Reduced Video Quality Measure Based on 3D Steerable Wavelet Transform and Modified Structural Similarity Index

Emil Dumic, Sonja Grgic
University of Zagreb, Faculty of Electrical Engineering and Computing
Department of Wireless Communications
emil.dumic@fer.hr

Abstract **- In this paper we present new reduced video quality measure (RVQM) based on 3D steerable wavelet transform and modified SSIM (Structural Similarity index) measure. RVQM is compared with other full and reduced reference quality measures and tested on LIVE video and LIVE mobile databases. Results show that proposed RVQM measure provides good correlation with subjective grades on both databases, while having fast calculation time.**

Keywords **- Riesz transform, SSIM, RVQM**

I. Introduction

Video quality evaluation plays an important role in video processing techniques, video compression, transmission, reproduction etc [1]. Video quality can be estimated using subjective or objective measures. Subjective measures are evaluated by making visual experiment under controlled conditions, in which human observers evaluate video sequence quality. Objective measures provide computed value of distortion severity and their main goal is good correlation with subjective grades. Depending on the development type, objective measures can be based on bottom-up or top-down approaches. In bottom-up approach, underlying premise is that the sensitivities of the human visual system (HVS) are different for different aspects of the visual signal that it perceives. Unlike these models, top-down approach is not affected by assumptions about HVS models, but is motivated instead by the need to capture the loss of visual structure in the signal that the HVS hypothetically extracts for cognitive understanding.

Objective quality measures, according to the reference information they are using, can be divided:

- full reference quality measures,
- reduced reference quality measures,
- no reference quality measures.

Classification and comparison of different already existing objective video quality assessment algorithms can be found in [2]. Generally, any image quality metric can be extended to video sequence analysis by calculating the average value of image quality for every frame. These metrics include well known PSNR (Peak Signal to Noise Ratio), SSIM (Structural Similarity index) [3], VSNR (Visual Signal to Noise Ratio) [4] etc. Since mentioned image based measures are not able to capture temporal distortions, different full reference or reduced reference video quality measures for simultaneously capturing of spatial and temporal distortions have been developed. MOVIE (Motion-based Video Integrity Evaluation) index [5] is a full-reference measure that utilizes a general, spatio-spectrally localized multiscale framework for evaluating dynamic video fidelity that integrates both spatial and temporal (and spatio-temporal) aspects of distortion assessment. VQM (Video Quality Measure), introduced by The National Telecommunications and Information Administration (NTIA) [6], is a reduced reference video quality measure that utilizes reduced reference parameters that are extracted from optimally-sized spatial-temporal regions of the video sequence. STRRED (Spatio-temporal Reduced Reference Entropic Differences) [7] is another measure with variable range of the reference information (from single scalar to the full reference information). It combines spatial RRED index [8], (SRRED) and temporal RRED index (TRRED).

In this paper we propose a novel video quality measure that uses different percent of original video pixels, RVQM (Reduced Video Quality Measure) which is based on top-down approach. Its details will be explained in Section II. Section III briefly describes tested video quality databases and their subjective scores. In Section IV correlation results will be presented for RVQM as well as for some other full and reduced reference quality measures. Section V draws the conclusion.

II. Reduced video quality measure (RVQM)

A. Modified SSIM and 3D steerable wavelet transform

Structural Similarity index (SSIM) is a method for measuring the similarity between two images [3]. It is computed from three image measurement comparisons: luminance, contrast and structure. At each calculation step, the local statistics and SSIM index are calculated within the local window. Resulting SSIM index map often exhibits undesirable "blocking" artifacts, so each window is filtered with normalized Gaussian weighting function (11x11 pixels) prior calculation of the three mentioned components. Luminance, contrast and structure terms can be calculated as:

$$SSIM\_lum = \frac{2 \cdot \mu_x \cdot \mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1} \tag{1}$$

$$SSIM\_cont = \frac{2\cdot\sigma_x\cdot\sigma_y + C_2}{\sigma_x^2+\sigma_y^2+C_2} \qquad (2)$$

$$SSIM\_struct = \frac{\sigma_{xy} + \frac{C_2}{2}}{\sigma_x\cdot\sigma_y + \frac{C_2}{2}} \qquad (3)$$

μ_x and μ_y are weighted mean values from original and degraded image blocks with size 11x11 pixels, using Gaussian weighting function. σ_x^2 and σ_y^2 are weighted variances and σ_{xy} is weighted covariance. C_1 and C_2 are constants defined as $C_1=(K_1L)^2$ and $C_2=(K_2L)^2$ where K_1 and K_2 are constants experimentally determined (K_1=0.01 and K_2=0.03) to improve measure stability when denominator in (1), (2) or (3) is close to zero. L is defined as maximal luminance (255 for 8 bit frames). The final local SSIM measure is the product of (1), (2) and (3).

We alter SSIM measure in the sense that only contrast and structure terms are calculated as concluded in [9] in 3 dimensions:

$$SSIM\,\mathrm{mod} = \frac{2\cdot\sigma_{xy} + C_2}{\sigma_x^2+\sigma_y^2+C_2} \qquad (4)$$

Applied Gaussian weighting function did not influence on correlation with subjective scores of video sequences so only averaging filter with different size will be used.

Since the human visual system percepts distortions differently at different frequency scales, it is preferable to compute the quality measure from coefficients obtained by using some type of transform. In calculating RVQM measure we used 3D steerable wavelets [10], which are obtained by applying a 3D version of the generalized Riesz transform to a primary isotropic wavelet frame. The novel transform is self-reversible (tight frame) and its elementary constituents (Riesz wavelets) can be efficiently rotated in any 3D direction by forming appropriate linear combinations. It may be seen as a generalization of Simoncelli's steerable pyramid [11] (which is used and successfully implemented in different image quality measures), giving access to a larger palette of steerable wavelets via a suitable parameterization. The Matlab code for this transform [12] for now supports only isotropic volumes, with size 2^n, which can be problematic because third dimension, time, may have different impact than spatial dimensions (height and width). Results show that even cube with specific size can be used to achieve good correlation results. While computing wavelet transform, volume is firstly prefiltered using lowpass and highpass prefilter (Simoncelli, Shannon or no prefiltering could be used in tested Matlab wavelet package). Lowpass scale can then be filtered using other filter: Simoncelli, Meyer, Papadakis, Aldroubi or Shannon filters are available in tested Matlab wavelet package. The transform can be calculated using different Riesz order. N-th order produces $(N + 2)\cdot(N + 1)/2$ components at each scale.

B. *Building the RVQM measure*

Using 3D steerable wavelet transform, original and distorted video sequence can be divided in different sizes and afterwards transformed. Each part should be isotropic, so tested sizes of cubes were 16, 32, 64 and 128 pixels. Cubes were extracted from original video sequences by calculating nearest neighbor interpolation for each frame and in third dimension every frame was used, disregarding sequence's fps type (25 or 50 fps). Other interpolation method (e.g. bilinear) did not produce any better correlation results, but made the algorithm slower.

The LIVE video quality database [13] was used to optimally design RVQM. Best correlation was obtained using Shannon prefilter, Simoncelli filter (although other filters produce nearly the same correlation results) and only third component from first Riesz order decomposition. From each component, modified SSIM measure was calculated using (4) and final measure for this block is the product of modified SSIM measures at each scale. Averaging filter was set to 8, 4 or 2 pixels in each direction. Step was set to half the size of tested cube (e.g. 8, 16, 32 and 64 for sizes 16, 32, 64 and 128) and final RVQM was calculated as average measure of all measures from each tested cube. RVQM calculation diagram is shown in Fig. 1.

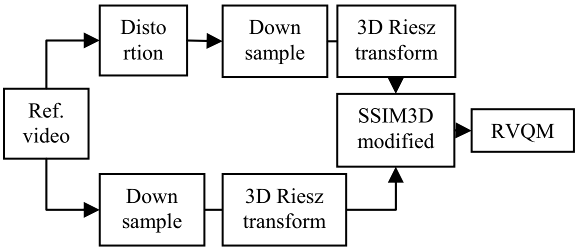

Figure 1. RVQM calculation diagram

III. SUBJECTIVE VIDEO DATABASES

For developing and testing of RVQM, the LIVE video quality database [13] was used (later called LIVE video). The database consists of 10 original video sequences, each having resolution of 768x432 pixels and 150 distorted video sequences (15 distorted sequences per one original) with 4 distortion types:

- Wireless distortions (four test videos per reference);
- IP distortions (three test videos per reference);
- H.264 compression (four test videos per reference);
- MPEG-2 compression (four test videos per reference).

Original video sequences have 8 bit planar YUV 4:2:0 format, while distorted video sequences have been converted back to the same format as the original. Six sequences have 250 frames (25 fps), one has 217 frames (25 fps) and three have 500 frames (50 fps).

The subjective study was conducted using a single stimulus procedure and the subjects indicated the video quality on a continuous scale. Subjects viewed each of the reference videos to facilitate computation of difference scores using hidden reference removal. Each video was viewed by 38 subjects. 9

subjects out of 38 were unreliable according to specifications in ITU-R BT 500.11 and the subjective data is provided from 29 valid subjects in the form of DMOS scores (Difference Mean Opinion Score).

To be able to test if a developed RVQM produces good correlation result, second video database was tested, LIVE Mobile Video Quality Database [14] (later called LIVE mobile), using mobile study only (tablet study was here also performed, however it was not compared here). 10 video files have planar YUV 4:2:0 format, with spatial resolution of 1280x720 pixels and 30 fps, each 15 seconds long. It exists of 20 degraded video sequences and 5 distortion types (all together 200 distorted video sequences):

- H.264 Compression (four test videos per reference);
- Wireless channel packet-loss (four test videos per reference);
- Frame-freezes (four test videos per reference);
- Rate Adaptation (three test videos per reference);
- Temporal Dynamics (five test videos per reference).

Because frame-freezes produces video sequence that may be longer than reference video sequence (e.g. in the case of live video delivery) and it is unclear how to compare such video sequences, this type of distortion was skipped in later results section. This means that 160 degraded video sequences were used to compare different objective quality measures.

The subjective study was conducted using a single stimulus procedure and the subjects indicated the quality of the video on a continuous scale. Subjects also viewed each of the reference videos to facilitate computation of difference scores using hidden reference removal. The videos were displayed on the Motorola Atrix smartphone (Atrix 4-inch Gorilla glass display with a screen resolution of 960x540 pixels). YUV files had to be compressed (with compression >18 Mbit/s) and embedded into 3gp container. The study was designed so that 18 subjective ratings from 36 subjects were obtained for each of the 200 videos in the study and the subject rejection procedure in ITU-R BT 500.11 was used to reject two subjects from the study.

IV. Results

A. Spearman's correlation for LIVE video quality database

In this section, results for RVQM and other existing video quality measures will be compared with subjective grades using Spearman's rank order correlation coefficient [15]. Spearman's coefficient assesses how well an arbitrary monotonic function describes the relationship between two variables without making any assumptions about the frequency distribution of the variables. Spearman's correlation coefficient is calculated like Pearson's correlation (normalized covariance between two variables) over ranked variables. Rank of the sample in variable is its sorted location in a row. In the case of tied ranks, positions of all tied samples are calculated as an arithmetic mean of their ranks.

Results of RVQM_64 (cube dimension is 64 pixels with 4 pixel averaging filter) measure were compared with other full and reduced reference video quality measures and are presented in Table I for LIVE video database: PSNR (which was calculated from average MSE from all frames in one video sequence), SSIM [3], MS-SSIM [16], VSNR [4], VQM [6], MOVIE [5] and STRRED [7] (Spearman's correlation results for VQM and MOVIE measures are taken from [7]). Best values are bolded. Depending on the reference information percentage, results are also presented in the Table I. Percentage is calculated as ratio between pixels per frame needed to calculate video quality measure and original size of video frames (disregarding if they are of type integer or real number). Time represents average timing for measure calculation for all 150 distorted video sequences, without importing them in memory. It should be noted that six video sequences have 250 frames, one has 217 frames and three have 500 frames (meaning 3217 frames in all video sequences), therefore timing was averaged for 1 frame. The average number of frames per video sequence is 321.7 frames. Depending on the step size (which was in all cases half the size of the cube) some of the frames were not used. However, this was not taken into account for average time per frame. Reading luminance part from YUV 420 average video sequence (consisting of 321.7 frames) has taken about 6.2s (or 19 ms per frame). Computer configuration which was used for calculating was: Intel Q6600 @2400 MHz, 4 GB RAM, Windows 7 with Matlab program. Results for VQM and MOVIE index were taken from [7] so timing comparison with these measures was skipped.

In Table II results for different cube sizes (directly influencing the percentage of reference information) and averaging filter sizes are presented. Best values are bolded. Percentage is calculated like in Table I.

TABLE I. Comparison of different video quality measures with RVQM measure

	PSNR	SSIM	MS-SSIM	VSNR	VIF	VQM	MOVIE	STRRED	RVQM_64
Wireless	0.6574	0.6553	0.7289	0.6951	0.5317	0.7214	0.8114	0.7857	**0.8317**
IP errors	0.4167	0.6182	0.6534	0.6930	0.5506	0.6383	0.7192	**0.7722**	0.7517
H.264	0.4585	0.7129	0.7313	0.6405	0.6349	0.6520	0.7797	**0.8193**	0.8017
MPEG-2	0.3862	0.6652	0.6684	0.5874	0.6331	0.7810	**0.8170**	0.7193	0.6654
Overall	0.5398	0.6947	0.7364	0.6726	0.5541	0.7026	**0.8055**	0.8007	0.8038
Percentage	100%	100%	100%	100%	100%	~15%	100%	0.17%	1.23%
Time (s) - per video sequence	2.2	13.8	69.7	30.1	757.2	-	-	330.3	10.0
Time (s) - per frame	0.007	0.043	0.217	0.094	2.35	-	-	1.02	0.031

TABLE II. RVQM WITH DIFFERENT CUBE SIZE AND AVERAGING FILTER SIZE

	RVQM_128	RVQM_128	RVQM_128	RVQM_64	RVQM_64	RVQM_64	RVQM_32	RVQM_32	RVQM_32	RVQM_16	RVQM_16
Averaging filter size	12	8	4	8	4	2	8	4	2	4	2
Wireless	0.8246	0.8244	0.8212	0.8259	0.8317	0.8236	0.8098	**0.8336**	0.8184	0.7257	0.7563
IP errors	**0.7539**	0.7362	0.7290	0.7477	0.7517	0.7099	0.7308	0.7313	0.7143	0.6641	0.6347
H.264	0.7842	0.7831	0.7711	**0.8167**	0.8017	0.7794	0.8060	0.8062	0.7841	0.7660	0.7396
MPEG-2	0.7196	0.7198	**0.7206**	0.6731	0.6654	0.6146	0.6258	0.6270	0.5834	0.5600	0.5086
Overall	**0.8100**	0.8085	0.8003	0.8058	0.8038	0.7723	0.7760	0.7828	0.7566	0.7092	0.6960
Percentage	4.94%	4.94%	4.94%	1.23%	1.23%	1.23%	0.31%	0.31%	0.31%	0.08%	0.08%
Time (s) - per video sequence	154.5	54.7	30.1	17.6	10.0	9.1	5.3	3.1	2.8	1.7	1.5
Time (s) - per frame	0.480	0.170	0.094	0.055	0.031	0.028	0.016	0.010	0.009	0.005	0.005

B. Comparison of overall Spearman's correlation for both tested databases

Comparison of LIVE video and LIVE mobile databases are presented in Fig. 2, while timing calculations from Table 1 are shown in Fig. 3. LIVE mobile database was not used in the building of RVQM measure.

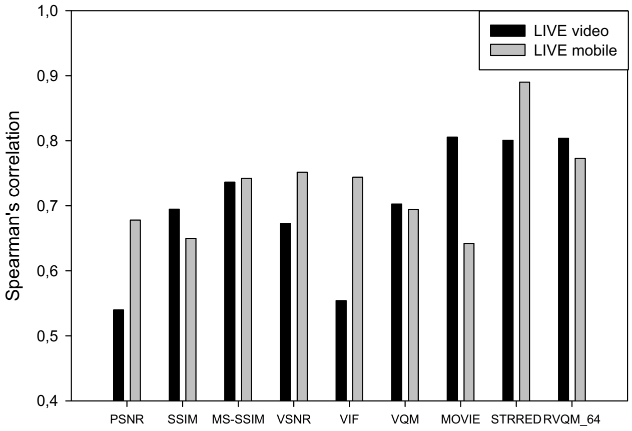

Figure 2. Overall Spearman's correlation for nine objective measures and 2 video databases

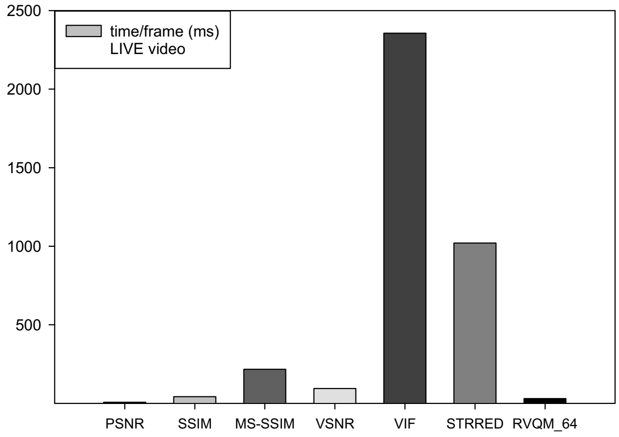

Figure 3. Timing calculation for seven objective measures per one frame, LIVE video database

C. Discussion of the results

In Table II it can be seen that optimal cube size was 64 or 32 pixels with 4 pixel averaging filter. RVQM_64 was compared with other video quality measures and produced similar results like STRRED or MOVIE algorithms. In comparison with MOVIE index, the RVQM_64 needs about 1.23% of reference information and compared with STRRED (downloaded from [17]), RVQM can be calculated much faster. STRRED needed on average 330.3s per video sequence (or 1.02s per frame, including reading frames into the memory) on the same computer configuration, which is much slower than proposed RVQM for any cube size. The STRRED can be used with different amount of reference information and also as a single number algorithm (one number of reference information per frame), unlike the RVQM which needs more information per frame. Comparing RVQM with same cube size and different averaging filer size, it can be seen that each cube size has its optimal averaging filter size which is about 8-16 time smaller than the cube dimension.

Comparing specific distortions in LIVE video database shows that the RVQM_64 has the best correlation for wireless distortions. In case of IP and H.264 distortion, the RVQM_64 has the second best correlation, while the best correlation is gained with STRRED. Comparison in case of MPEG-2 distortion shows possible improvement of overall correlation by improving correlation with MPEG-2 distortions since correlation for RVQM_64 is the lowest in this specific case.

For real time applications, RVQM_32 (4 pixels averaging filter) could be used for 25 fps sequences, while 50 fps have near real time calculation for RVQM_32 and RVQM_64 (with added 19 ms time needed to read sequence in memory) and using mentioned computer configuration. However, it could be possible to further reduce calculation time (e.g. by computing RVQM and reading frames simultaneously), thus making it useful in any real time scenario.

It is useful to compare correlation between objective and subjective measures on different databases. As LIVE mobile database uses different viewing screen (mobile), video sequences, frame rates, resolution and some different degradation types, it can be used as a tested database to check correlation from the LIVE video database. Measure that correlates well with subjective grades should have nearly equal (stable) correlation in any video database. From the Fig. 2 these measures could be MS-SSIM, VQM and RVQM_64. MS-SSIM represents robust image quality measure [18] so it obviously performs fairly well in video databases. VQM represents video measure that gives stable results, however they're even lower than some of the image quality measures. MOVIE index gives good results in LIVE video database, but its correlation drops considerably in LIVE mobile database. RVQM_64 gives nearly stable results in both video databases and are better than any of the image measures tested and most video measures, only STRRED gives better results in LIVE mobile database (from the tested quality measures). STRRED gives good results in LIVE video database, but excellent results in LIVE mobile database so it should be tested further. However, it could yet not be used in some real time application, Fig. 3. Comparison of specific degradation types on LIVE mobile database was not performed yet.

V. Conclusion

In this paper, simple and efficient new reduced video quality measure (RVQM) was proposed. The new measure produces similar correlation results with state of the art full and other reduced video quality measures, while having lower computational time per frame that provides real time calculation. None of the existing algorithms perform excellent on both tested video databases so more effort should be directed towards obtaining better correlation with subjective testing.

In future research, proposed RVQM measure should be tested on different video databases to check whether correlation will be similar to the LIVE video and mobile databases. Also, RVQM with different parameters could be computed on LIVE mobile database. It would be useful to determine optimal cube size for RVQM measure in proportion with frame size. E.g. from the LIVE video database results, this ratio could be about 1:10 since percentage of the reference information for RVQM_64 is 1.23%. Research can be directed to some specific distortion type or maybe to some real time application. 3D objective measure could also be developed, using RVQM for left and right view independently, or by combining each view in one 3D cube for each frame. It may also be interesting to develop objective measure for frame-freezing degradation type (only for this type or combined with other types of degradation), where frame length of original and degraded sequence is not equal.

Acknowledgment

The work described in this paper was conducted under the research project "Picture Quality Management in Digital Video Broadcasting" (036-0361630-1635), supported by the Ministry of Science, Education and Sports of the Republic of Croatia.

References

[1] H. R. Wu and K. R. Rao, "Digital Video Image Quality and Perceptual Coding", CRC Press, 2005

[2] S. Chikkerur, V. Sundaram, M. Reisslein, and L. J. Karam, "Objective video quality assessment methods: A classification, review, and performance comparison", IEEE Transactions on Broadcasting, vol. 57, no. 2, pp. 165-182, Feb. 2011

[3] Z. Wang, A.C. Bovik, H.R. Sheikh, E.P. Simoncelli, "Image Quality Assessment: From Error Visibility to Structural Similarity", IEEE Trans. on Image Proc., Vol. 13, No. 4, pp. 600-612, 2004

[4] 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, No. 9, pp. 2284-2298, 2007

[5] K. Seshadrinathan and A. C. Bovik, "Motion Tuned Spatio-temporal Quality Assessment of Natural Videos", vol. 19, no. 2, pp. 335-350, IEEE Transactions on Image Processing, Feb. 2010

[6] M. H. 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

[7] Rajiv Soundararajan and Alan C. Bovik, "Video Quality Assessment by Reduced Reference Spatio-temporal Entropic Differencing", IEEE Trans. Image Process., vol. 23, no. 4, pp. 684 - 694, April 2013

[8] R. Soundararajan and A. C. Bovik, "RRED indices: Reduced reference entropic differencing for image quality assessment", IEEE Trans. Image Process., vol. 21, no. 2, pp. 517–526, Feb. 2012

[9] E. Dumic, I. Bacic and S. Grgic, "Simplified structural similaritxy measure for image quality evaluation", 19th International Conference on Systems, Signals and Image Processing IWSSIP-2012, pp. 456-461, 2012

[10] N. Chenouard, M. Unser, "3D Steerable Wavelets in practice", IEEE Transactions on Image Processing, Vol. 21, Num. 11, pp 4522-4533, Nov 2012

[11] E. P. Simoncelli, W. T. Freeman, "The Steerable Pyramid: A Flexible Architecture for Multi-Scale Derivative Computation", 2nd IEEE International Conference on Image Processing, Vol. 3, pp. 444-447., 1995

[12] http://bigwww.epfl.ch/demo/steerable-wavelets-3d/

[13] K. Seshadrinathan, R. Soundararajan, A. C. Bovik, and L. K. Cormack, "Study of subjective and objective quality assessment of video", IEEE Trans. Image Process., vol. 19, pp. 1427–1441, Jun. 2010

[14] A. K. Moorthy, L. K. Choi, A. C. Bovik and G. deVeciana, "Video Quality Assessment on Mobile Devices: Subjective, Behavioral and Objective Studies", IEEE Journal of Selected Topics in Signal Processing, vol.6, no.6, pp. 652-671, October 2012

[15] J. Hauke and T. Kossowski, "Comparison of values of Pearson's and Spearman's correlation coefficient on the same sets of data", Proceedings of the MAT TRIAD 2007 Conference, Bedlewo, Poland, 2007

[16] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multiscale structural similarity for image quality assessment", in Proc. 37th Asilomar Conf. on Signals, Systems and Computers, Pacific Grove, CA 2003

[17] http://live.ece.utexas.edu/research/quality/strred.rar

[18] N. Ponomarenko, V. Lukin, A. Zelensky, K. Egiazarian, M. Carli, F. Battisti, "TID2008 - A Database for Evalu-ation of Full-Reference Visual Quality Assessment Metrics", Advances of Modern Radioelectronics, vol. 10, pp. 30-45, 2009, http://www.ponomarenko.info/tid2008.htm, access: 2011.