Evaluation of objective video quality assessment methods on video sequences with different spatial and temporal activity encoded at different spatial resolutions

Original Scientific Paper

Jelena Vlaović
J. J. Strossmayer University of Osijek,
Faculty of Electrical Engineering, Computer Science and Information Technology Osijek,
Kneza Trpimira 2b, Osijek, Croatia
jvlaovic@ferit.hr

Snježana Rimac-Drlje
J. J. Strossmayer University of Osijek,
Faculty of Electrical Engineering, Computer Science and Information Technology Osijek,
Kneza Trpimira 2b, Osijek, Croatia
rimac@ferit.hr

Drago Žagar
J. J. Strossmayer University of Osijek,
Faculty of Electrical Engineering, Computer Science and Information Technology Osijek,
Kneza Trpimira 2b, Osijek, Croatia
dzagar@ferit.hr

Mario Vranješ
J. J. Strossmayer University of Osijek,
Faculty of Electrical Engineering, Computer Science and Information Technology Osijek,
Kneza Trpimira 2b, Osijek, Croatia
mvranjes@ferit.hr

Abstract *– With the development of Video on Demand applications due to the availability of high-speed internet access, adaptive streaming algorithms have been developing and improving. The focus is on improving the user's Quality of Experience (QoE) and taking it into account as one of the parameters for the adaptation algorithm. Users often experience changing network conditions, so the goal is to ensure stable video playback with a satisfying QoE level. Although subjective Video Quality Assessment (VQA) methods provide more accurate results regarding user's QoE, objective VQA methods cost less and are less time-consuming. In this article, nine different objective VQA methods are compared on a large set of video sequences with various spatial and temporal activities. VQA methods used in this analysis are: Peak Signal-to-Noise Ratio (PSNR), Structural Similarity Index (SSIM), MultiScale Structural Similarity Index (MS-SSIM), Video Quality Metric (VQM), Mean Sum of Differences (DELTA), Mean Sum of Absolute Differences (MSAD), Mean Squared Error (MSE), Netflix Video Multimethod Assessment Fusion (Netflix VMAF) and Visual Signal-to-Noise Ratio (VSNR). The video sequences used for testing purposes were encoded according to H.264/AVC with twelve different target coding bitrates, at three different spatial resolutions (resulting in a total of 190 sequences). In addition to objective quality assessment, subjective quality assessment was performed for these sequences. All results acquired by objective VQA methods have been compared with subjective Mean Opinion Score (MOS) results using Pearson Linear Correlation Coefficient (PLCC). Measurement results obtained on a large set of video sequences with different spatial resolutions show that VQA methods like SSIM and VQM correlate better with MOS results compared to PSNR, SSIM, VSNR, DELTA, MSE, VMAF and MSAD. However, the PLCC results for SSIM and VQM are too low (0.7799 and 0.7734, respectively), for the usage of these methods in streaming services instead of subjective testing. These results suggest that more efficient VQA methods should be developed to be used in streaming testing procedures as well as to support the video segmentation process. Furthermore, when comparing results obtained for different spatial resolutions, it can be concluded that the quality of video sequences encoded at lower spatial resolutions in cases of lower target coding bitrate is higher compared to the quality of video sequences encoded at higher spatial resolutions at the same target coding bitrate, particularly when video sequences with higher spatial and temporal information are used.*

Keywords *– spatial activity, spatial resolution, temporal activity, video streaming, video quality assessment*

1. INTRODUCTION

The major increase in Internet accessibility over the last decade has resulted in high demand of different multimedia content availability that is subject to changing network conditions. The prediction is that Internet video traffic will increase from the current 105 EB per month to 240 EB per month by 2022 and consumer Video on Demand (VoD) traffic will nearly double by 2022 [1]. Various VoD applications and adaptive bitrate (ABR) streaming algorithms have been developed, which has solved some of the problems with adapting to changing network conditions, but there is still room for improvement. Nowadays ABR algorithms take into account various parameters to adapt to changing network conditions like variations in bandwidth, video segment size, and buffer fullness. Quality of video sequences that are played back to the user should be tested so that the amelioration of ABR algorithms could be verified.

Regardless of the network conditions, users request the highest possible Quality of Experience (QoE), which is still a challenging task. As opposed to QoE, the parameters in the Quality of Service (QoS) specification are selected depending on the type of application and are related to technical aspects. QoS was used to quantify the quality in multimedia services for many years, but the question arose whether the technical parameters of the network correspond to the user's perception of video quality, because QoS does not take into account user's subjectivity. Currently a large amount of video streaming services, including social media, use objective Video Quality Assessment (VQA) methods to measure and control the video quality they deliver to end-users. Objective VQA methods can be also used as inputs to QoE predictors [2-3]. Still the correlation between objective and subjective VQA methods should be more thoroughly analyzed.

Human-Computer Interaction (HCI) researchers were first to point out that QoE takes into account emotions, relationships, context, and expectations [4]. The QoE is defined by ITU-T as the user's subjective acceptance of service [5]. The QoE is also defined as a measure of user's satisfaction or bother with the service [6] which differs from subjective VQA, which is focused on predicting user's responses to visible distortions. Authors in [7] state that QoE is a multidisciplinary area based on engineering and cognitive science, economics, and social psychology. The QoE is affected by various factors that can be divided into human, context and system. Human factors include user preferences, different sensorial and cognitive processes. Context factors are economic and social aspects, as well as time and space in which a service is used. System factors are software (SW) and hardware (HW) limitations of electronic devices that are being used [8]. To ensure the highest possible QoE level, the playback of video content should be seamless, without delays and rebuffering events.

Developers of novel services are taking into account subjective and objective VQA methods that are used to quantify user's QoE [9]. The reliability of an objective VQA method is usually verified and quantified by comparing its results to the results of subjective testing. Objective VQA methods are still a vital part of service testing because subjective testing is time-consuming, expensive and cannot provide the measurement of video quality fast enough [10]. Thus, analysis of objective VQA methods on various video sequences with different spatial and temporal activity and with different spatial and temporal resolution is important to further optimize ABR algorithms testing procedures. The server side in streaming services stores video sequences in segments encoded with various target coding bitrates and with different spatial resolutions. The client side in streaming services connects the received video segments and prepares the video sequence for playback. Depending on the playback device, spatial resolutions of all video segments have to be adjusted. The idea of this article is to analyze the efficiency of VQA methods on the server side, in order to improve the selection of proper target coding bitrates and spatial resolutions with respect to network conditions, but also the spatial and temporal activity of a given video sequence.

For this article ten different video sequences with different spatial and temporal activity were encoded with various parameters regarding target coding bitrates and spatial resolutions. Nine VQA methods have been tested on encoded video sequences after they were scaled to Full HD spatial resolution. All results were then compared to results acquired from subjective testing.

This article is constructed as follows: related work is given in section 2. Section 3 gives information about test setup, selected coding parameters, target coding bitrates and calculated values of spatial and temporal information of selected video sequences. Section 4 that includes results and discussion is followed by the conclusion.

2. INTRODUCTION

Based on the final video quality score of distorted video sequence which can be determined by a computer or a user, VQA methods can be divided into objective and subjective methods. Objective VQA methods can further be divided into:

- full-reference (FR) objective VQA methods: require full reference video sequence for analysis [10]
- reduced-reference (RR) objective VQA methods: require only a number of features from reference video sequence for evaluation of distorted video sequence quality [11, 12],
- no-reference (NR) objective VQA methods: predict quality without using any reference information and do not need any information about the original video [13, 14].

In this article several FR objective VQA methods are analyzed: Mean Sum of Differences (DELTA), Mean Sum of Absolute Differences (MSAD), Mean Squared Error (MSE), Peak Signal-to-Noise Ratio (PSNR) [15], Structural Similarity index (SSIM) [16], MultiScale Structural Similarity Index (MS-SSIM) index [17], Visual Signal-to -noise Ratio (VSNR) [18], Netflix Video Multimethod Assessment Fusion (Netflix VMAF) [15], RR metric Video Quality Metric (VQM) [19].

DELTA uses the mean difference of the color components in the correspondent points of image for quality assessment. MSAD uses the mean absolute difference of the color components in the correspondent points of image for quality assessment [20].

MSE, DELTA, MSAD and PSNR are VQA metrics that do not take into account Human Visual System (HVS) characteristics. MSE and PSNR are based on statistical processing of the mean value of the square of the pixel difference, thus they do not achieve a high correlation with subjective VQA scores. Metrics that achieve better correlation with subjective VQA scores are SSIM and MS-SSIM because they take into account HVS which is suitable for acquiring structural information. MS-SSIM uses structural distortion calculation but on multiple scales. SSIM and MS-SSIM compare structural features extracted from the video sequences as opposed to comparing the pixel values, [15, 16,18].

VSNR uses low-level properties of HVS of contrast sensitivity and visual masking and mid-level properties of global precedence in the decision-making process. It takes into account the near-threshold and supra-threshold properties of HVS to determine the visual accuracy of the analyzed video. Both the near-threshold and supra-threshold properties are presented as Euclidean distances. VSNR is the linear sum of those distances. The characteristic of VSNR are efficiency regarding the memory requirements and computational complexity [18].

Netflix VMAF is a metric that correlates well with Mean Opinion Score (MOS) due to the fact that it combines multiple elementary video quality features. The estimation model was trained using a large MOS dataset. VMAF focuses on quality degradation resulted from rescaling and compression. The perceived quality score is calculated by computing scores from various VQA algorithms and combining them using a support vector machine [15].

VQM metric uses the Discrete Cosine Transform (DCT) to calculate the distortion in the video sequences. The value of VQM increases with the level of degradation in the analyzed video sequence [19].

Before analyzing the correlation between different objective and subjective VQA methods, it is necessary to investigate what affects the QoE of end-users the most. The QoS used for analyzing various multimedia services depends on the type of service for which the parameters are analyzed. Nowadays it is well known that user's expectation of delivered video quality often do not meet the information acquired from measured network parameters. It is useful to gather information about network parameters to get the knowledge of certain events in the network, but that information will not present the user's experience of those events [21].

Video sequence processing procedure consists of recording or generating the video sequence, coding, compression, transferring, decoding, and reproducing. During those processing procedures different factors can influence both the QoE and the QoS and vary the user's experience. Considering that various technical and non-technical factors influence the user's experience, optimization and measurement of video quality is an elaborate task [22].

As previously mentioned, there are different technical factors that can decrease the quality of the video sequence in processing procedure. Users generally perceive high contrast video sequences as video sequences with better video quality, whereas video sequences with low brightness, contrast and sharpness as video sequences with low video quality [23]. ITU-T states that subjective testing conducted with a group of at least 15 individuals is the most precise VQA method [24]. For researchers to conduct such testing, ITU-T provides regulation concerning viewing conditions, evaluation procedures, criteria for selecting users and materials and methods of data analysis [25].

Taking into account that subjective VQA methods give more accurate results, it is important to state that the QoE is lower if there is substantial initial delay or there is deterioration in the first temporal segment of the video sequence. Also the QoE is lower in the cases of up-switching between successive video quality levels compared to instantaneous up-switching between quality levels by more than one step. Authors in [26] state that quality level switching among video sequences with different spatial resolutions affects the QoE more than switching among video sequences with different temporal resolutions. Nevertheless, some authors point out the drawbacks of subjective testing like the fact that previous experiences of the participants can compromise the test results [27]. Also, subjective VQA results can be affected by the participants' preferences [28].

Authors in [26] state that results acquired by subjective VQA metric like PSNR do not always correlate with the results acquired from the subjective VQA methods.

Although authors in [18] state that objective VQA metrics like MSE and PSNR have low correlation with subjective testing because they do not take into account the properties of HVS, authors in [29] state that PSNR and SSIM have good correlation with subjective testing.

Authors in [30] compared PSNR and SSIM to MS-SSIM and VMAF. They state that PSNR has the worst results due to the fact that it does not consider perceptual information. SSIM performed somewhat better compared to PSNR especially in cases of images with various supra-threshold distortions. Considering the MS-SSIM has multiscale properties, it performed better than SSIM. VMAF performed the best, but only if Netflix dataset is used because it captures scaling compression artifacts and does not perform well with unseen distortions. Authors in [31, 32] state that MS-SSIM and VQM achieve roughly the same results.

Compared to related work, this article compares results of nine full reference VQA methods to the results of subjective testing using ten video sequences with different spatial and temporal information encoded on three spatial resolutions and various target coding bitrates. The idea is to use a larger set of video sequences to identify the best VQA method to be used in the future for optimizing video segmentation procedures at server side as well as testing ABR algorithms that use scaling of spatial resolutions.

3. TEST SETUP

Testing conducted for this article was done in order to compare different VQA methods and determine which of them gives the most similar results to those obtained by subjective testing when using video sequences with different spatial and temporal information encoded on different spatial resolutions, scenario often seen in streaming services. Ten various video sequences (often used in testing of ABR algorithms) were selected for testing purposes based on their spatial and temporal activity. All sequences were encoded at three spatial resolutions and various target coding bitrates, from 600 kbps to 12600 kbps. The VQA methods analyzed in this article are MSE, PSNR, VSNR, SSIM, MS-SSIM, DELTA, MSAD, VQM and Netflix VMAF.

Results were acquired using MSU Quality Measurement Tool [20]. All selected video sequences are available in FullHD spatial resolution with 25 fps. The core sequences used in this article are titled: BlueSky (BS), Chimera1102353 (C53), Station2 (S2), PedestrianArea (PA), Chimera1102347 (C47), CosmosLaundromat (CL), ElFuenteDance (ED), MeridianConversation (MC), Skateboarding (SK) and Soccer (SO). BS, S2, PA were obtained from the dataset published in [33]. Other video sequences were obtained from the dataset published in [34].

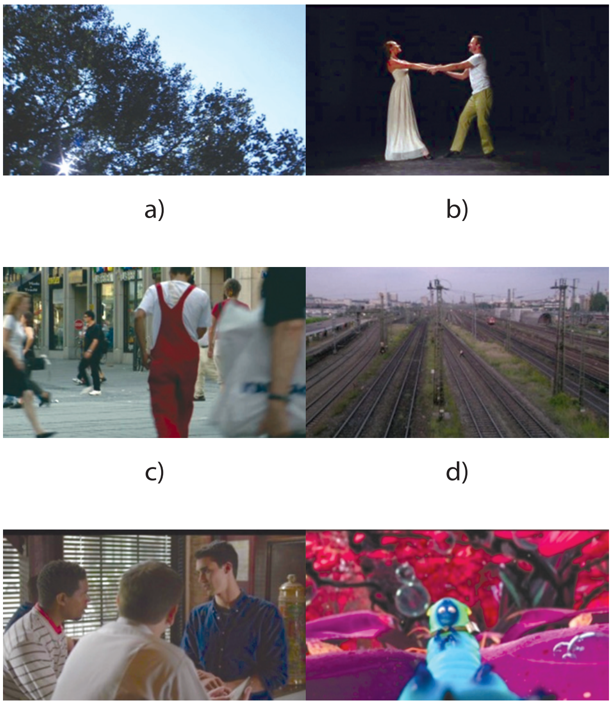

Fig.1. Frames from video sequences: a)Blue Sky (BS) b) Chimera1102353 (C53) c) PedestrianArea (PA) d) Station2 (S2) e) Chimera1102347 (C47) f) CosmosLaundromat (CL) g) ElFuenteDance (ED) h) MeridianConversation (MC) i) Skateboarding (SK) j) Soccer (SO)

Temporal and spatial activity parameters titled Temporal Perceptual Information (TI) and Spatial Perceptual Information (SI) were calculated based on [35] for Y color component of video sequences in YUV format. TI and SI values for all sequences are presented in Fig. 2. From Fig. 2. it can be seen that PA and S2 have similar SI but different TI. Sequences C53 and C47 have similar TI but different SI. The same stands for ED and SO. SO is the sequence with the highest SI, BS is the sequence with the highest TI, whereas C53 is the sequence with lowest SI, MC is the sequence with the lowest TI. This information of SI and TI shall be vital for the analysis of VQA methods.

All selected video sequences were encoded at three spatial resolutions: nHD (640x360), HD (1280x720) and FullHD (1920x1080), according to H.264/AVC video compression standard with coding parameters given in Tab.1, using open-source program called FFmpeg [36]. Spatial resolution downscaling was done using the medium preset and CRF 0 (lossless). FFmpeg uses the scale filter that changes the output sample aspect ratio.

During the coding process, the preset was set to slow because it presents a compromise between time needed for compression and its efficiency. When creating core video sequences, Constant Rate Factor (CRF) was set to zero because it ensures the best possible quality of the output video.

Other sequences encoded from core sequences were encoded using CRF of 23, which is the default value in FFmpeg. Since the CRF was used, the achieved target coding bitrate can vary i.e. be higher or lower than target coding bitrate, because CRF focuses on delivering the requested quality level by using the bitrate close to target coding bitrate. Video sequences with high SI and TI are expected to have the highest achieved coding bitrate for equal target coding bitrate.

All selected target coding bitrates are given in Tab. 2. There are in total 19 different combinations of spatial resolution and target coding bitrate listed in Tab. 2. All ten video sequences with different SI and TI were encoded at those 19 different combinations, thus there are 190 encoded video sequences.

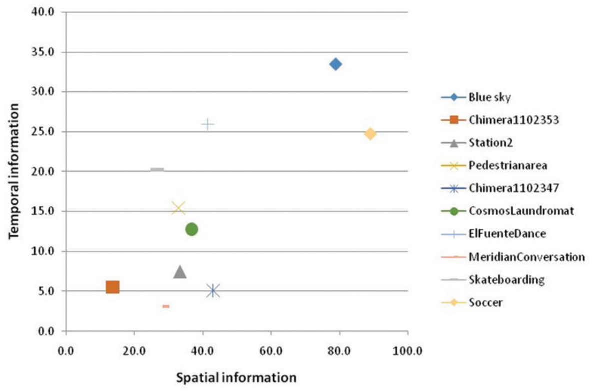

Fig.2. Spatial and temporal activity of video sequences used in our experiments

Tab.1. H.264/AVC coding parameters used in our experiments.

Coding parameter	Value
overall encoding strategy	2-pass variable bitrate
profile	high
coding	CABAC entropy coding
level	4.0
peak rate	1080p Superbit
quantiziercurve compression	0.9
minimum quantization	3

Tab.2. Target coding bitrates.

Spatial resolution	Target coding bitrate [kbps]
nHD	600; 800; 1000; 1400; 2400; 5400
HD	600; 1400; 2400; 3200; 4000; 5600; 9000
Full HD	600; 1400; 5400; 7200; 9000; 12600

In order to compare the quality of sequences with different spatial resolutions, before the objective and subjective quality evaluation, video sequences with nHD and HD resolution were up-scaled to Full HD. Bilinear interpolation was selected as scaling method due to its low complexity but overall satisfying results. The bilinear interpolation was also done using FFmpeg. In this way, we simulated conditions in which ABR may request from the server video segments with a lower resolution due to changes in network throughput, but the video sequence is always presented to a viewer with the same (the highest possible) resolution.

4. RESULTS AND DISCUSSION

Fig. 3. to Fig. 6. present test results for VQA methods PSNR, SSIM, VQM, and Netflix VMAF. All figures show measurement results for ten video sequences encoded at Full HD spatial resolution with 600, 1400, 5400, 7200, 9000, 12600 target coding bitrates. From Fig. 3 it can be seen that video sequences with higher SI and TI (Fig. 2.) have lower values of PSNR due to the fact that they are more complex to encode with the selected coding parameters. Furthermore, it can be concluded that video sequences with higher SI have lower PSNR values in cases when comparing two video sequences with similar TI like C53 and C47. Looking at values from S2 and PA video sequences, it can be concluded that in case of video sequences with similar SI, video sequence with higher TI has lower PSNR values. All video sequences have a value drop at the target coding bit rate of 600 kbps due to exceedingly low target coding bitrate for the analyzed spatial resolution. The lower the values of SI and TI are, the lower the drop in PSNR value is. Similar conclusions considering the relation of measurement results to SI and TI can be gathered from Fig. 4. to Fig. 6. The only main difference is the scale the VQA metric uses, thus the curves look more or less scattered. In addition to objective quality testing, subjective testing was performed for all 190 video sequences. The subjective VQA measurement was done in a controlled environment with 26 inexperienced viewers [37]. After conducting the experiment, MOS was calculated as a mean value of gathered results for each video sequence.

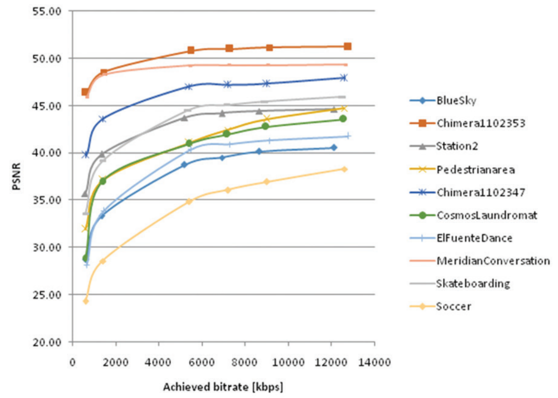

Fig.3. PSNR values for Full HD spatial resolution

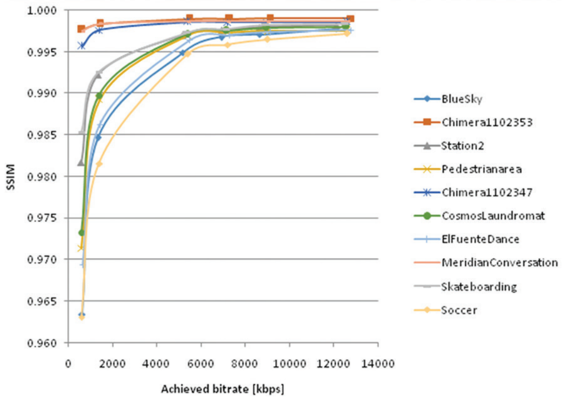

Fig.4. SSIM values for Full HD spatial resolution

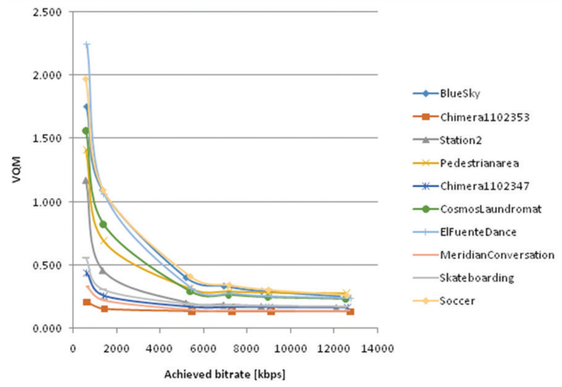

Fig.5. VQM values for Full HD spatial resolution

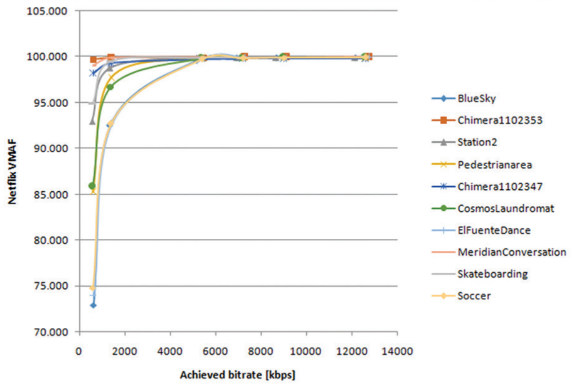

Fig.6. Netlix VMAF values for Full HD spatial resolution

Pearson Linear Coefficient (PLCC) was calculated to determine the correlation between measurement results from all objective VQA methods and subjective test results [38]. PLCC results have been calculated for every spatial resolution for every video sequence separately, then for all video sequences for certain spatial resolution, and for all video sequences for certain objective VQA method.

Results for PLCC for all video sequences at certain spatial resolution and for all video sequences in total for each objective VQA method are given in Tab. 3. PLCC results for nHD and HD have been calculated for the upscaled video sequences. In total, 190 video sequences have been analyzed for each objective VQA method. The results showed that the correlation between measurement results of subjective and objective VQA is, for most VQA methods, the lowest for video sequences encoded with HD resolution. This can be explained by the fact that the number of target coding bitrates used in the encoding process for this resolution was the largest.

In general, none of the metrics used achieved a very high PLCC value for all video sequences. However, SSIM and VQM have the highest PLCC (0.7794 and 0.7734, respectively), and they achieve overall the best results when different spatial resolutions are used. Although SSIM has the highest PLCC for all sequences, VQM achieves the best results for HD resolution, i.e. when a larger number of target coding bitrates is used.

Considering that SSIM and VQM take into account some of the HVS characteristics, they outperform PSNR, DELTA, MSAD, MSE and VSNR as can be seen from Tab. 3. The unexpectedly low PLCC results are obtained for Netflix VMAF, comparable to PSNR results, which has only a bit lower PLCC. Although Netflix VMAF combines multiple elementary video quality features and has been trained on streaming video sequences, it does not perform well on our experimental setup. Partially, this can be explained with different settings for subjective measurements used for VMAF development, i.e. model for VMAF is based on the assumption that the video sequences are presented in 1080p resolution TV display and that viewers are on the viewing distance of 3x the screen height (3H).

Tab.3. PLCC results for video sequences at nHD, HD and Full HD spatial resolutions.

Objective metric	Spatial resolution	PLCC
PSNR	360p	0.685044
	720p	0.644172
	1080p	0.715153
	All video sequences	0.708182
SSIM	360p	0.888921
	720p	0.715210
	1080p	0.831389
	All video sequences	0.779926
MS-SSIM	360p	0.718876
	720p	0.681271
	1080p	0.769829
	All video sequences	0.711381
VQM	360p	0.867630
	720p	0.755780
	1080p	0.831150
	All video sequences	0.773982
DELTA	360p	0.823293
	720p	0.634594
	1080p	0.830959
	All video sequences	0.649855
MSAD	360p	0.608980
	720p	0.605221
	1080p	0.684087
	All video sequences	0.699860
MSE	360p	0.726784
	720p	0.691223
	1080p	0.697680
	All video sequences	0.666198
Netflix VMAF	360p	0.670822
	720p	0.585973
	1080p	0.718319
	All video sequences	0.718851
VSNR	360p	0.543860
	720p	0.525684
	1080p	0.712838
	All video sequences	0.708905

In our experiment, viewers watched video sequences on a computer monitor at a distance of less than 3H. When analyzing the results for each video sequence it can be concluded that objective VQA method results obtained for video sequences with higher spatial and temporal information have lower PLCC which states for all VQA methods (Tab. 4. presents SSIM results for C53, CL and SO, but similar results were obtained for all VQA methods).

The lower PLCC values for video sequences with high SITI (a product of SI and TI values, SITI=SI.TI), are a result of spatial and temporal masking of the errors that occur due to a large number of details and fast motions in video sequences with high SI and TI. Objective methods compare video sequences frame by frame (full reference metrics) and do not take properly the effects of spatial and temporal masking into account, so they generally rate the quality more rigorously compared to human viewers for these fast and complex sequences. That can cause the metrics to overestimate the visible impairments and give lower objective VQA scores compared to subjective scores.

Taking into account that ABR algorithms use video segments with different spatial resolution, analysis has been done on results given for every spatial resolution separately. Fig.7. presents SSIM results for video sequences CL, C53 and SO for all three spatial resolutions. CL, C53 and SO are selected because C53 has the lowest SITI, SO has very high SITI and the highest SI, and SITI of CL is in the middle compared to C53 and SO (Fig. 2.). Although results are given for SSIM, all other metrics give similar results. From Fig. 7. it can be concluded that in general the SSIM values are higher when achieved bit rate is higher and that video sequences encoded at higher spatial resolution have higher SSIM values. Still, in cases when spatial and temporal activity of video sequence are too high to encode it on low target coding bitrate and with HD or Full HD spatial resolutions, SSIM values can be higher in the case of nHD spatial resolution. For example, in the case of the SO video sequence that has high spatial and temporal activity, SSIM values for the target bit rate of 600 kbps are lower for both HD and Full HD compared to nHD. In the case of CL that has a lower SI and TI at 600 kbps, SSIM value for HD spatial resolutions is higher than for Full HD spatial resolution. Taking into account SI and TI, it can be seen that the bit rate at which the overlap occurs is higher for video sequences with higher SI and TI. This conclusion can help with selecting the most suitable target coding bitrates for each spatial resolution, thus improving encoding and segmentation process when preparing video sequences for adaptive streaming. For the C53 sequence, which has the lowest SITI, there are no overlapping effects for nHD resolution. SSIM results for this resolution are lower than for HD and Full HD even for 600 kbps, because due low spatial and temporal activity of this video sequence, coder can successfully encode at higher spatial resolutions. It can be expected that overlapping for sequences with such low SITI occurs at even lower target coding bitrates. The MOS results for video sequences C53, CL and SO, given in Fig. 8, confirm that the overlapping of curves occurs. For CL and SO video sequences it can be seen that at lower target coding bitrates, MOS as well as SSIM can be higher for lower spatial resolution. The C53 sequence has much lower MOS results at nHD than at HD and Full HD resolutions, even at 600 kbps, because the loss of details caused by a decrease in spatial resolution cannot be masked due to the low spatial and temporal activity of the video content.

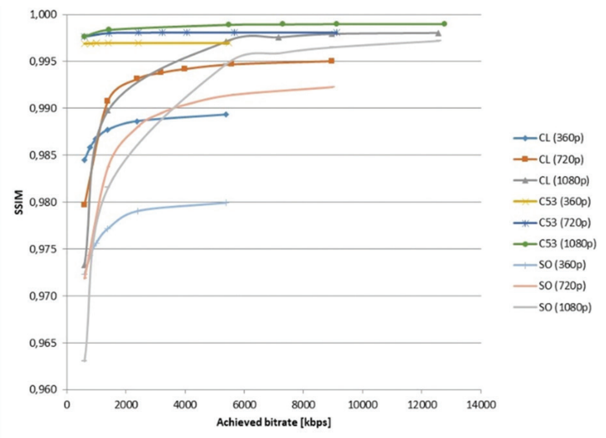

Fig.7. SSIM values for three spatial resolutions for video sequences Chimera1102353 (C53), CosmosLaundromat (CL) and Soccer (SO)

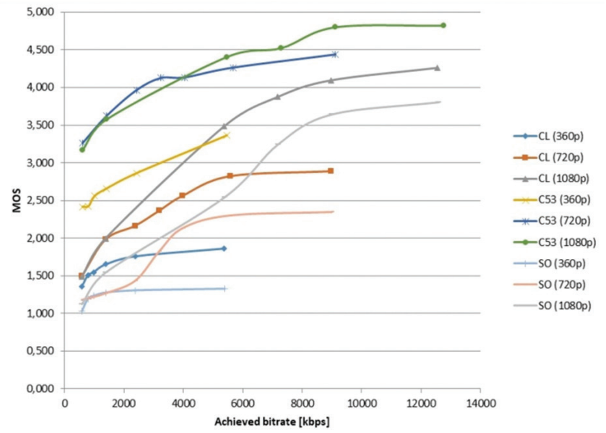

Fig.8. MOS values for three spatial resolutions for video sequences Chimera1102353 (C53), CosmosLaundromat (CL) and Soccer (SO)

Tab. 4. PLCC results for SSIM and SITI for Chimera1102353 (C53), CosmosLaundromat (CL) and Soccer (SO) video sequences.

Objective metric	Video sequence	SITI	PLCC
SSIM	Chimera1102353 (C53)	75.066	0.921
	CosmosLaundromat(CL)	470.342	0.842
	Soccer (SO)	2200.608	0.821

5. CONCLUSION

In order to improve datasets and the segmentation process for video streaming purposes in the future, different variants of encoded video sequences were used. In this paper, nine subjective VQA methods were analyzed using ten video sequences with different SI and TI that were encoded at three spatial resolutions and various target coding bitrates. From our results, video sequences with higher SI and TI have lower values of PSNR. When comparing the sequences with similar TI, it can be concluded that sequences with higher SI have lower PSNR values. Sequences with higher TI and similar SI have lower PSNR values. PSNR values drop considerably at coding bit rate of 600 kbps, especially for video sequences with higher SI and TI. Similar results are obtained for all analyzed objective VQA metrics. Results also show that in cases of higher SI and TI and low target coding bitrate, subjective VQA scores can be higher for lower spatial resolutions. The same conclusion can be made from MOS results. On selected dataset, SSIM achieves the best overall correlation to PLCC results calculated based on MOS thus it is the best in cases when video sequences are encoded with various spatial resolutions and various target coding bitrates. Video sequences with higher SI and TI have lower PLCC results due to spatial and temporal masking, which objective VQMs fail to capture well. VQM acquires the best results for HD resolution which was calculated for larger set of encoded sequences, though SSIM has the highest overall PLCC. Compared to PSNR, DELTA, MSAD, MSE and VSNR, SSIM and VQM have higher values of PLCC because they consider HVS characteristics. From MOS results and results obtained with objective VQA methods it can also be concluded that when sequences with high SI and TI encoded at low target coding bitrate are used, video quality can be higher in case of nHD spatial resolution compared to HD and Full HD spatial resolutions. This situation does not occur when video sequences with low SI and TI are used. In the future work those conclusions shall be used to shape parameters for new adaptive streaming algorithm and to propose a new dataset for adaptive streaming purposes.

6. ACKNOWLEDGMENT

The current archival periodical article is based on the conference presentation [37].

7. REFERENCES

[1] Cisco Visual Networking Index: Forecast and Methodology, 2016–2021, San Jose, CA, USA, Sep. 2017.

[2] C. G. Bampis and A. C. Bovik. (2018). “Feature-based prediction of streaming video QoE: Distortions rebuffering and memory.” Signal Processing: Image Communication. vol. 68, no. 7, pp. 218–228, Oct. 2018.

[3] C. G. Bampis, Z. Li, I. Katsavounidis, and A. C. Bovik, “Recurrent and dynamic models for predicting streaming video quality of experience,” IEEE Trans. Image Process., vol. 27, no. 7, pp. 3316–3331, Jul. 2018.

[4] Y. Chen, K. Wu, and Q. Zhang, “From QoS to QoE: A tutorial on video quality assessment,” IEEE Communications Surveys & Tutorials, vol. 17, pp. 1126–1165, October 2015.

[5] “ITU-T G.1080 Quality of experience requirements for IPTV services,” International Telecommunication Union, pp. 1-44, 2008.

[6] P. L. Callet, S. Moller, and A. Perkis, “Qualinet white paper on definitions of quality of experience (2012),” Eur. Netw. Qual. Exp. Multimedia Syst. Services (COST Action IC 1003), White Paper, 2013.

[7] G. Pibiri, C. Mc Goldrick, and M. Huggard, “Expected Quality of Service (eQoS) A network metric for capturing end-user experience,” IFIP Wirel. Days, November 2012.

[8] U. Reiter et al. “Factors influencing Quality of Experience,” in Quality of Experience, S. Moeller and A. Raake, Eds. London: Springer, 2014, pp. 55-72.

[9] H-J. Park and D-H. Har, “Subjective Image Quality Assessment based on Objective Image Quality Measurement Factors,” IEEE Trans. On Consumer Electron., vol. 57, no. 3, pp. 1176-1184, Aug. 2011.

[10] S. Chikkerur, V. Sundaram, M. Reisslein, and L. J. Karam, “Objective video quality assessment methods: A classification, review, and performance comparison,” IEEE Trans. Broadcast., vol. 57, no. 2, pp. 165–182, Jun. 2011.

[11] Z. Wang, L. Lu, and A. C. Bovik, “Video quality assessment based on structural distortion measurement,” Signal Process., Image Commun., vol. 19, no. 2, pp. 121–132, 2004.

[12] Z. Wang and E. P. Simoncelli, “Reduced-reference image quality assessment using a wavelet-domain natural image statistic model,” Electron. Imag., vol. 5666, pp. 149–159, Mar. 2005.

[13] A. Mittal, M. A. Saad, and A. C. Bovik, “A completely blind video integrity oracle,” IEEE Trans. Image Process., vol. 25, no. 1, pp. 289–300, Jan. 2016.

[14] Nighi, N. Aggarwal, “A review on Video Quality Assessment”, 2014 Recent Advances in Engineering and Computational Sciences (RAECS), Chandigarh, 2014, pp. 1-6, March 2014.

[15] R. Rassool, “VMAF reproducibility: Validating a percep-

tual practical video quality metric," 2017 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB), pp. 1-2, June 2017.

[16] Z. Wang, A. Bovik, and H. Sheikh, "Image Quality Assessment: From Error Visibility to Structural Similarity," IEEE Transactions on Image Processing, vol. 13, No. 4, Apr 2004.

[17] C. Yang, L. Zhao and Z. Liao, "Objective Quality Metric Based on Perception for Video," International Conference on Computer Engineering and Technology, pp. 20-23, January 2009.

[18] D. M. Chandler and S. S. Hemami, "VSNR: A wavelet-based visual signal-to-noise ratio for natural images", IEEE Transactions on Image Processing, vol. 16, pp. 2284-2298, August 2007.

[19] M. Pinson and S. Wolf, "A new standardized method for objectively measuring video quality," IEEE Trans. Broadcast., vol. 50, no. 3, pp. 312–322, Sep. 2004.

[20] D. Vatolin, M. Smirnov, A. Ratushnyak and V. Yoockin, www.compression.ru/video/quality_measure/video_measurement_tool.html (accessed: 2020)

[21] P. Orosz, T. Skopkó, Z. Nagy, P. Varga, and L. Gyimóthi, "A case study on correlating video QoS and QoE," IEEE Network Operations and Management Symposium (NOMS), pp. 1-5, May 2014.

[22] V. Deksnys, E. Sakalauskas and G. Činčikas, "New integral quality of TV service criterion construction based on quality of experience statistical estimation," 14th Biennial Baltic Electronic Conference (BEC), pp. 149–152, October 2014.

[23] R. Koshimura, Y. Ito and Y. Nomura, "Empirical study on clarification of relationship between QoS and QoE for Web services by path analysis," 3rd Global Conference on Consumer Electronics (GCCE), pp. 10-11, 2014, October 2014.

[24] "ITU-T P.10 Vocabulary for performance, quality of service and quality of experience.", International Telecommunication Union, pp. 1-22, 2017.

[25] "ITU-T P.912: Subjective video quality assessment methods for recognition tasks.", International Telecommunication Union, pp. 1-22, 2016.

[26] T. Zinner, O. Hohlfeld, Osama Abboud and T. Hossfeld, "Impact of Frame Rate and Resolution on Objective QoE Metrics," Second International Workshop on Quality of Multimedia Experience (QoMEX), pp. 29-34, June 2010.

[27] P. Orosz, T. Skopkó, Z. Nagy, P. Varga and L. Gyimóthi, "A Case Study on Correlating Video QoS and QoE," IEEE Network Operations and Management Symposium (NOMS), pp 1-5, May 2014.

[28] D. Z. Rodríguez, R. L. Rosa, E. A. Costa, J. Abrahão, and G. Bressan, "Video Quality Assessment in Video Streaming Services Considering User Preference for Video Content," IEEE Transactions on Consumer Electronics, vol. 60, pp. 570 - 571, March 2014.

[29] M. Vranješ, S. Rimac-Drlje, D. Žagar, "Subjective and Objective Quality Evaluation of the H.264/AVC Coded Video," 15th International Conference on Systems, Signals and Image Processing, pp. 287 – 290, June 2008.

[30] C. G. Bampis, Z. Li and A. C. Bovik, "Spatiotemporal Feature Integration and Model Fusion for Full Reference Video Quality Assessment," IEEE Transactions on Circuits and Systems for Video Technology, vol. 29, no. 8, pp. 2256-2270, Aug. 2019.

[31] D. Vranješ, D. Žagar and O. Nemčić, "Comparison of objective quality assessment methods for scalable video coding," Proceedings ELMAR-2012, pp. 19-22, September 2012.

[32] Z. Duanmu, K. Ma and Z. Wang, "Quality-of-Experience for Adaptive Streaming Videos: An Expectation Confirmation Theory Motivated Approach," IEEE Transactions on Image Processing, vol. 27, no. 12, pp. 6135-6146, 2018.

[33] https://media.xiph.org/ (accessed: 2020)

[34] C. G. Bampis, Z.Li, I. Katsavounidis, TY Huang, C. Ekanadham and A. C. Bovik, "Towards Perceptually Optimized End-to-end Adaptive Video Streaming," submitted to IEEE Transactions on Image Processing.

[35] "ITU-T P.911 Subjective audiovisual quality assessment methods for multimedia applications," International Telecommunication Union, pp. 1-27, 1998.

[36] https://www.ffmpeg.org/ (accessed: 2020)

[37] J. Vlaović, M. Vranješ, D. Grabić and D. Samardžija, "Comparison of Objective Video Quality Assessment Methods on Videos With Different Spatial Resolutions," 2019 International Conference on Systems, Signals and Image Processing (IWSSIP), pp. 287-292, June 2019.

[38] W. Kirch, Pearson's Correlation Coefficient, Springer, 2008.