INTERNATIONAL ORGANISATION FOR STANDARDISATION
ORGANISATION INTERNATIONALE DE NORMALISATION
ISO/IEC JTC 1/SC 29/AG 5

ISO/IEC JTC 1/SC 29/AG 5 m57118
July 2021, Online

Title: **Performance of full-reference objective metrics in immersive video**
Source: **Adrian Dziembowski, Dawid Mieloch, Jakub Stankowski**
Poznań University of Technology

Abstract

This informative contribution presents the results of the experiments on assessing the performance of several state-of-the-art full-reference objective quality metrics in immersive video applications. The results presented in this document were obtained for the purposes of the journal paper about the IV-PSNR quality metric (currently under review).

1 Experiments

1.1 Evaluated quality metrics

1. IV-PSNR [WG04 N0013],
2. $PSNR_Y$ – PSNR for luma component,
3. $PSNR_{YUV}$ – weighted average of PSNR for 3 color components with luma weight 6 times higher than weights for both chroma components, as described in [611],
4. WS-$PSNR_Y$ – weighted-to-spherically uniform version of $PSNR_Y$, adapted for ERP videos [WS-PSNR],
5. WS-$PSNR_{YUV}$ – WS- version of $PSNR_{YUV}$,
6. VMAF – Video Multimethod Assessment Fusion [VMAF],
7. SSIM – Structural Similarity index [SSIM],
8. MS-SSIM – Multi-Scale version of SSIM [MS-SSIM],
9. VIF-P – the pixel-based version of VIF, Visual Information Fidelity metric [VIF-P],
10. PSNR-HVS – modified PSNR, which considers the Human Visual System properties [PSNR-HVS],
11. PSNR-HVS-M – PSNR-HVS modified by considering the visual masking of DCT coefficients [PSNR-HVS-M],
12. SFF – Sparse Feature Fidelity index [SFF],
13. PSNR-HA – PSNR-HVS modified by considering change of contrast and mean value [PSNR-HA],
14. PSNR-HMA – PSNR-HVS-M modified in the same way [PSNR-HA],
15. VSNR – wavelet-based Visual Signal-to-Noise Ratio [VSNR],
16. WSNR – Weighted Signal-to-Noise Ratio [WSNR],
17. SAM – Spectral Angle Mapper [SAM],
18. SRE – Signal to Reconstruction Error ratio [SRE],
19. FSIM – Feature-based Similarity index [FSIM],

20. UIQ – Universal Image Quality [UIQ],
21. MP-PSNR-F – Morphological Pyramid PSNR, full variant [MP-PSNR-F],
22. MP-PSNR-R – reduced version of MP-PSNR [MP-PSNR-R],
23. MW-PSNR-F – Morphological Wavelet PSNR, full variant [MW-PSNR],
24. MW-PSNR-R – reduced version of MW-PSNR [MP-PSNR],
25. 3DSwIM – 3D Synthesized view Image Metric [3D-SWIM],
26. LPIPS – Learned Perceptual Image Patch Similarity [LPIPS].

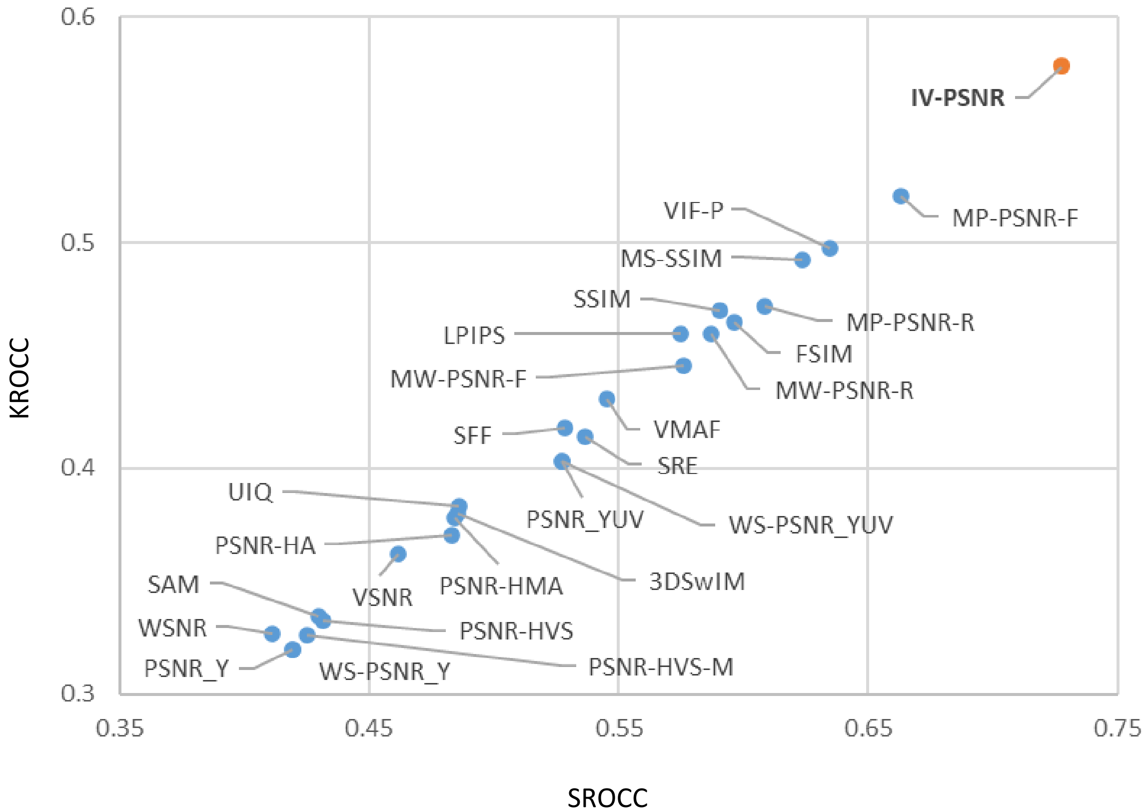
1.2 Methodology

For each quality metric, the correlation between its results and the subjective quality was estimated. As some metrics do not linearly correlate with MOS, we have used two rank-based coefficients: KROCC and SROCC (Kendall and Spearman rank correlation coefficients), which evaluate the monotonicity of the relationship between two datasets.

We did not implement any metric, for all quality metrics, existing implementations were used. However, some of them have to be wrapped in order to properly handle yuv files with multiple frames.

1.3 Immersive video coding: CfP on 3DoF+ Visual [WG11 N18145]

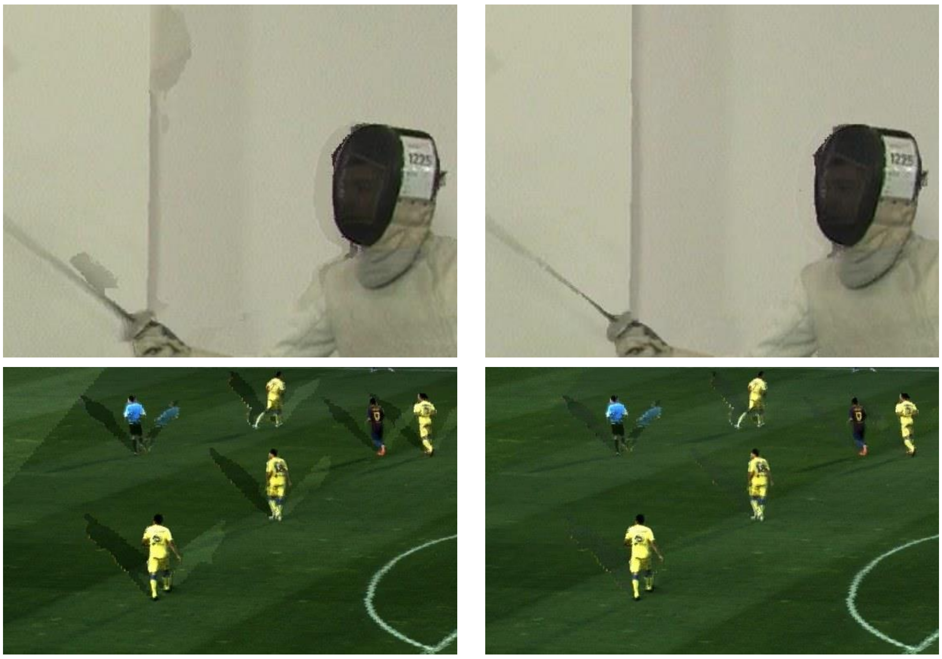
1. 7 different immersive video coding techniques (2 anchors + 5 proposals: [m47407], [m47372], [m47179], [m47445], [m47684]),
2. reliable subjective tests: [WG11 N18353],
3. 5 test sequences, 4 rate points,
4. subjective quality evaluated on posetraces, objective – on a subset of synthesized input views.



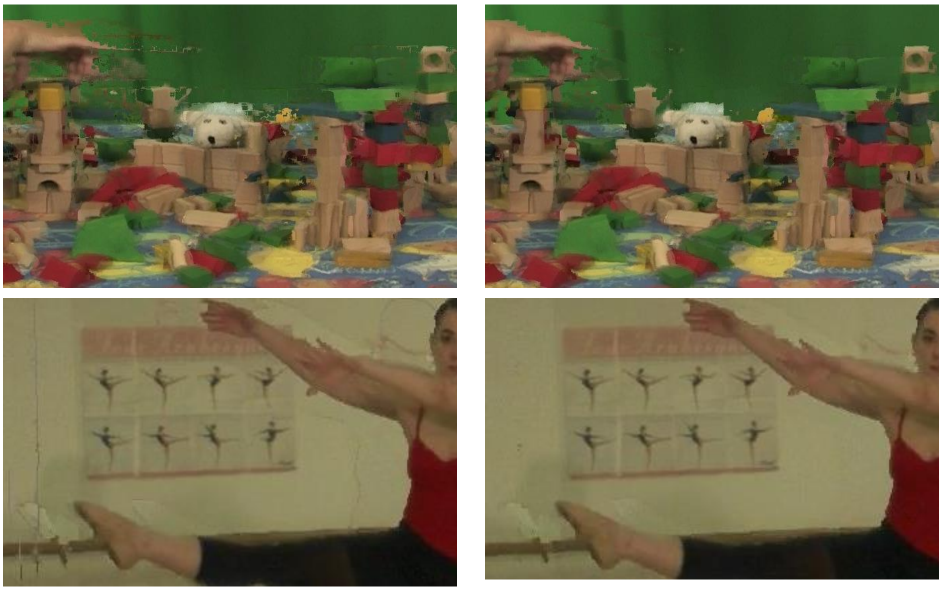
1.4 Immersive video processing – 3 common processing types

1. different synthesizers: VSRS [m40657] vs. PUT’s synthesizer [MVS]):

2. color correction: with vs. without:

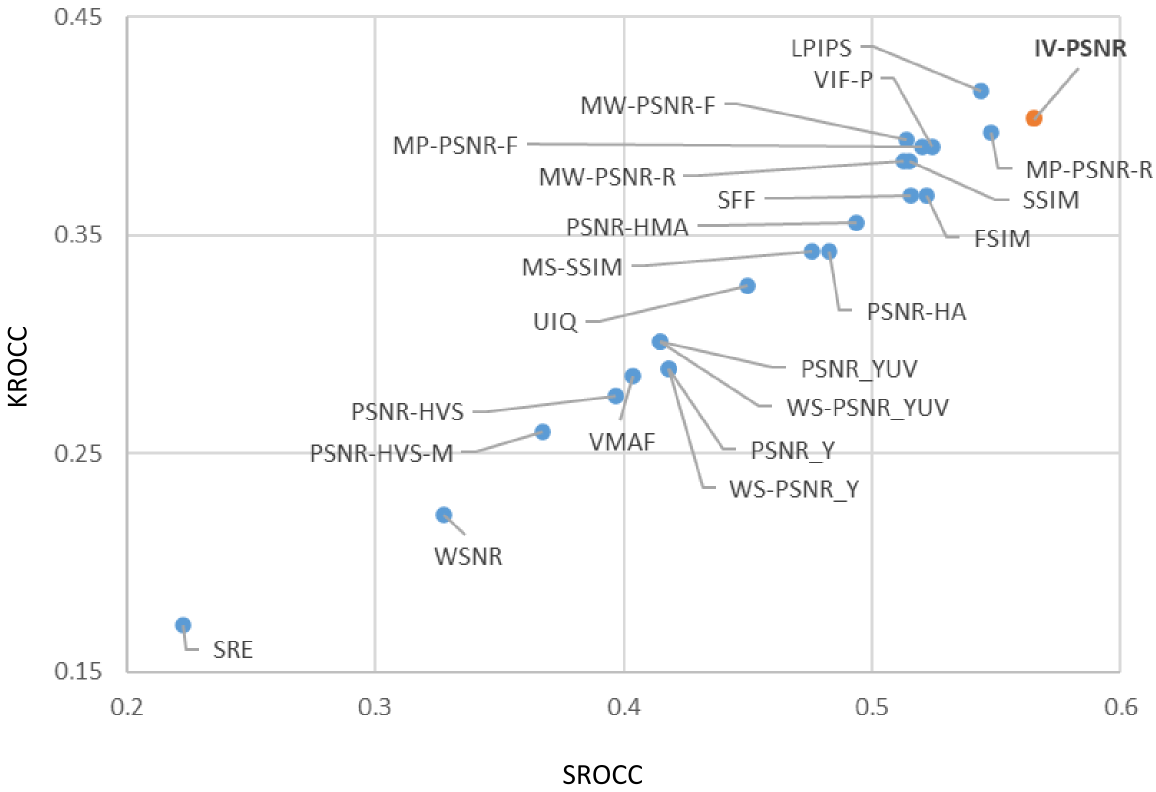


3. filtration during synthesis:



- 44 naïve viewers, PairComparison method with scale [-3, 3],
- 12 test sequences.

As we did not use the ACR method, but the differential one, the correlation between objective and subjective quality was calculated using quality differences, not the absolute values. For example, for an influence of the color correction, the viewers were comparing two videos (side-by-side), one synthesized with color correction and one without color correction. For objective quality metrics, the difference was also calculated, e.g. for PSNR the ΔPSNR was calculated as: $\Delta PSNR = PSNR_{cc} - PSNR_{nocc}$.



1.5 Other applications: TID2013 database [TID2013]

- non-immersive video applications,
- 24 different types of distortions, 5 magnitudes for each,
- 25 test images (static), low resolution.

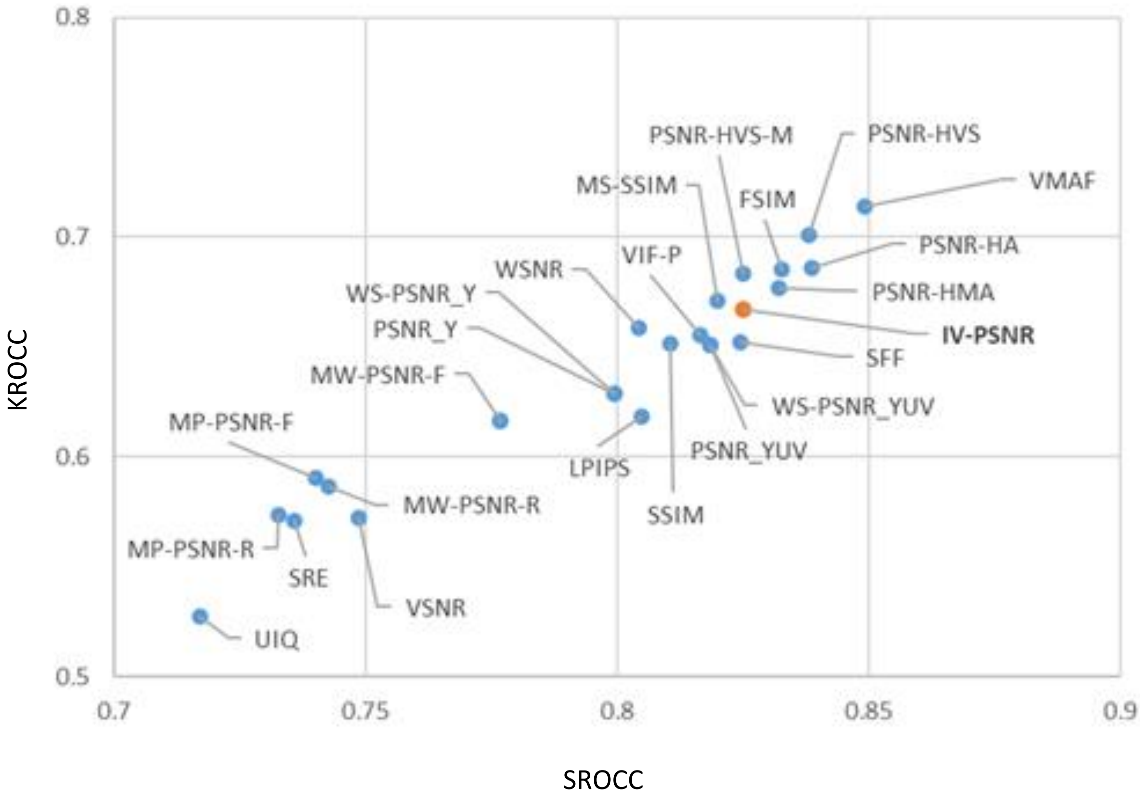


Table. Ranks for all considered metrics (the best metric for each distortion type is highlighted); WS-PSNR$_Y$ and WS-PSNR$_{YUV}$ metrics were skipped, as for perspective content they perform identically to PSNR$_Y$ and PSNR$_{YUV}$.

	Distortion type \ Quality metric	**IV-PSNR**	PSNR$_Y$	PSNR$_{YUV}$	VMAF	SSIM	MS-SSIM	VIF-P	PSNR-HVS	PSNR-HVS-M	SFF	PSNR-HA	PSNR-HMA	VSNR	WSNR	SAM	SRE	FSIM	UIQ	MP-PSNR-R	MP-PSNR-F	MW-PSNR-R	MW-PSNR-F	3DSwIM	LPIPS
#	Average SROCC	6	15	10	1	12	9	11	3	7	8	2	5	17	14	24	20	4	22	21	19	18	16	23	13
1	Additive Gaussian noise	1	8	7	11	18	17	19	4	9	12	2	6	20	15	24	13	14	22	10	16	5	3	23	21
2	Noise in color comp.	1	2	4	8	18	19	17	3	13	15	7	11	20	10	24	12	14	22	9	16	6	5	23	21
3	Spatially correl. noise	2	9	8	11	16	18	19	7	5	13	1	3	21	15	24	12	14	22	10	17	6	4	23	20
4	Masked noise	14	4	2	19	7	8	3	15	20	6	9	11	21	22	24	5	12	18	16	17	10	1	23	13
5	High freq. noise	5	10	11	3	19	17	18	2	13	15	1	8	20	12	24	7	14	22	9	16	6	4	23	21
6	Impulse noise	9	6	3	8	14	16	11	5	10	15	2	7	17	4	24	1	12	13	23	21	22	20	18	19
7	Quantization noise	6	8	9	7	19	10	21	4	2	12	3	1	14	5	24	13	11	22	15	20	17	16	23	18
8	Gaussian blur	16	14	15	2	5	1	4	11	13	3	10	12	8	9	24	22	7	19	21	17	20	18	23	6
9	Image denoising	1	5	6	8	17	13	21	4	7	15	2	3	16	9	24	20	12	22	19	10	18	11	23	14
10	JPEG compression	14	16	6	8	19	9	17	1	12	13	2	4	21	11	24	20	10	22	7	18	3	5	23	15
11	JPEG2000 compression	18	21	15	7	19	8	13	4	2	6	3	1	16	9	24	22	5	20	10	14	12	17	23	11
12	JPEG transm. errors	17	18	19	3	10	4	8	13	14	2	7	6	16	20	24	23	5	21	12	15	9	11	22	1
13	JPEG2000 transm. errors	19	9	15	17	13	8	14	3	4	12	1	2	18	10	24	22	6	21	11	20	5	7	23	16
14	Non ecc. patt. noise	1	21	15	8	13	3	9	18	14	10	16	12	22	6	24	7	4	20	5	19	11	17	23	2
15	Local block-wise dist.	22	15	19	7	1	6	5	13	17	14	18	20	16	23	21	24	4	3	12	10	11	9	2	8
16	Mean shift	9	7	8	18	1	3	17	6	5	16	13	12	21	2	24	19	14	15	20	10	22	4	23	11
17	Contrast change	10	14	16	1	11	8	2	12	15	9	4	3	20	17	24	23	7	22	19	5	18	6	21	13
18	Change of color saturation	4	18	13	12	9	10	8	5	6	1	14	15	16	7	19	17	11	2	21	20	23	22	24	3
19	Multipl. Gaussian noise	2	7	8	10	17	15	18	5	9	13	1	4	19	11	24	3	12	22	16	21	14	6	23	20
20	Comfort noise	16	18	5	4	21	17	19	2	8	10	1	6	15	7	24	3	9	22	13	20	12	14	23	11
21	Lossy compr. of noisy images	4	15	12	1	17	16	19	6	3	10	9	2	20	5	24	7	8	22	14	21	13	11	23	18
22	Image color quant. w. dither	1	3	6	13	18	15	22	2	7	10	4	5	14	8	24	21	11	23	20	16	17	12	9	19
23	Chromatic aberrations	21	3	12	9	2	4	1	7	16	8	19	20	14	17	24	23	13	15	18	5	10	11	22	6
24	Sparse sampl. and reconstr.	5	19	13	10	20	8	14	4	2	7	3	1	12	9	24	21	6	22	18	17	16	15	23	11

Acknowledgement

The research was supported by the Ministry of Science and Higher Education of Republic of Poland.

The Authors would like to thank Vittorio Baroncini from EVATech and Giacomo Baroncini from GBTech for providing the results of the subjective quality evaluation of MPEG Call for Proposals. The Authors would like to thank also the proponents of MPEG CfP: Bart Kroon and Bart Sonneveldt (Philips Research Eindhoven), Vinod Kumar Malamal Vadakital (Nokia), Lu Yu and Bin Wang (Zhejiang University), Gérard Briand, Julien Fleureau, and Renaud Doré (Technicolor/ InterDigital) for providing executables and/or videos generated using their proposals for the evaluation of the IV-PSNR metric.

References

[WS-PSNR] Y. Sun, A. Lu, and L. Yu, "Weighted-to-Spherically-Uniform Quality Evaluation for Omnidirectional Video," IEEE Signal Processing Letters, vol. 24, no. 9, Sep. 2017.

[611] Y. Huang, H. Qi, B. Li, and J. Xu, "Adaptive weighted distortion optimization for video coding in RGB color space," *IEEE International Conference on Image Processing, ICIP 2014*, Paris, France, Oct. 2014.

[VMAF] Z. Li, A. Aaron, I. Katsavounidis, A. Moorthy, and M. Manohara, "Toward a practical perceptual video quality metric," Netflix Technology Blog, Tech. Rep., 2016.

[SSIM] Z. Wang, A.C. Bovik, H.R. Sheikh, and E.P. Simoncelli, "Image quality assessment: From error measurement to structural similarity," IEEE Transactions on Image Processing, vol. 13, Jan. 2004.

[MS-SSIM] Z. Wang, E.P. Simoncelli, and A.C. Bovik, "Multiscale structural similarity for image quality assessment," The 37th Asilomar Conference on Signals, Systems & Computers, Pacific Grove, USA, Nov. 2003.

[VIF-P] H.R. Sheikh and A.C. Bovik, "Image information and visual quality," IEEE Transactions on Image Processing, vol. 15, pp. 430-444, 2006.

[PSNR-HVS] K. Egiazarian, J. Astola, N. Ponomarenko, V. Lukin, F. Battisti, and M. Carli, "New full-reference quality metrics based on HVS," 2nd International Workshop on Video Processing and Quality Metrics, Scottsdale, USA, 2006.

[PSNR-HVS-M] N. Ponomarenko, F. Silvestri, K. Egiazarian, M. Carli, J. Astola, and V. Lukin, "On between-coefficient contrast masking of DCT basis functions," 3rd International Workshop on Video Processing and Quality Metrics, Scottsdale, USA, 2007.

[SFF] H.W. Chang, H. Yang, Y. Gan, and M.H. Wang, "Sparse feature fidelity for perceptual image quality assessment," IEEE Transactions on Image Processing, vol. 22, pp. 4007-4018, 2013.

[PSNR-HA] N. Ponomarenko, O. Ieremeiev, V. Lukin, K. Egiazarian, and M. Carli, "Modified image visual quality metrics for contrast change and mean shift accounting," 11th International Conference on the Experience of Designing and Application of CAD Systems in Microelectronics (CADSM), Polyana-Svalyava, Ukraine, pp. 305-311, 2011.

[VSNR] 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, 2007.

[WSNR] T. Mitsa and K.L. Varkur, "Evaluation of contrast sensitivity functions for the formulation of quality measures incorporated in halftoning algorithms," IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Minneapolis, USA, pp. 301-304, 1993.

[SAM] F.A. Kruse, A.B. Lefkoff, J.B. Boardman, K.B. Heidebrecht, A.T. Shapiro, P.J. Barloon, and A.F.H. Goetz, "The Spectral Image Processing System (SIPS) - Interactive Visualization and

Analysis of Imaging spectrometer Data," Remote Sensing of Environment, vol. 44, pp. 145-163, 1993.

[SRE] C. Lanaras, J. Bioucas-Dias, S. Galliani, E. Baltsavias, and K. Schindler, "Super-resolution of Sentinel-2 images: Learning a globally applicable deep neural network," ISPRS Journal of Photogrammetry and Remote Sensing, vol. 146, 2018.

[FSIM] L. Zhang, L.Zhang, X.Mou, and D.Zhang, "FSIM: a feature similarity index for image quality assessment," IEEE Transactions on Image Processing, vol. 20, pp. 2378-2386, 2011.

[UIQ] Z. Wang, A.C. Bovik, "A universal image quality index," IEEE Signal Processing Letters, vol. 9, pp. 81-84, 2002.

[MP-PSNR] D. Sandić-Stanković, D. Kukolj, P. Le Callet, "DIBR-synthesized image quality assessment based on morphological multi-scale approach," J Image Video Proc, vol. 4, 2016.

[MP-PSNR-F] D. Sandić-Stanković, D. Kukolj, and P. Le Callet, "DIBR synthesized image quality assessment based on morphological pyramids," *3DTV-CON Immersive and interactive 3D media experience over networks*, Lisbon, Portugal, Jul. 2015.

[MP-PSNR-R] D. Sandić-Stanković, D. Kukolj, and P. Le Callet, "Multi–Scale Synthesized View Assessment Based on Morphological Pyramids," *Journal of Electrical Engineering*, vol. 67 (1), pp. 1–9, 2016.

[MW-PSNR] D. Sandić-Stankowić, D. Kukolj, P. Le Callet, "DIBR synthesized image quality assessment based on morphological wavelets", *International Workshop on Quality of Multimedia Experience QoMEX*, Costa Navarino, Greece, May 2015.

[3D-SWIM] F. Battisti, E. Bosc, M. Carli, P. Le Callet, S. Perugia, "Objective image quality assessment of 3D synthesized views," Signal Processing: Image Communication, vol. 30, pp. 78-88, 2015.

[LPIPS] R. Zhang, P. Isola, A. A. Efros, E. Shechtman and O. Wang, "The Unreasonable Effectiveness of Deep Features as a Perceptual Metric," 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018, pp. 586-595.

[MVS] A. Dziembowski, A. Grzelka, D. Mieloch, O. Stankiewicz, K. Wegner, and M. Domański, "Multiview Synthesis – improved view synthesis for virtual navigation," 32nd Picture Coding Symposium, PCS 2016, Nürnberg, Germany, Dec. 2016.

[TID2013] N. Ponomarenko et al., "Image database TID2013: Peculiarities, results and perspectives," Signal Proc.: Image Comm., vol. 30, pp. 57-77, 2015.