PERFORMANCE ANALYSIS OF FULL-REFERENCE OBJECTIVE IMAGE AND VIDEO QUALITY ASSESSMENT METRICS

Boban P. Bondžulić[a], *Boban* Z. Pavlović[b], *Vladimir* S. Petrović[c]

[a] University of Defence in Belgrade, Military Academy, Department of Telecommunications and Informatics, Belgrade, Republic of Serbia, e-mail: bondzulici@yahoo.com, ORCID iD: http://orcid.org/0000-0002-8850-9842

[b] University of Defence in Belgrade, Military Academy, Department of Telecommunications and Informatics, Belgrade, Republic of Serbia, e-mail: bobanpav@yahoo.com, ORCID iD: http://orcid.org/0000-0002-5476-7894

[c] University of Novi Sad, Faculty of Technical Sciences, Novi Sad, Republic of Serbia, e-mail: v.petrovic@manchester.ac.uk, ORCID iD: http://orcid.org/0000-0002-8299-8999

http://dx.doi.org/10.5937/vojtehg66-12708

FIELD: Telecommunications
ARTICLE TYPE: Original Scientific Paper
ARTICLE LANGUAGE: English

Summary:

This paper presents the performance evaluation of image and video quality assessment metrics on two publicly available datasets with subjective quality ratings. In addition to the performance analysis at the global level – at the level of complete datasets, the paper presents the objective measures performance evaluation on subsets of signals inside them. The image dataset contains five subsets created by using different types of JPEG compression, while the video dataset contains six subsets of sequences – four created by compression of original sequences, and two subsets are with video signal transmission characteristic degradations. To determine the success of objective measures, i.e. comparison of subjective and objective quality scores, there were used measures accepted by the International Telecommunication Union – ITU (linear correlation coefficient, rank-order correlation coefficient, mean absolute error, root mean squared error and outlier ratio). It was shown that objective quality measures can reach a high level of agreement with the results of subjective tests on subsets of datasets. Objective measures performances depend on the type of degradation which significantly affects the performance at the complete dataset level. The difference in

performances is more pronounced on video sequences due to considerable visual differences in sequences created by using compression, packet losses and additive Gaussian noise. Therefore, we can say that a universal objective measure, i.e. measure that is useful for different types of signal degradation, for different degradation levels, and for different applications currently does not exist.

Key words: JPEG compression, H.264 and H.265 video compression, objective image and video quality assessment.

Introduction

Images, or videos, pass numerous processing and transmission phases before being viewed by the observer, and each of the phases may enter degradations that affect the quality of the final presentation. Image and video degradation in the recording process can occur due to the characteristics of the optics, sensor noise, color calibration, exposure time, motion of the camera, etc. After recording, an image or video adapts to the bandwidth of the transmission system through the compression process. The introduction of a high degree of compression is carried out at the expense of greater signal degradation. During transmission through a communication channel or during archiving, bit errors can also lead to degradation. Finally, end-user devices can affect the subjective quality impression (poor resolution, calibration, etc.) (Bovik, 2013).

As a human is an observer and user of the largest number of imaging systems, subjective assessment is the most reliable method for evaluating the quality of visual signals. In order to avoid subjective assessment, a procedure for automatic, computational image/video quality evaluation is required. An automated assessment procedure is called objective assessment and is useful in many applications where visual effects on images during recording, processing, compression, transmission and archiving need to be evaluated. Objective assessment measures do not require testing equipment, there is no complex organization of viewers, and with software implementation, time estimates are reduced to real-time (Bovik, 2010).

The basic goal of the research in the field of objective quality assessment is the development of a quantitative measure that (algorithmically, automatically) evaluates the quality of images/videos and is in good correlation with the mean opinion score (MOS). The ideal objective quality assessment measure should be applicable to different types of distortion, quantitatively covering different degrees of distortion and a wide range of content of the source signal. In practical applications, due to demands for real-time work, besides the conditions listed, computational complexity is also important (Wang et al, 2002).

Objective quality assessment measures have three types of application. The first is to monitor the image quality to control the transmission system. In this sense, quality assessment algorithms can be used to improve image quality through the perceptual optimization of the recording process, change of the transfer rate, resource reallocation with the goal of balancing quality through the network, through post-processing or by combining such approaches. Second, they can be used to select the system and image processing algorithms. Third, they can be embedded in an image processing system to optimize the algorithms and parameters used. For example, achieve a minimum degree of signal degradation for the given bit rate using compression, or achieve an acceptable level of signal degradation with the lowest bit rate (Wang et al, 2002).

According to the amount of information of the original (source, reference) image used in the quality assessment process on the receiving side (observer side), objective image quality evaluation can be divided into three categories: no-reference (NR), full-reference (FR) and reduced-reference (RR) (Bovik, 2013).

NR objective measures do not require knowledge of the original image and the assessment depends entirely on the human perception of the test image. Such measures can be used in all applications where quality measurement is required. Reliable NR quality assessment is currently possible if the type of distortion is known (JPEG compression, JPEG2000 compression, blurring), and in recent times general NR techniques for quality assessment appear (Wang & Bovik, 2011).

FR measures require full knowledge of the original image information. In this case, the quality assessment system can be considered as a communication system in which the original image is on the transmitting side and a test image (image with degradation) is on the receiving side. The basis of FR measures is comparison of the two images (source and test images) at the pixel level, region level and/or frequency characteristics level. However, in some real-world applications, it is not possible to know the original image on the receiving side. Also, objective FR image/video quality measures usually require precise spatial and temporal registration.

The original image/video usually comes from a high quality sensor and as such requires much more resources than image/video after compression. For this reason, FR quality assessment measures are used in laboratory tests to select image and video processing techniques.

RR techniques are between the previous two categories and are designed to provide practical solutions in quality assessment while

retaining the accuracy of the quality assessment. In these techniques, only the most important source information is sent from the transmitting to the receiving side. Since in this case the amount of additional information is not large, the requirements regarding the bandwidth of the channel are not significantly changed.

RR and NR quality assessment algorithms can be used as agents in the network data transfer, for installation in routers, set boxes, smart phones, tablets, or laptops. Through them, feedback information can be obtained for source adaptation and control mechanisms resource allocation, source coding and other network parameters (Wang & Bovik, 2011).

In this paper, we analyzed the performance of full-reference objective image quality and video quality measures on two publicly available datasets with subjective quality impressions – JPEG XR Image Dataset (De Simone et al, 2009) and CSIQ Video Dataset (Vu & Chandler, 2014). The JPEG XR Image Dataset is selected because it contains high spatial resolution images, while the CSIQ Video Dataset is selected because it contains sequences with recently introduced H.265 compression. The analysis was carried out at the level of complete datasets (global level), as well as at the level of subsets (types of degradation) of the test signals within the datasets (degradation level). In the image quality assessment, eight objective measures were analyzed: peak signal to noise ratio (PSNR), universal image quality index (UIQI) (Wang & Bovik, 2002), structural similarity index (SSIM) (Wang et al, 2004) and its multi-scale version (MS-SSIM) (Wang et al, 2003), visual information fidelity (VIF) (Sheikh & Bovik, 2006), visual signal to noise ratio (VSNR) (Chandler & Hemami, 2007), most apparent distortion (MAD) (Larson & Chandler, 2010) and gradient-based objective image quality assessment measure Q^{AB} (Bondzulic & Petrovic, 2011).

Video quality estimation can be obtained directly by applying an image quality assessment measure frame by frame. Thus, the quality of the video sequence was carried out by averaging the frame quality values obtained using PSNR (Frame PSNR) and SSIM (Frame SSIM) measures.

Additionally, the paper also analyzes the performance of the objective video quality assessment measure VQ^{AB} (Bondžulić, 2016), which was created by the extension of the objective image quality assessment measure Q^{AB} (Bondzulic & Petrovic, 2011). The VQ^{AB} measure in its original form was used to evaluate the quality of the sequences in low bit rate transmission systems.

Some of the objective measures were analyzed for the first time in the quality assessment of high-resolution images (JPEG XR Image Dataset), and in the quality evaluation of the sequences with H.265 compression (CSIQ Video Dataset).

The work is organized in six sections. After the introduction, in the second part of the paper, the methods of generalization of the results of subjective tests are described. The third part of the paper presents the criteria used to evaluate the performance of objective image and video quality measures, which have been adopted by the International Telecommunication Union (ITU). The fourth and fifth parts of the paper present the performance of objective quality assessment measures on the JPEG XR Image Dataset and the CSIQ Video Dataset. The last part of the paper is a conclusion.

Subjective quality assessment

Procedures and standards for subjective quality assessment of speech, audio and video signals are available throughout the years. Subjective tests for assessing the quality of visual signals have been formalized in the recommendations (International Telecommunication Union, 2008), (International Telecommunication Union, 2012) and (International Telecommunication Union, 2016), which suggest observation conditions, criteria for the choice of observers and test materials, and methods of data analysis.

The outputs of subjective experiments are the observations of the quality of the observers, which are consolidated after the test and represented through the mean opinion score (MOS). MOS is the most commonly used method of subjective scores generalization and is used as the basis for the development of objective quality assessment measures:

$$MOS_i = \frac{1}{N_S}\sum_{n=1}^{N_S} SQ(n,i) \tag{1}$$

where are:

i – index (label) of the video with degradation in a subjective test,

$SQ(n,i)$ – the subjective quality given by the observer n to the sequence i,

N_S – the number of observers in the subjective test.

It can be said that MOS is a "democratic" measure that treats each subjective score equally and really represents a mean (average) opinion.

MOS values can be used to compare to the values obtained by applying objective measures.

Since the MOS value is obtained from several individual quality scores, a quality evaluation can be associated with a certain level of statistical uncertainty. If there are significant fluctuations of subjective scores, this uncertainty is high. The uncertainty of subjective scores can be measured in many ways, but standard deviation, variance and standard error are commonly used.

Standard deviation is determined on the basis of individual subjective video quality scores, SQ_i (index n corresponding to the observer is ignored) and the mean opinion score of that video:

$$\sigma_i = \sqrt{\mathrm{E}[SQ_i^2]\text{-}(\mathrm{E}[SQ_i])^2} = \sqrt{\mathrm{E}[SQ_i^2]\text{-}MOS_i^2} \tag{2}$$

where E[X] is the expected value of the variable X.

The standard error of the MOS is determined from the standard deviation of the subjective scores and takes into account the number of measurements in determining the measured value. In this case, that is the number of observers who participated in subjective tests, N_S:

$$SE_i = \frac{\sigma_i}{\sqrt{N_S}} \tag{3}$$

The standard error is usually displayed along with the mean value as its positive and negative deviation ($MOS_i \pm 2 \cdot SE_i$) – confidence interval, but it does not speak about the reasons for the deviation.

Except through the MOS, the results of subjective tests can also be presented through the difference mean opinion scores (DMOS) obtained by subtracting the subjective evaluation of the signal with a distortion from the average subjective estimation of the corresponding original signal:

$$DMOS_i = MOS_i^{original} - MOS_i^{distorted} \tag{4}$$

Criteria for performance evaluation of objective image and video quality assessment metrics

The most common attributes reflecting the success of the objective measure in the image/video quality prediction are: (1) the accuracy of the prediction, (2) the prediction monotonicity and (3) the consistency of the prediction (International Telecommunication Union, 2004). Objective and subjective tests provide numerical (scalar) values for each original and

test signal that indicate the relationship between the quality of the original signal and the test signal. From the subjective tests results, mean values of the scores (MOS/DMOS) and confidence intervals are used.

The results of the objective quality assessment metrics – video quality ratings (VQR) are compared with the results of subjective tests that can be delivered through DMOS or MOS quality scores. The connection between VQR and DMOS/MOS scores does not have to be linear as the results of subjective tests can have non-linear compression (scaling) of scores around the extreme values of the used quality range. In order to eliminate the nonlinearities introduced by the subjective evaluation process (Figure 1), the relationship between predictions of objective measurements and subjective quality scores is observed using nonlinear regression between VQR and DMOS sets. By introducing non-linear mapping, the accuracy and consistency of the prediction change, and the prediction monotonicity remains the same.

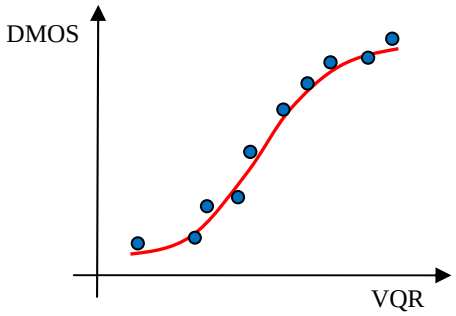


Figure 1 – An example of a connection between VQR and DMOS
Рис. 1 – Пример связи между VQR и DMOS
Слика 1 – Пример везе између VQR и DMOS

Nonlinear regression is performed over [VQR, DMOS] datasets, and must be monotone in the range of VQR scores. The shape of nonlinear regression is not critical, with the condition that it is a monotone, generally acceptable and has as few free parameters as possible to facilitate the interpolation of data. In regression, different forms are used for each of the objective measures and chooses the one that is the most appropriate (with minimal mean square error) for a given measure.

The most common functions used in regression are (International Telecommunication Union, 2004):

– third-order polynomial with four parameters

$$y = Quality(x) = A_0 + A_1 x + A_2 x^2 + A_3 x^3 \quad (5)$$

– logistic function with four parameters

$$y = Quality(x) = \frac{\beta_1 - \beta_2}{1 + \exp\left(\frac{x - \beta_3}{|\beta_4|}\right)} + \beta_2 \quad (6)$$

– logistic function with five parameters

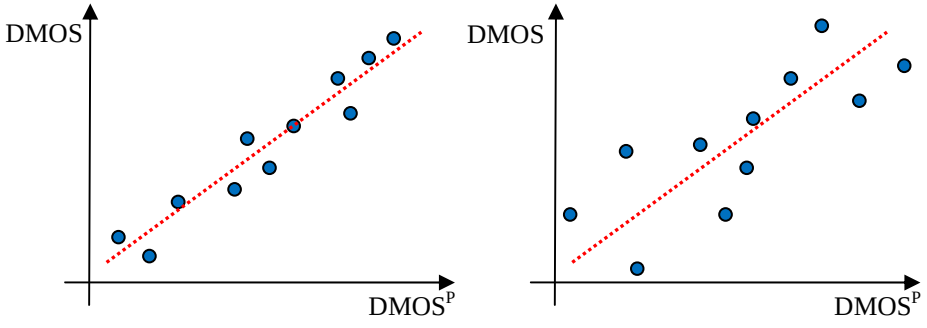
$$y = Quality(x) = \beta_1 \left(\frac{1}{2} - \frac{1}{1 + \exp\left(\beta_2 \left(x - \beta_3\right)\right)}\right) + \beta_4 x + \beta_5 \quad (7)$$

where x represents a set of VQR values ($x \equiv$VQR) while y represents a set of DMOS predictions, $DMOS^P$ ($y \equiv DMOS^P$), which are then compared with subjective DMOS scores.

For nonlinear mapping between objective estimates x and subjective scores y, the most commonly used function is accepted – the logistic function with four parameters (6).

Prediction accuracy

Accuracy is the ability of an objective quality estimation measure for the prediction of DMOS subjective scores with a minimal average error (International Telecommunication Union, 2004). Figure 2(a) shows the results of the measure with a smaller mean error between $DMOS^P$ and DMOS scores in relation to the measure of Figure 2(b). Therefore, the accuracy of the objective measure in Figure 2(a) is better.



(a) measure with greater accuracy *(b) measure with less accuracy*

Figure 2 – Objective quality assessment metrics comparison through prediction accuracy
Рис. 2 – Сравнение показателей объективности оценки качества методом точного прогнозирования
Слика 2 – Поређење мера објективне процене квалитета кроз тачност предикције

Numerous metrics can be used to determine the mean error, and the typically used ones are the mean absolute error (MAE) and the root-mean-squared error (RMSE) between objective estimates after non-linear mapping and subjective scores. If the difference between DMOS values and their predictions is determined as:

$$Perror_i = DMOS_i - DMOS_i^P \tag{8}$$

where the index *i* refers to the serial number of the analyzed test sequence, the mean absolute error and the root-mean-squared error for the set of *N* test sequences are obtained as:

$$MAE = \frac{1}{N}\sum_{i=1}^{N}\left|Perror_i\right| \tag{9}$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(Perror_i\right)^2} \tag{10}$$

The linear correlation coefficient (Pearson Linear Correlation Coefficient, LCC), although not a direct measure of the mean error, is also a common metric used to determine the accuracy of the prediction. The lower values of the MAE and RMSE (ideally equal to zero) correspond to the higher values of the correlation coefficient (ideally equal to the unit).

For a set of *N* pairs (x_i,y_i), the linear correlation coefficient is defined as:

$$LCC = \frac{\sum_{i=1}^{N}\left(x_i-\overline{x}\right)\left(y_i-\overline{y}\right)}{\sqrt{\sum_{i=1}^{N}\left(x_i-\overline{x}\right)^2\sum_{i=1}^{N}\left(y_i-\overline{y}\right)^2}} \tag{11}$$

where $\overline{x}$ and $\overline{y}$ are the mean values of the subsets *x* and *y*, which here represent DMOS and $DMOS^P$ sets.

Prediction monotonicity

Monotonicity is the degree to which objective predictions agree with the relative amplitudes of subjective assessments (International Telecommunication Union, 2004). $DMOS^P$ values derived from the results of an objective measure should ideally be completely monotone compared to the paired DMOS values, i.e. changing the $DMOS^P$ value should have the same sign as the change of the DMOS scores.

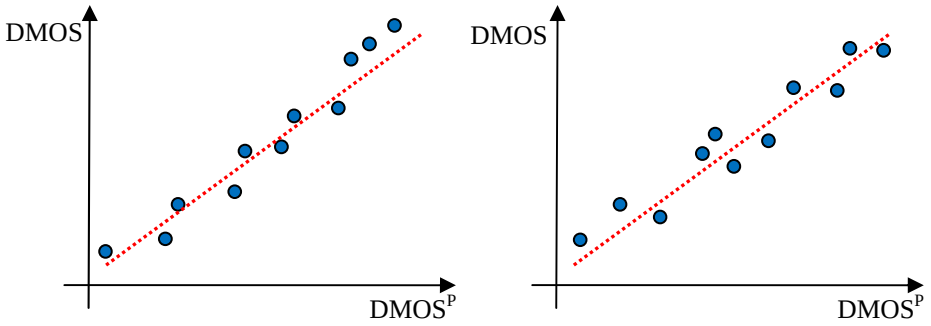
Figure 3 illustrates hypothetical connections between $DMOS^P$ and DMOS for two measures of different monotonicity. Both measures have approximately the same accuracy but the prediction monotonicity in Figure 3(a) is better. The monotonicity prediction in Figure 3(b) is worse. This is seen through the fall of the $DMOS^P$ values relative to what the observers actually see and what is being done through increasing the DMOS values.

Prediction monotonicity can be measured by the Spearman rank-order correlation coefficient (SROCC), which is defined as:

$$SROCC = \frac{\sum_{i=1}^{N}(\chi_i - \bar{\chi})(\gamma_i - \bar{\gamma})}{\sqrt{\sum_{i=1}^{N}(\chi_i - \bar{\chi})^2 \sum_{i=1}^{N}(\gamma_i - \bar{\gamma})^2}} = 1 - \frac{6 \cdot \sum_{i=1}^{N}(\chi_i - \gamma_i)^2}{N\left(N^2 - 1\right)} \qquad (12)$$

where χ_i and γ_i are x_i and y_i ranks, while $\bar{\chi}$ and $\bar{\gamma}$ are their mean values.

This parameter compares the change between adjacent pairs of DMOS values with a change between the corresponding $DMOS^P$ values. As the SROCC only works with the rankings (order) of the data and ignores the relative distances between them, it is taken as a measure of correlation with less sensitivity and is typically used if the number of points (samples) is small. With the increase in the SROCC value, the monotonicity of the objective quality assessment measure (ideally, SROCC=1) grows.



(a) measure with more monotonicity *(b) measure with less monotonicity*

Figure 3 – Objective quality assessment metrics comparison through prediction monotonicity

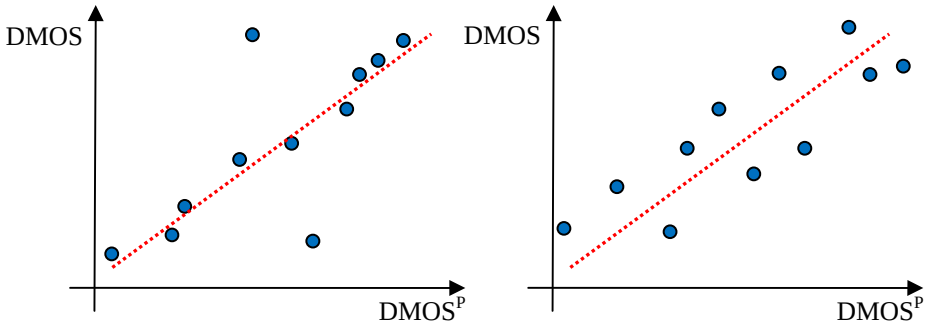
Рис. 3 – Сравнение показателей объективности оценки качества методом монотонного прогнозирования

Слика 3 – Поређење мера објективне процене квалитета кроз монотоност предикције

Prediction consistency

This attribute is related to the extent to which the objective quality assessment measure preserves the prediction accuracy through a whole set of analyzed video sequences (International Telecommunication Union, 2004). An objective measure should be consistent for all types of video sequences, that is, there is no significant deviation for subset of the analyzed sequences.

Figure 4 shows the results derived from two objective quality estimation measures with approximately the same values of the MAE/RMSE between DMOS and $DMOS^P$ datasets. Figure 4(a) is an example of a measure which has precise prediction for most of the sequences, but also has large prediction errors for two points in the middle part of the image. Figure 4(b) is an example of a measure that has balanced prediction errors – for most sequences it is not accurate as a measure in Figure 4(a), but its consistency is evident through acceptable predictions for all sequences.



(a) big prediction errors *(b) balanced prediction errors*

Figure 4 – Objective quality assessment metrics comparison through prediction consistency
Рис. 4 – Сравнение показателей объективности оценки качества методом согласованного прогнозирования
Слика 4 – Поређење мера објективне процене квалитета кроз конзистентност предикције

The consistency of an objective quality estimation measure can be determined by the number of points for which the prediction error is greater than the adopted threshold. The threshold is usually twice the DMOS scores standard error, i.e. the prediction error exists if:

$$\left|Perror_i\right| > 2 \cdot DMOS_i^{SE} \tag{13}$$

The number of prediction errors relative to the total number of points is called the outlier ratio, OR, and is used as a measure of the consistency of objective quality assessments. For an objective measure with a smaller OR ratio, it is said to be more consistent.

Analysis of results on the JPEG XR Image Dataset

The JPEG XR (eXtended Range) Image Dataset is designed to compare JPEG2000, JPEG and JPEG XR compression algorithms (De Simone et al, 2009). Compression of 24-bit high-resolution images (1280x1600 pixels) was considered through subjective tests carried out in four sessions. In subjective tests, 16 observers participated. The results of the subjective tests show a high degree of consistency, and can be used to compare the performance of the analyzed compression algorithms. The subjective quality evaluation was done in the laboratory of the Multimedia Signal Processing Group (MMSP) of the Ecole Politechnique Federale de Lausanne (Switzerland) academic institution (EPFL).

The dataset contain 10 original images of different content, the distribution of color and texture. Four original images were used in the observer training stage, while the remaining six original images were used in subjective tests.

In the JPEG2000 encoding, two configurations were used, which differ according to the sampling in the color channels (4:2:0 and 4:4:4). The JPEG XR compression was performed using two implementations obtained from Microsoft Corporation (MS) and Pegasus Corporation (PS). JPEG XR is a block-based compression developed to optimize image quality and compression efficiency with a small encoder/decoder complexity.

Test images were obtained using six degrees of compression – 0.25, 0.50, 0.75, 1.00, 1.25 and 1.50 bpp. In this way, 30 degraded images are obtained for each original image (5 compression algorithms x 6 degrees of compression), and the dataset contains 180 test images that are subjectively evaluated.

The observers gave subjective assessments by simultaneous observation of the original and test image, with the task of first detecting the test image, and then evaluating it. The impressions of quality were given using a continuous scale at an interval from 0 to 100. The results of subjective tests were presented through the mean subjective scores (MOS) and confidence intervals.

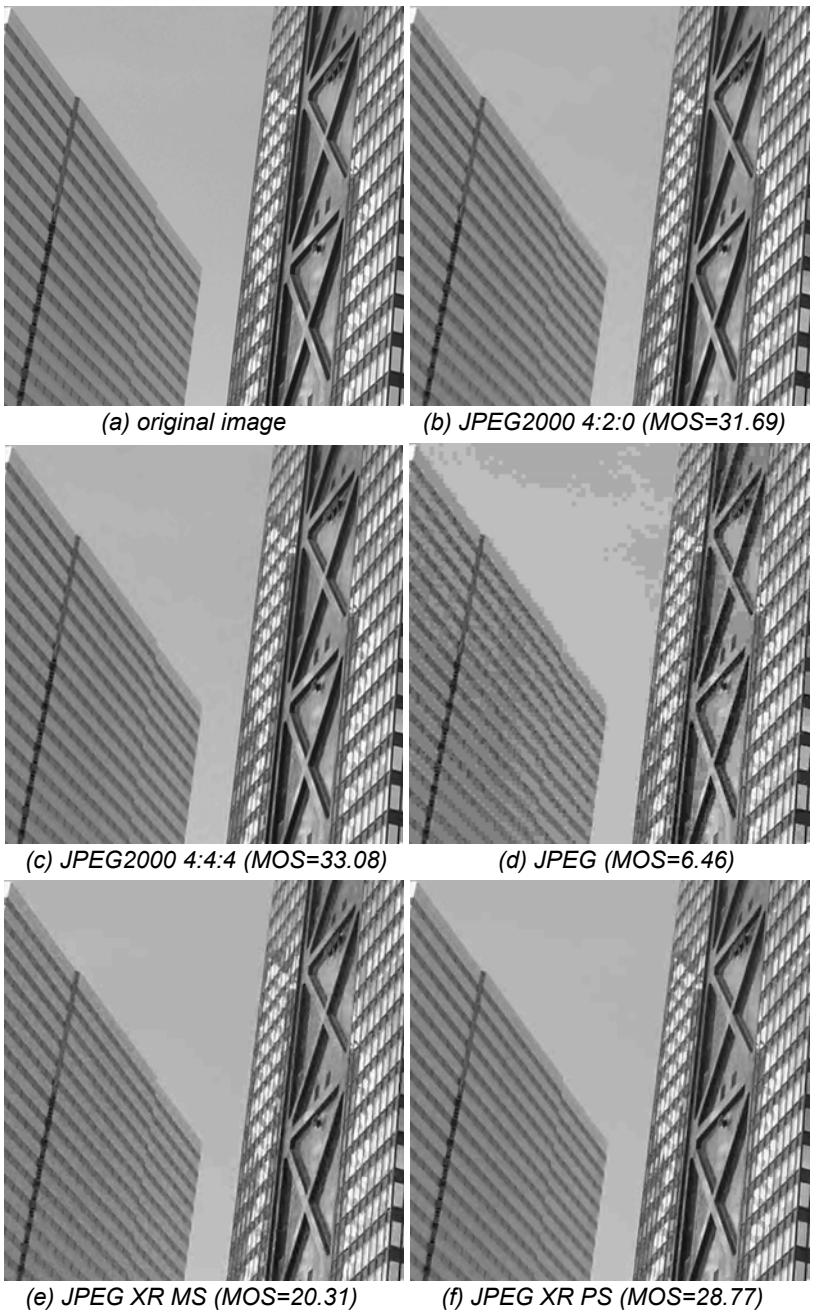
The results of the subjective tests have shown that for situations in which the pixel is represented with more than one bit (1 bpp) the quality of compressed images is approximately the same for all codecs. Also, JPEG2000 4:4:4, JPEG XR MS and JPEG XR PS compression algorithms have shown stable behavior for different content and different bit rates. On the other hand, the performance of JPEG2000 4:2:0 and JPEG compression algorithms significantly depend on the content of the image and the degree of compression (De Simone et al, 2009).

Figure 5 shows the image degradation examples of the JPEG XR Image Dataset (the original image "p10_orig.bmp" and the corresponding images with degradation at 0.25 bpp). In order to better understand the differences in quality, parts of images are shown (700x700 pixels). Subjective quality impressions are also given. The observers rated the JPEG compressed image with the smallest grade, in which the blocking effects are very noticeable. Blocking effects are less visible in the JPEG XR compression algorithms thanks to the adaptive quantization techniques used in them. The images with the JPEG2000 compression show typical artifacts created by wavelet compression – blurring and ringing. Blurring occurs due to the attenuation of high spatial frequencies of the image, while the ringing effect arises due to the quantization of high frequency coefficients in transformational coding. These effects are perceived through the spreading of edges, the loss of detail and waves around the boundaries of the regions.

The performance of objective quality assessment measures on the complete JPEG XR Image Dataset is given in Table 1. Eight objective quality assessment measures were analyzed. The performance of two measures with the best results are marked with bold font. The best matching of subjective and objective quality scores is obtained by using the MAD and VIF objective measures, while the PSNR is with the worst performance.

Table 1 – Performance of objective quality assessment metrics on the JPEG XR dataset
Таблица 1 – Эксплутационные характеристики показателей объективности оценки качества на основании JPEG XR базы
Табела 1 – Перформансе објективних мера процене квалитета на JPEG XR бази

Measure	LCC	SROCC	MAE	RMSE	OR (%)
PSNR	0.7819	0.7980	12.8737	16.5360	35.5556
UIQI	0.8621	0.8186	9.5605	13.4404	23.3333
SSIM	0.8744	0.8435	9.6144	12.8684	23.8889
MS-SSIM	0.9309	0.8930	7.0745	9.6863	14.4444
VIF	**0.9389**	**0.9130**	**6.8067**	**9.1278**	13.3333
VSNR	0.8765	0.7803	10.1065	12.7692	23.3333
MAD	**0.9466**	**0.9406**	**6.2598**	**8.5498**	**11.1111**
Q^{AB}	0.9269	0.8995	6.8809	9.9561	**11.6667**



(a) original image *(b) JPEG2000 4:2:0 (MOS=31.69)*

(c) JPEG2000 4:4:4 (MOS=33.08) *(d) JPEG (MOS=6.46)*

(e) JPEG XR MS (MOS=20.31) *(f) JPEG XR PS (MOS=28.77)*

Figure 5 – Degradation examples introduced in the JPEG XR original image
Рис. 5 – Примеры искажения исходного изображения из JPEG XR базы
Слика 5 – Примери деградација унетих у изворну слику JPEG XR базе

Figure 6 shows the scatter plots of subjective (MOS) versus MAD/VIF objective quality values on the JPEG XR Image Dataset, where each point represents a single test image (lower MAD, or higher VIF values correspond to better subjective quality). Vertical and horizontal axes represent MOS and the obtained objective estimates, respectively. There is an almost linear relationship between subjective and objective quality scores, with constant dissipation of quality scores around the interpolation curve in a complete range of quality.

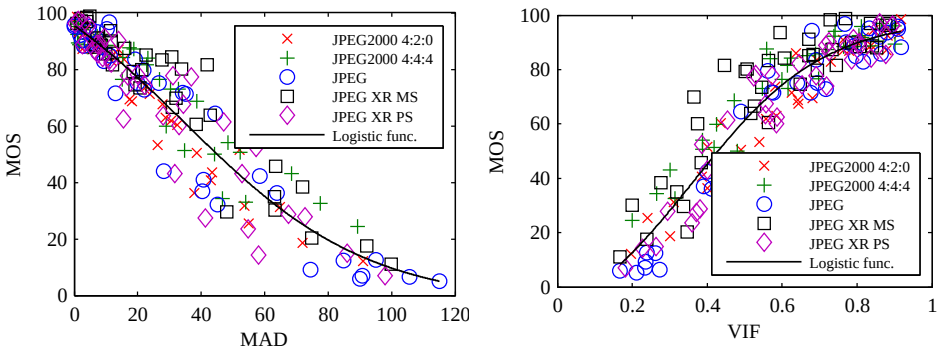


Figure 6 – Scatter plots of subjective (MOS) versus objective quality predictions on the JPEG XR Image Dataset
Рис. 6 – Диаграмма распространения субъективных (MOS) и объективных показателей на основании JPEG XR базы
Слика 6 – Дијаграми расипања субјективних (MOS) и објективних вредности квалитета на JPEG XR бази слика

Tables 2 and 3 show the performance of objective measures on image subsets within the JPEG XR Image Dataset, through the linear correlation coefficient (LCC) and the correlation of the ranks (SROCC). The performance of objective measures depends on the choice of a subset of the JPEG XR Image Dataset. Thus, the degree of agreement between subjective and VSNR objective quality scores (measured through SROCC) in subgroups ranges from 64% to 84%. The performance of other objective measures also varies from a subgroup to a subgroup, with the most stable performance of the MAD objective measure (the correlation coefficient over all subsets is greater than 93%).

Perhaps the most important aspect of the performance of an objective measure is its ability to reliably evaluate and rank various image compression systems, which can be achieved through costly and

time-long subjective tests. The subjective codec evaluation at different compression levels is available from subjective tests and can be directly compared to objective scores. Methodologically, this includes separating the image quality scores obtained by different compression types for the specific compression rate and adding them to one score that reflects this type of compression (e.g. JPEG at 0.25 bpp). This is done for subjective and objective quality scores for all types of compression, after which a correlation of the resulting (mean) quality scores is determined.

Table 2 – Linear correlation coefficient (LCC) between subjective and objective quality scores (after nonlinear regression) on the JPEG XR Image Dataset
Таблица 2 – Коэффициент линейной корреляции (LCC) субъективных и объективных показателей (после нелинейной регрессии) на основании JPEG XR базы
Табела 2 – Коефицијент линеарне корелације (LCC) субјективних и објективних скорова (након нелинеарне регресије) на JPEG XR бази слика

Measure	JPEG2000 4:2:0	JPEG2000 4:4:4	JPEG 4:2:0	JPEG XR MS 4:2:0	JPEG XR PS 4:2:0
PSNR	0.8990	0.8359	0.7876	0.7743	0.7628
UIQI	0.8564	0.8420	0.9009	0.8780	0.8770
SSIM	0.9460	0.8773	0.8956	0.8609	0.8504
MS-SSIM	**0.9706**	0.9500	0.9389	0.9167	**0.9430**
VIF	0.9683	**0.9544**	**0.9643**	**0.9247**	**0.9657**
VSNR	0.9020	0.8822	0.9247	0.9221	0.9167
MAD	**0.9748**	**0.9608**	**0.9647**	**0.9516**	0.9306
Q^{AB}	0.9483	0.9225	0.9432	0.9231	0.9266

Table 3 – Spearman rank-order correlation coefficient (SROCC) between subjective and objective quality scores on the JPEG XR Image Dataset
Таблица 3 – Ранговая корреляция (SROCC) субъективных и объективных показателей (после нелинейной регрессии) на основании JPEG XR базы
Табела 3 – Корелација рангова (SROCC) субјективних и објективних скорова на JPEG XR бази слика

Measure	JPEG2000 4:2:0	JPEG2000 4:4:4	JPEG 4:2:0	JPEG XR MS 4:2:0	JPEG XR PS 4:2:0
PSNR	0.8888	0.8719	0.7640	0.7732	0.7938
UIQI	0.8234	0.8318	0.8049	0.8505	0.8278
SSIM	0.9284	0.8578	0.8005	0.8293	0.8239
MS-SSIM	0.9539	0.8981	0.8674	0.8698	0.8963
VIF	**0.9601**	**0.9268**	**0.9107**	0.8824	**0.9305**
VSNR	0.8376	0.6435	0.8414	0.8376	0.8095
MAD	**0.9665**	**0.9624**	**0.9428**	**0.9390**	**0.9315**
Q^{AB}	0.9323	0.9225	0.8746	**0.9148**	0.9012

Figure 7 shows the mean values of the subjective and objective quality scores of different types of compression, at different degrees of compression. In this case, averaging is carried out by taking the appropriate values of the subjective, i.e. objective quality values for six visual contents (generated from six original images). The degree of agreement between subjective and objective quality scores is given in Table 4. Although Table 4 shows an extremely high degree of agreement between subjective and objective quality scores, none of the objective quality assessment measures has reached a full compliance with the rankings of subjective quality impressions. This is also evident from Figure 7, where there is an almost constant quality of all compression algorithms for bitstreams greater than 1 bpp. As the compression rate increases, subjective and objective differences become greater. Objective measures correctly detected the worst quality obtained by using JPEG compression for the degree of compression of 0.25 and 0.5 bpp. However, while the observers in these situations preferred the JPEG2000 4:4:4 compression type, according to objective quality assessment measures, priority is given to the JPEG2000 4:2:0 compression. This is due to the introduction of the quantization tables of visual significance in the JPEG2000 4:4:4 compression, which gave the viewers better visual quality but "confused" the objective quality assessment measures (made an objective difference between test images and original images). Also, from Figure 7 it can be observed that the observers between two implementations of JPEG XR compression (with different quantization techniques) favor the MS implementation for all degrees of compression. However, objective measures favor the PS implementation.

Although the results/conclusions are derived from averaging on six visual content, it can be concluded that there is still a need for the development of new and improvement of existing techniques of objective image quality assessment. Bearing in mind the results on the JPEG XR Image Dataset, the possible direction for improvement is through the implementation of different characteristics of the human visual system in objective measurements (VSNR, MAD, VIF and Q^{AB} objective measures include some of the human visual system characteristics). As the image datasets mainly contain color images (as well as the JPEG XR Image Dataset), and objective measures are mainly designed to work only in the intensity channel, another possible direction of improving the performance of objective measures is through their extension to analyze the preservation of information from the color channels.

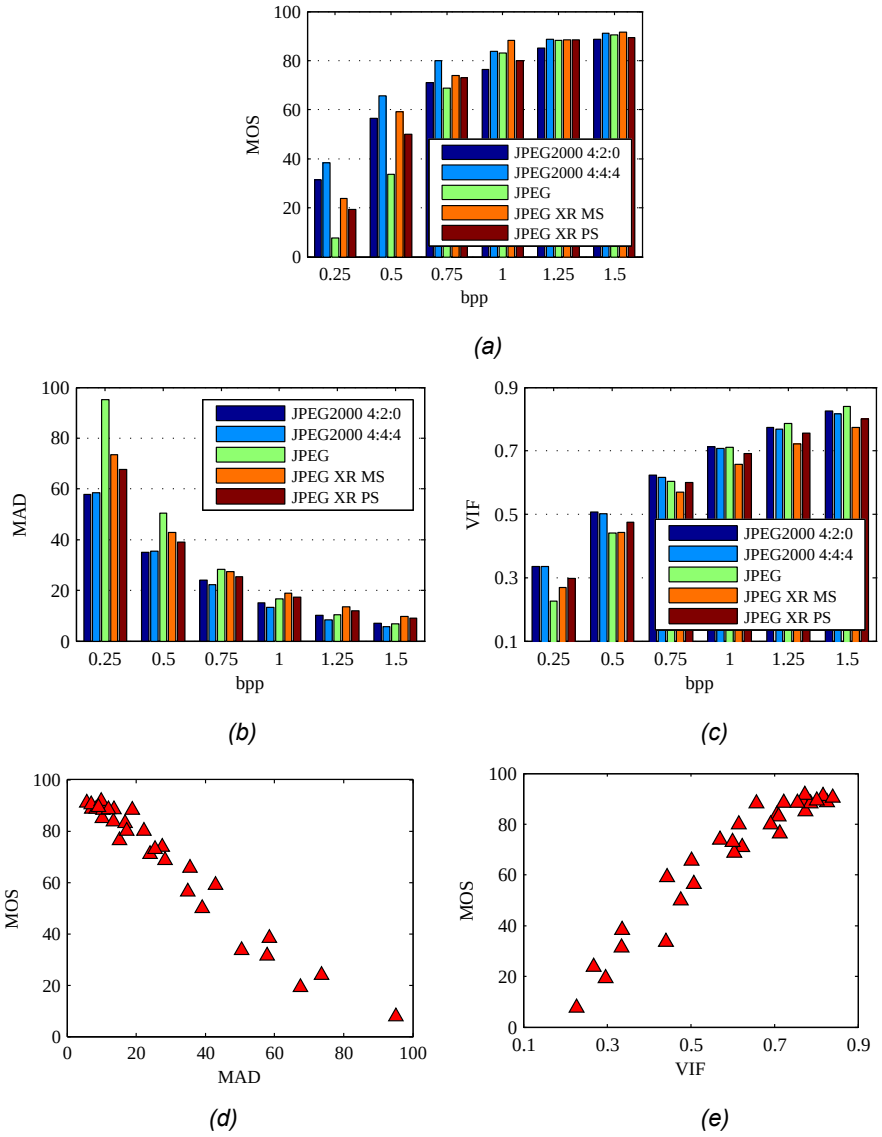


Figure 7 – (a) mean values of subjective quality scores (MOS), (b) (c) mean values of MAD and VIF objective quality scores and (d) (e) scatter plots of subjective versus objective quality predictions after averaging

Рис. 7 – (a) средние значения субъективных (MOS) показателей качества, (b) (c) средние значения MAD и VIF показателей качества и (d) (e) диаграммы распространения субъективных и объективных показателей качества после усреднения

Слика 7 – (a) средње вредности субјективних (MOS) скорова квалитета, (b) (c) средње вредности MAD и VIF скорова квалитета и (d) (e) дијаграми расипања субјективних и објективних скорова квалитета након усредњавања

Table 4 – Comparison of objective quality assessment metrics performances after averaging subjective/objective scores on the JPEG XR Image Dataset
Таблица 4 – Сравнение характеристик объективности оценки качества после усреднения субъективных/объективных показателей качества на основании JPEG XR базы
Табела 4 – Поређење перформанси објективних мера процене квалитета након усредњавања субјективних/објективних скорова квалитета на JPEG XR бази

Measure	LCC	SROCC
PSNR	0.9228	0.9138
UIQI	0.9756	0.9834
SSIM	**0.9769**	0.9670
MS-SSIM	0.9401	**0.9839**
VIF	0.9603	0.9552
VSNR	0.9462	0.8918
MAD	**0.9797**	0.9692
Q^{AB}	0.9676	**0.9879**

Analysis of the results on the CSIQ Video Dataset

The CSIQ Video Dataset (Vu & Chandler, 2014) was developed by the Oklahoma State University in order to obtain a set of data for the validation of objective video signal quality assessment measures. The dataset consists of 12 high quality reference videos and 216 degraded videos. Degradations were made by using six different types of distortion. In the subjective tests, 35 observers participated.

The video sequences in the dataset are in the raw YUV420 format, with 832x480 resolution, 10 seconds duration and a wide tempo range: 24, 25, 30, 50 and 60 fps. All types of distortion were applied to each reference video signal, in three different degrees. In this way, for each original video, there are 18 test video sequences. Four compression distortions and two distortions that occur during transmission were used:

– H.264 compression (H.264/AVC),
– HEVC/H.265 compression (HEVC),
– Motion JPEG compression (MJPEG),
– compression based on a wavelet transformation using the SNOW codec (Wavelet),
– packet losses caused by the wireless transmission of the H.264 compressed bitstreams (H.264/PLR) and
– additive white Gaussian noise (White noise).

The compression systems generally provide uniform distortion/quality in video, both spatial and temporal. The uniformity of the distortion is also characteristic for the sequences with the additive Gaussian noise. Packet losses, on the other hand, cause short-term distortions in the video in the form of flicker, both spatial and temporal.

Figure 8 shows the representative frames of the CSIQ Video Dataset test sequences, illustrating all six types of degradation used in this dataset. The figure shows that the visual effects of distortion are different. The video sequences with compression have typical compression artefacts, such as blocking effect, blurring, ringing effect and poor compensation of movement around the edges of the objects. It is also interesting to highlight the differences in the degradation of the resulting MJPEG and H.264/H.265 compressions, where the blocking effect is greatly reduced in the H.264/H.265 sequences. The errors in the packet networks are short-term and appear as sudden transitions in the video. Test sequences and frames were selected randomly, with the aim of illustrating the diversity of the contents of the original sequences and typical artefacts of their degradation. Through this set of selected sequences, the observers gave the highest quality to the H.264/AVC compressed sequence (DMOS=21.84) – Figure 8(a), and the lowest score was obtained for the H.265 (HEVC) compressed sequence (DMOS=75.79) – Figure 8(f).

Three objective quality assessment measures – Frame PSNR, Frame SSIM and VQ^{AB} were tested on the CSIQ Video Dataset, and their performance analysis is given in Table 5. Among the three measures, the best matching between subjective and objective quality scores is obtained with the VQ^{AB} objective measure. It is noted that the performance of objective measures is significantly worse than the performance of objective measures on the JPEG XR Image Dataset. A better understanding of these results can be obtained by analyzing the performance of objective measures on subsets of the CSIQ Video Dataset. The performance of objective measures in subgroups is given in Tables 6 and 7, through linear correlation (LCC) and rank-order correlation (SROCC) between subjective and objective quality scores. In addition to the three tested measures, the performance of four additional measures is also given – VQM (Pinson & Wolf, 2004), MOVIE (Seshadrinathan & Bovik, 2010), TQV (Narwaria et al, 2012) and VIS_3 (Vu & Chandler, 2014). The performance of additional measures was taken from (Vu & Chandler, 2014). Additionally, Figure 9 shows the scatter plots of subjective and objective quality scores on the CSIQ Video Dataset.

(a) H.264/AVC compression
BasketballDrive sequence (DMOS=21.84)

(b) H.264 with packet losses
Kimono sequence (DMOS=57.36)

(c) MJPEG compression
BQTerrace sequence (DMOS=66.48)

(d) wavelet compression (SNOW codec)
Timelapse sequence (DMOS=65.81)

(e) additive Gaussian noise
Keiba sequence (DMOS=42.06)

(f) HEVC compression
PartyScene sequence (DMOS=75.79)

Figure 8 – Example frames of the test sequences used in the CSIQ Video Dataset
Рис. 8 – Примеры кадров тестовых последовательностей на основании CSIQ видеобазы
Слика 8 – Репрезентативни кадрови тест секвенци CSIQ видео-базе

Table 5 – Performance comparison of various objective quality assessment metrics on the complete CSIQ Video Dataset
Таблица 5 – Сравнение характеристик объективности оценки качества на основании полной CSIQ видеобазы последовательностей
Табела 5 – Поређење перформанси објективних мера процене квалитета на комплетној CSIQ видео-бази секвенци

Measure	LCC	SROCC	MAE	RMSE	OR (%)
Frame PSNR	0.5820	0.5957	10.7469	13.5212	13.8889
Frame SSIM	0.6441	0.5769	10.1931	12.7189	12.0370
VQ^{AB}	**0.7160**	**0.6418**	**9.2039**	**11.6078**	**7.8704**

Table 6 – Linear correlation coefficient (LCC) between subjective and objective quality scores (after nonlinear regression) on the CSIQ Video subsets
Таблица 6 – Коэффициент линейной корреляции (LCC) субъективных и объективных показателей качества (после нелинейной регрессии) на основании подмножеств CSIQ видеобазы
Табела 6 – Коефицијент линеарне корелације (LCC) субјективних и објективних скорова квалитета (након нелинеарне регресије) на подскуповима CSIQ видео-базе

Measure	H.264/AVC	H.264/PLR	MJPEG	Wavelet	White noise	HEVC
Frame PSNR	0.8232	0.8236	0.6872	0.7713	**0.9494**	0.7851
Frame SSIM	0.8779	0.7666	0.8304	0.7878	**0.9446**	0.8258
VQ^{AB}	**0.9640**	0.6586	**0.9397**	**0.9000**	0.8815	**0.9589**
VQM	0.916	0.806	0.641	0.840	0.918	0.915
MOVIE	0.904	**0.882**	**0.882**	0.898	0.855	**0.937**
TQV	**0.965**	0.784	0.871	0.846	0.930	0.913
VIS_3	0.918	**0.850**	0.800	**0.908**	0.916	0.933

Table 7 – Spearman rank-order correlation coefficient (SROCC) between subjective and objective quality scores on the CSIQ Video subsets
Таблица 7 – Ранговая корреляция (SROCC) субъективных и объективных показателей качества на основании подмножеств последовательностей CSIQ видеобазы
Табела 7 – Корелација рангова (SROCC) субјективних и објективних скорова квалитета на подскуповима секвенци CSIQ видео-базе

Measure	H.264/AVC	H.264/PLR	MJPEG	Wavelet	White noise	HEVC
Frame PSNR	0.7949	0.8172	0.6530	0.7493	0.9053	0.7552
Frame SSIM	0.8582	0.7712	0.8196	0.7586	**0.9236**	0.8118
VQ^{AB}	**0.9627**	0.6792	**0.9393**	0.8937	0.8409	**0.9387**
VQM	0.919	0.801	0.647	0.874	0.884	0.906
MOVIE	0.897	**0.886**	**0.887**	**0.900**	0.843	**0.933**
TQV	**0.955**	0.842	0.870	0.831	0.908	0.902
VIS_3	0.920	**0.856**	0.789	**0.908**	**0.928**	0.917

Tables 6 and 7 show that the performance of objective quality assessment measures significantly depends on the type of degradation of the original sequences. Thus, the performance of the Frame PSNR is the worst for the MJPEG compression sequences, while the performance of the VQ^{AB} measure is the worst for a subset of sequences with packet losses (in this subset VQ^{AB} measure has the smallest agreement with subjective impressions among all other objective measures). The newly-proposed VQ^{AB} measure has very good performance for a subsets of sequences with compression, and slightly lower performance for the sequences with additive noise. This measure was tested on the H.265 compression sequences for the first time, where the best performance among the analyzed measures was achieved.

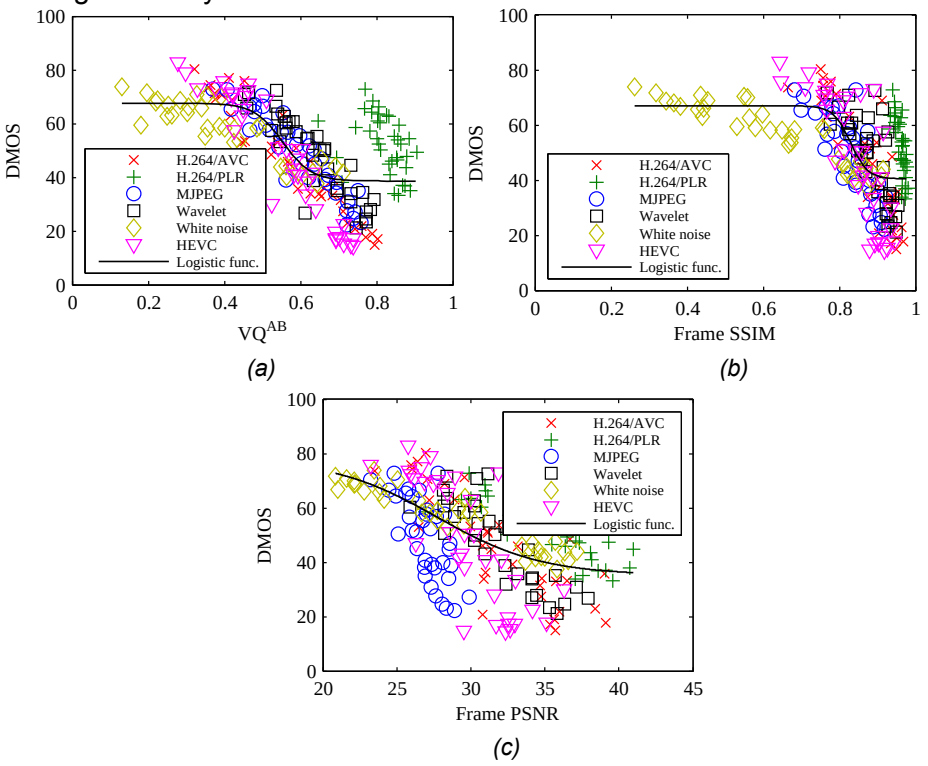


Figure 9 – Scatter plots of subjective (DMOS) versus objective quality predictions on the CSIQ Video Dataset: (a) VQ^{AB}, (b) Frame SSIM and (c) Frame PSNR

Рис. 9 – Диаграммы распространения субъективных (DMOS) и объективных показателей качества на основании CSIQ видеобазы последовательностей: (a) VQ^{AB}, (b) Frame SSIM и (c) Frame PSNR

Слика 9 – Дијаграми расипања субјективних (DMOS) и објективних вредности квалитета на CSIQ видео-бази секвенци: (a) VQ^{AB}, (b) Frame SSIM и (c) Frame PSNR

The scatter plot of the subjective and VQ^{AB} objective quality scores, shown in Figure 9(a), confirms that the bad VQ^{AB} measure results at the global level (on a dataset) originate from the worse results for a subset of sequences with packet losses, in which the objective values VQ^{AB} deviate from the trend of other scores. Also, with the frame SSIM objective measure, the deviation trend of the sequences with the packet losses compared to the sequences with compression is confirmed, and additionally there is the deviation trend of the sequences with additive Gaussian noise – Figure 9(b). With the Frame PSNR objective measure, it is possible to talk about an isotropic cloud in the space of subjective-objective quality scores – Figure 9(c).

Conclusion

This paper analyzed a possibility of using the full-reference objective quality assessment measures of visual signals – images and videos. The analysis was conducted on the two publicly available datasets with subjective quality impressions, with a representative number of visual signals (180 test images and 216 test sequences). These datasets contain the subsets of images/videos created by the characteristic degradations typical for processing and transmission of visual signals.

An analysis of the objective measures performance at the level of subsets of the signals inside the datasets has shown that the performance of objective measures depends on the choice of a subset, i.e. type of degradation. The difference between the performance is more pronounced on the video dataset due to significant visual differences inside the test sequences. Therefore, it can be said that there is no objective measure of quality assessment that will be useful in all situations – for different types of degradation, for different degrees of degradation, for various applications, etc. (universal measure).

It has been shown that the objective measures applied on the subsets can reach a high degree of agreement with the results of subjective tests. The maximum level of agreement between objective and subjective quality scores (measured through linear correlation) on the analyzed image subsets is 97.48% – MAD objective measure on the JPEG2000 4:2:0 subset. The same measure on complete image dataset also provided the maximum agreement with subjective quality scores – LCC=94.66%.

Within the video subsets with compression, the best results were achieved by the spatial and temporal gradient-based information

preservation measure, VQ^{AB}. At four subsets with compression, the degree of agreement between VQ^{AB} and the subjective quality scores ranges from LCC=90% to LCC=96.40%. The MOVIE objective measure provided the best results on the subset of sequences with the packet losses (LCC=88.20%), while the Frame PSNR measure is more suitable for quality evaluation of video sequences with the additive Gaussian noise (LCC=94.94%).

Objective quality evaluation is a very complex problem, but it is possible to solve it and reach high performance using approaches that have been proposed for image and video quality evaluation. Improvement of objective quality assessment at the global level is possible to be achieved by a fusion of objective quality assessment measures suitable for different types of degradation of the original signal.

References

Bondzulic, B., & Petrovic, V. 2011. Edge-based objective evaluation of image quality. In: *IEEE International Conference on Image Processing (ICIP),* Brussels, Belgium, pp.3305-3308. September 11-14. Available at: http://dx.doi.org/10.1109/ICIP.2011.6116378.

Bondžulić, B. 2016. *Procena kvaliteta slike i videa kroz očuvanje informacija o gradijentu*. Novi Sad: University in Novi Sad. Ph.D. thesis (in Serbian).

Bovik, A.C. 2010. Perceptual video processing: Seeing the future. *Proceedings of the IEEE*, 98(11), pp.1799-1803. Available at: http://dx.doi.org/10.1109/JPROC.2010.2068371.

Bovik, A.C. 2013. Automatic prediction of perceptual image and video quality. *Proceedings of the IEEE*, 101(9), pp.2008-2024. Available at: http://dx.doi.org/10.1109/JPROC.2013.2257632.

Chandler, D.M., & Hemami, S.S. 2007. VSNR: A wavelet-based visual signal-to-noise ratio for natural images. *IEEE Transactions on Image Processing*, 16(9), pp.2284-2298. pmid:17784602. Available at: http://dx.doi.org/10.1109/TIP.2007.901820.

De Simone, F., Goldmann, L., Baroncini, V., & Ebrahimi, T. 2009. Subjective evaluation of JPEG XR image compression. In: *Proc. of SPIE*, 7443, San Diego, CA, pp.1-12 74430L. August 02. Available at: http://dx.doi.org/10.1117/12.830714.

-International Telecommunication Union. 2004. *ITU TUTORIAL: Objective perceptual assessment of video quality: Full reference television*. Geneva, Switzerland.

-International Telecommunication Union. 2008. *ITU-T Recommendation P.910: Subjective video quality assessment methods for multimedia applications*. Geneva, Switzerland.

-International Telecommunication Union. 2012. *ITU-R Recommendation BT.500-13: Methodology for the subjective assessment of the quality of television pictures*. Geneva, Switzerland.

-International Telecommunication Union. 2016. *ITU-T Recommendation P.913: Methods for the subjective assessment of video quality, audio quality and audiovisual quality of Internet video and distribution quality television in any environment*. Geneva, Switzerland.

Larson, E.C., & Chandler, D.M. 2010. Most apparent distortion: Full reference image quality assessment and the role of strategy. *Journal of Electronic Imaging*, 19(1), pp.1-21, 011006. Available at: http://dx.doi.org/10.1117/1.3267105.

Narwaria, M., Lin, W., & Liu, A. 2012. Low-complexity video quality assessment using temporal quality variations. *IEEE Transactions on Multimedia*, 14(3), pp.525-535. Available at: http://dx.doi.org/10.1109/TMM.2012.2190589.

Pinson, M.H., & Wolf, S. 2004. A new standardized method for objectively measuring video quality. *IEEE Transactions on Broadcasting*, 50(3), pp.312-322. Available at: http://dx.doi.org/10.1109/TBC.2004.834028.

Seshadrinathan, K., & Bovik, A.C. 2010. Motion tuned spatio-temporal quality assessment of natural videos. *IEEE Transactions on Image Processing*, 19(2), pp.335-350, pmid:19846374. Available at: http://dx.doi.org/10.1109/TIP.2009.2034992.

Sheikh, H.R., & Bovik, A.C. 2006. Image information and visual quality. *IEEE Transactions on Image Processing*, 15(2), pp.430-444, pmid:16479813. Available at: http://dx.doi.org/10.1109/TIP.2005.859378.

Vu, P.V., & Chandler, D.M. 2014. ViS3: An algorithm for video quality assessment via analysis of spatial and spatiotemporal slices. *Journal of Electronic Imaging*, 23(1), pp.1-24, 01316. Available at: http://dx.doi.org/10.1117/1.JEI.23.1.0130.

Wang, Z., & Bovik, A.C. 2002. A universal image quality index. *IEEE Signal Processing Letters*, 9(3), pp.81-84. Available at: http://dx.doi.org/10.1109/97.995823.

Wang, Z., & Bovik, A.C. 2011. Reduced- and no-reference image quality assessment. *IEEE Signal Processing Magazine*, 28(6), pp.29-40. Available at: http://dx.doi.org/10.1109/MSP.2011.942471.

Wang, Z., Bovik, A.C., & Lu, L. 2002. Why is image quality assessment so difficult? In: *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP),* Orlando, FL, pp.3313-3316. May 13-17. Available at: http://dx.doi.org/10.1109/ICASSP.2002.5745362.

Wang, Z., Bovik, A.C., Sheikh, H.R., & Simoncelli, E.P. 2004. Image quality assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4), pp.600-612, pmid:15376593. Available at: http://dx.doi.org/10.1109/TIP.2003.819861.

Wang, Z., Simoncelli, E.P., & Bovik, A.C. 2003. Multi-scale structural similarity for image quality assessment. In: *Conference Record of the 37th Asilomar Conference on Signals, Systems and Computers, Pacific Grove,* CA, pp.1398-1402. November 9-12. Available at: http://dx.doi.org/10.1109/ACSSC.2003.1292216.

АНАЛИЗ ХАРАКТЕРИСТИК ОБЪЕКТИВНЫХ ПОКАЗАТЕЛЕЙ ОЦЕНКИ КАЧЕСТВА ФОТОГРАФИЙ И ВИДЕО НА ОСНОВАНИИ БАЗЫ ДАННЫХ

Бобан П. Бонджулич[а], *Бобан* З. Павлович[а], *Владимир* С. Петрович[б]

[а] Университет обороны в г. Белград, Военная академия, Кафедра телекоммуникаций и информатики, г. Белград, Республика Сербия

[б] Университет в г. Нови-Сад, Факультет технических наук, г. Нови-Сад, Республика Сербия

ОБЛАСТЬ: телекоммуникации
ВИД СТАТЬИ: оригинальная научная статья
ЯЗЫК СТАТЬИ: английский

Резюме:

В данной статье представлены рабочие характеристики объективных показателей оценки качества фотографий и видео, на основании двух соответствующих баз открытого доступа, с приведенными субъективными оценками качества. Наряду с анализом характеристик на глобальном уровне – полных баз, в данной работе представлен анализ и характеристики показателей подмножеств сигналов в их рамках. База фотографий состоит из пяти подмножеств, сформированных благодаря применению различных типов JPEG сжатия, в то время как видео последовательности содержат шесть подмножеств – четыре из которых образованы сжатием исходных последовательностей, а два подмножества представлены с искажениями, характерными для передачи видеосигналов. При определении успешности сравнения объективных и субъективных показателей качества применялись методы, утвержденные Международным телекоммуникационным содружеством (коэффициент корреляции, ранговая корреляция, средняя абсолютная ошибка и среднеквадратическое отклонение прогнозирования). В работе показано, что объективные показатели подмножеств сигналов из базы данных могут в большой степени совпадать с результатами субъективных тестов. Характеристики объективных показателей зависят от выбора подмножеств – типа искажения, и соответственно значительно влияют на характеристики на глобальном уровне, то есть, на уровне полной базы. Разница характеристик в большей мере выражена в видео последовательностях, вследствие значительных зрительных разниц в последовательностях, образовавшихся в процессе сжатия, потери пакетов и воздействия гауссовского шума. В заключение можно сказать, что на данный момент не существует единых универсальных объективных показателей, применяемых при различных типах искажений сигналов, и при различной степени искажений, и для различного назначения.

Ключевые слова: JPEG сжатие, H.264 и H.265 видеосжатие, объективная оценка качества фотографий и видео.

АНАЛИЗА ПЕРФОРМАНСИ ОБЈЕКТИВНИХ МЕРА ПРОЦЕНЕ КВАЛИТЕТА СЛИКА И ВИДЕА СА ПОТПУНИМ РЕФЕРЕНЦИРАЊЕМ

Бобан П. Бонџулић[а], *Бобан* З. Павловић[а], *Владимир* С. Петровић[б]

[а] Универзитет одбране у Београду, Војна академија, Катедра телекомуникација и информатике, Београд, Република Србија
[б] Универзитет у Новом Саду, Факултет техничких наука, Нови Сад, Република Србија

ОБЛАСТ: телекомуникације
ВРСТА ЧЛАНКА: оригинални научни чланак
ЈЕЗИК ЧЛАНКА: енглески

Сажетак:

У раду су представљене перформансе објективних мера процене квалитета слика и видеа на две јавно доступне базе са субјективним импресијама квалитета. Поред анализе перформанси на глобалном плану – нивоу комплетних база, у раду су анализиране и перформансе мера на подскуповима сигнала унутар њих. База слика садржи пет подскупова насталих применом различитих типова JPEG компресије, док база видео- секвенци садржи шест подскупова – четири настала компресијом изворних секвенци и два подскупа са деградацијама карактеристичним за пренос видео-сигнала. За одређивање успешности објективних мера, тј. поређење субјективних и објективних скорова квалитета, коришћене су мере које је прихватио ITU (коефицијент корелације, корелација рангова, средња апсолутна грешка, средња квадратна грешка и стандардна девијација процена). Показано је да објективне мере на подскуповима сигнала из база могу достићи висок степен слагања са резултатима субјективних тестова. Перформансе објективних мера зависе од типа деградације, што знатно утиче на перформансе на нивоу комплетне базе. Разлика у перформансама је израженија на бази видео-секвенци због знатних визуелних разлика у секвенцама насталим компресијом, пакетским губицима и додавањем Гаусовог шума. Због тога се може рећи да универзална објективна мера, тј. мера која ће бити употребљива код различитих типова деградације сигнала, за различите степене деградације, за различите примене и сл., тренутно не постоји.

Кључне речи: JPEG компресија, H.264 и H.265 видео-компресије, објективна процена квалитета слике и видеа.

Paper received on / Дата получения работы / Датум пријема чланка: 14.12.2016.
Manuscript corrections submitted on / Дата получения исправленной версии работы / Датум достављања исправки рукописа: 10.01.2018.
Paper accepted for publishing on / Дата окончательного согласования работы / Датум коначног прихватања чланка за објављивање: 12.01.2018.