholarship at the Universidad Carlos III de Madrid. After returning to Poland, he was employed at the Department of Telecommunications as a research and teaching assistant, and in 2006, he successfully defended his doctoral dissertation as an assistant professor. His current research interests are focused on multimedia data analysis and processing systems, with particular emphasis on QoE. He has participated more than 20 major research projects, including FP4, FP5, FP6, FP7, Horizon 2020, OPiE, Culture 2000, PHARE, eContent+, and Eureka!. Between 2009 and 2014, he was the administrator of the major international INDIRECT research project, dealing with solutions for intelligent surveillance and automatic detection of suspicious behavior and violence in urban environments. He is a member of VQEG, IEEE, and GAMA. He (co-)authored approximately 130 scientific publications.
E-mail: leszczuk@agh.edu.pl
AGH University of Science and Technology, Department of Telecommunications
Mickiewicza av. 30 30-059 Krakow, Poland

Lucjan Janowski is an assistant professor with the Department of Telecommunications, AGH University of Science and Technology. He received his Ph.D. degree in telecommunications in 2006 from the AGH. In 2007, he worked in a postdoctoral position at the Centre National de la Recherche Scientifique (CNRS), LAAS (Laboratory for Analysis and Architecture of Systems of CNRS) in France, where he prepared both malicious traffic analysis and anomaly detection algorithms. In 2010–2011, he spent half a year in a postdoctoral position at the University of Geneva, working on quality of experience (QoE) for health applications. In 2014–2015, he spent half a year in a postdoctoral position at The Telecommunications Research Center Vienna (FTW), working on quality of experience for IPTV customers. His main interests are statistics and probabilistic modeling of subjects and subjective rates used in QoE evaluation.
E-mail: janowski@kt.agh.edu.pl
AGH University of Science and Technology, Department of Telecommunications
Mickiewicza av. 30 30-059 Krakow, Poland

Study of No-Reference Video Quality Metrics for HEVC Compression

Kais Rouis, Mikolaj Leszczuk, Lucjan Janowski, Zdzislaw Papir, and Jamal Bel Hadj Tahar

**Zdzisław Papir** is professor and a deputy chair at Department of Telecommunications, AGH University of Science and Technology. Between 1999–2006 he was a guest co-editor for IEEE Communications Magazine responsible for the Broadband Access Series. He has been participating in several R&D IST European projects being responsible for network performance evaluation and quality assessment of communication services. He has also been appointed as an ICT expert by the European Commission. His current research interests include modeling of telecommunication networks/services and measuring quality of experience.

E-mail: papir@kt.agh.edu.pl  
AGH University of Science and Technology  
Department of Telecommunications  
Mickiewicza av. 30  
30-059 Krakow, Poland

**Jamal Bel Hadj Tahar** received his Ph.D. in 1993 at the Polytechnic Institute of Grenoble, INPG, and became master conference in 2011 at the Higher School of Communications of Tunis, on Information and communication technologies TIC. Currently, he is professor at the National Engineering School of Sousse and responsible for the training of masters research in telecommunications engineering. His research focuses on two areas one is one wireless communication techniques and systems while the other is optical systems.

E-mail: belhadj.tahar@supcom.rnu.tn  
Innov‘COM Lab, Sup‘Com  
University of Carthage  
Ecole Nationale d’Ingénieurs de Tunis  
Université de Tunis El Manar  
Tunisia