INTERNATIONAL ORGANISATION FOR STANDARDISATION
ORGANISATION INTERNATIONALE DE NORMALISATION
ISO/IEC JTC1/SC29/WG11
CODING OF MOVING PICTURES AND AUDIO

ISO/IEC JTC1/SC29/WG11 MPEG2019/M48093
July 2019, Göteborg, Sweden

Source **Poznań University of Technology (PUT), Poznań, Poland**
Status **Input**
Title **[MPEG-I Visual] Objective quality metric for immersive video**
Author Adrian Dziembowski, Marek Domański

1 Introduction

This document presents a technical description of proposed objective quality metric adapted for Immersive Video applications, which correlates to subjective quality much higher, than PSNR.

2 Overview of the proposed technique

Two assumptions were stated before creating proposed objective quality metric:

1. It should be insensitive on typical artifacts appearing during virtual view synthesis.
2. It should be as simple as possible.

In order meet these requirements, we decided to modify PSNR quality metric. Because of a block-based character, we gave it a provisional name BSNR (Block-PSNR).

We have added two major modifications to regular PSNR:

1. Corresponding pixel shift.
2. Global color shift.

2.1 Corresponding pixel shift

Because of rounding errors, edges of the objects in the virtual view may be little shifted when compared to the real view. This phenomenon significantly reduces the PSNR value, but it is unnoticeable for the viewer.

Typically, PSNR value is calculated as:

$$\mathrm{PSNR} = 10 \cdot \log\left(\frac{\mathrm{MAX}^2}{\mathrm{MSE}}\right),$$

where:

$$\mathrm{MSE} = \frac{1}{W \cdot H} \sum_{y=0}^{H-1} \sum_{x=0}^{W-1} \left(c_T(x,y) - c_R(x,y)\right)^2,$$

where: W and H are width and height of the image, MAX is the maximum value of the color component (e.g. 1023 for 10-bit video), $c_T(x, y)$ and $c_R(x, y)$ are values of color component in the position (x, y) in test image and reference image, respectively.

Proposed BSNR value is calculated in a very similar way:

$$\mathrm{BSNR} = 10 \cdot \log\left(\frac{\mathrm{MAX}^2}{\mathrm{BMSE}}\right),$$

where:

$$\mathrm{BMSE} = \frac{1}{W \cdot H} \sum_{y=0}^{H-1} \sum_{x=0}^{W-1} \min_{\substack{x_R \in \left[x-\frac{B}{2}, x+\frac{B}{2}\right] \\ y_R \in \left[y-\frac{B}{2}, y+\frac{B}{2}\right]}} \left(c_T(x, y) - c_R(x_R, y_R)\right)^2 ,$$

where B is the size of analyzed block in the reference view. For each pixel in the test image the most similar pixel within that colocated block in the reference image is chosen. Then, the squared component difference is calculated for such pair of pixels.

In the experiments we chose $B = 5$, what corresponds to 2-pixel shift of objects' edges.

We have also tried different scenarios, where the pixel shift is globally or locally optimized in order to preserve consistency of the shift and eliminate random shift directions for neighboring pixels. However, at this moment it is an academic research and in this document we want to report, that presented, very simple method allows to produce surprisingly good results.

2.2 Global color shift

Different input views may have various color characteristics, e.g. one of them is slightly darker. When such a view is reprojected to the position of another input view, there is a slight color difference for all the pixels. Obviously, it significantly decreases PSNR value. However, if all the color change for all the pixels is similar, the viewer does not notice it.

Therefore, we proposed to consider this phenomenon when estimating quality of the virtual view.

At the first step, for each color component we calculate the global difference between two images:

$$\mathrm{GCD} = \frac{1}{W \cdot H} \sum_{y=0}^{H-1} \sum_{x=0}^{W-1} \min_{\substack{x_R \in \left[x-\frac{B}{2}, x+\frac{B}{2}\right] \\ y_R \in \left[y-\frac{B}{2}, y+\frac{B}{2}\right]}} \left(c_T(x, y) - c_R(x_R, y_R)\right) ,$$

searching for most similar pixel within a $B \times B$ block in the reference view for each pixel in the test view. GCD computation is performed separately for all the color components.

Then, color of each pixel in the test view is modified by:

$$c_T{}'(x, y) = c_T(x, y) - \max(\mathrm{GCD}, \mathrm{MUD}) ,$$

where MUD is the Maximum Unnoticeable Difference. We assumed MUD = 1% for all the color components.

2.3 Minor additions

2.3.1 WS-BSNR

In order to provide better quality assessment for omnidirectional video, we applied WS-PSNR technique. According to [6], errors for pixels located at different distances from the equator are differently weighted:

$$\text{WS-BMSE} = \frac{\sum_{y=0}^{H-1}\sum_{x=0}^{W-1} \min_{\substack{x_R \in \left[x-\frac{B}{2}, x+\frac{B}{2}\right] \\ y_R \in \left[y-\frac{B}{2}, y+\frac{B}{2}\right]}} \left(c_T(x,y) - c_R(x_R, y_R)\right)^2 \cdot w_{x,y}}{\sum_{y=0}^{H-1}\sum_{x=0}^{W-1} w_{x,y}},$$

where weight $w_{x,y}$ is calculated as:

$$w_{x,y} = \cos\frac{\left(y + 0.5 - \frac{H}{2}\right)\cdot \pi}{H},$$

where x, y is a position of the pixel in ERP image and H is height of this image.

2.3.2 Chroma component analysis

In order to better simulate human perception system, the quality of both chroma components should be assessed, as well as luma. The BSNR value is calculated independently for each color component and finally BSNR for all components are averaged:

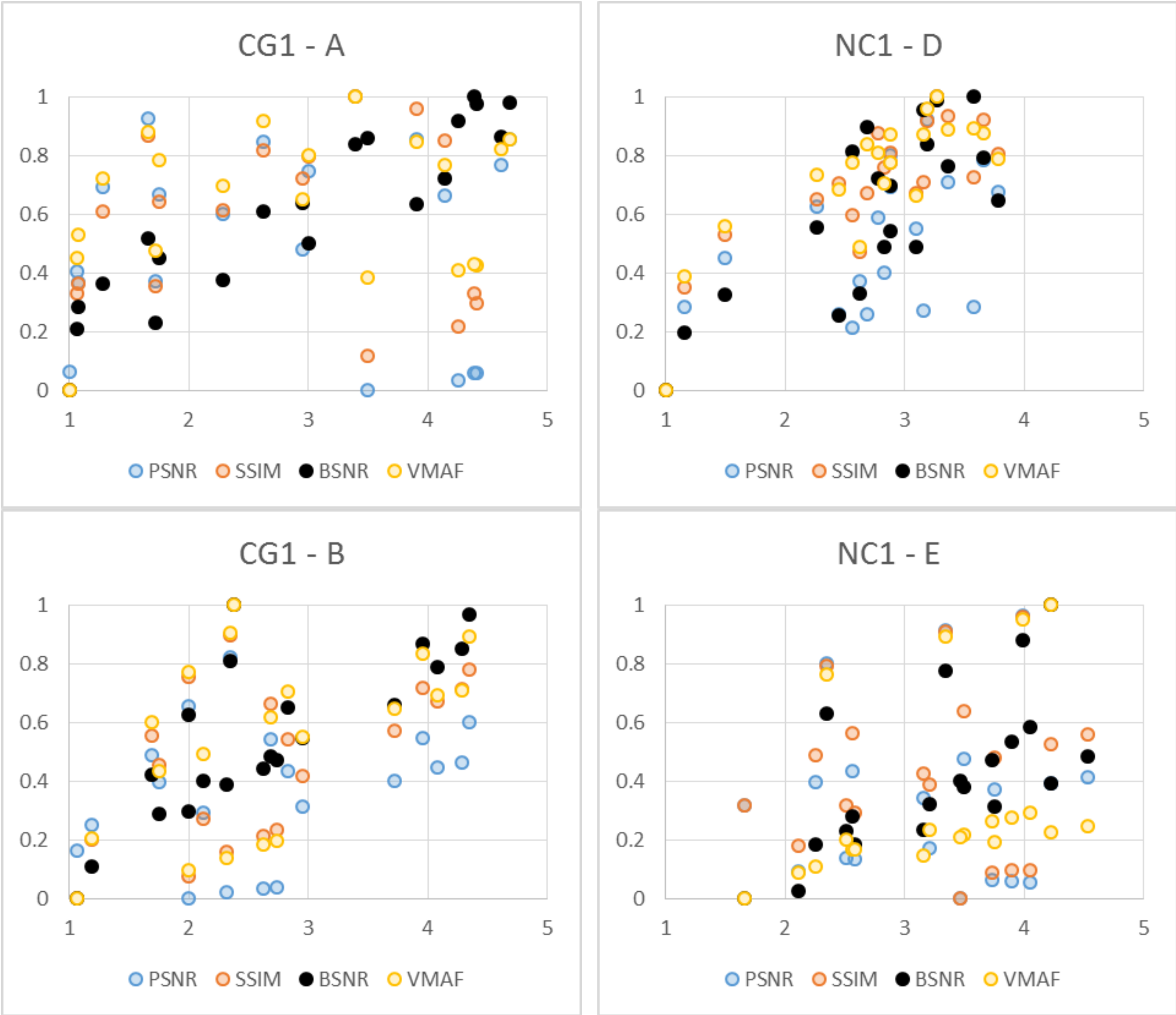
$$\text{BSNR}_{YUV} = \frac{\text{BSNR}_Y + \text{BSNR}_U \cdot w_U + \text{BSNR}_V \cdot w_V}{1 + w_U + w_V}.$$

Weights w_U and w_V are set to 0.25 (in 4:2:0 chrominance subsampling format the number of luma samples is 4 times higher).

3 Experimental results

The tests require good subjective quality results. Such tests are very time-consuming and laborious work, therefore we decided to use existing MOS scores. However, in order to study influence of different types of artifacts in synthesized views, we also needed results obtained for different methods/algorithms.

Therefore, proposed method was tested on the subjective quality results for CfP responses [1]. The test set contained results for 2 anchors and 3 CfP participants [3], [4], [5].



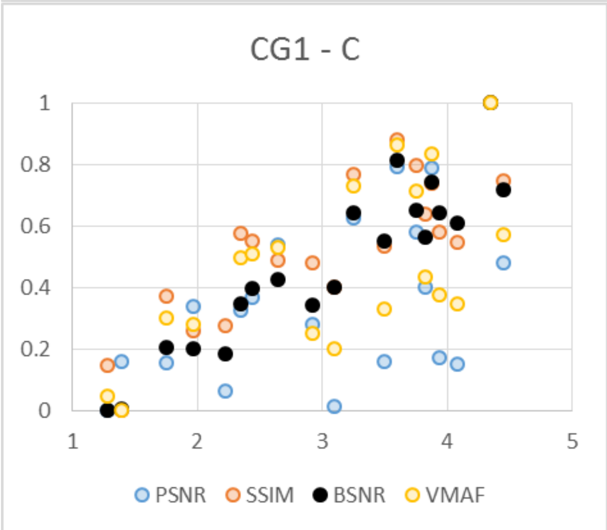


Fig. 1. MOS vs. objective quality metrics for computer-generated sequences (left column) and natural content (right column);
horizontal axis: MOS,
vertical axis: normalized objective quality.

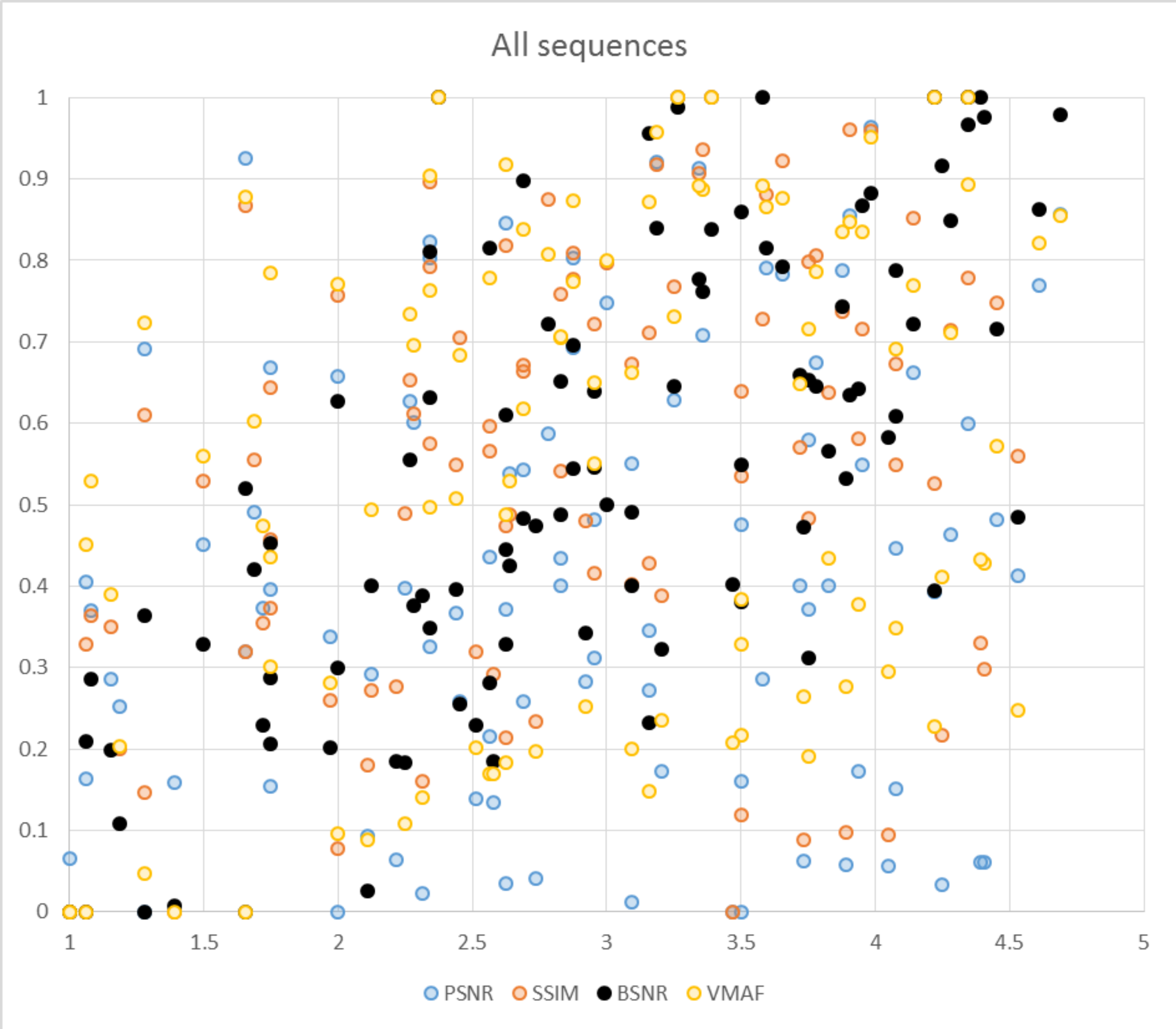


Fig. 2. MOS vs. objective quality metrics for all sequences used in CfP [2], horizontal axis: MOS, vertical axis: normalized objective quality.

Correlation coefficient between MOS and all 4 quality metrics are presented in Table 1.

Table 1. Correlation coefficient between MOS and objective quality metrics.

Quality metric	Correlation with MOS
PSNR	0.2109
SSIM	0.4225
VMAF	0.3383
BSNR (proposed)	0.7177

Presented results were obtained for one chosen view; the quality for 10 first frames for each sequence was assessed.

4 Acknowledgement

This work was supported by the Ministry of Science and Higher Education.

We would like to thank Bart Kroon from Philips and Julien Fleureau, Gérard Briand, Renaud Doré and Franck Thudor from Technicolor, who provided their CfP [4], [5] data, executables and support.

5 Recommendations

Performed tests indicate, that proposed simple quality metric allows to simulate subjective quality assessment results for Immersive Video applications.

We encourage the group members to test proposed solution on their own results and, when they will find it a good solution, to include it into CTC.

6 References

[1] V. Baroncini, G. Baroncini, “Report of the evaluation of the responses to the CfP on 3DoF+ Visual”, ISO/IEC JTC1/SC29/WG11 MPEG/N18353, Mar. 2019, Geneva, Switzerland.
[2] "Call for Proposals on 3DoF+ Visual" ISO/IEC JTC1/SC29/WG11 MPEG/N18145, January 2019, Marrakesh, MA.
[3] M. Domański, A. Dziembowski, D. Mieloch, O. Stankiewicz, J. Stankowski, A. Grzelka, G. Lee, J. Seo, “Technical description of proposal for Call for Proposals on 3DoF+ Visual prepared by Poznan University of Technology (PUT) and Electronics and Telecommunications Research Institute (ETRI)”, ISO/IEC JTC1/SC29/WG11 MPEG/M47407, Mar. 2019, Geneva, Switzerland.
[4] B. Kroon, B. Sonneveldt, “Philips response to 3DoF+ Visual CfP”, ISO/IEC JTC1/SC29/WG11 MPEG/M47179, Mar. 2019, Geneva, Switzerland.
[5] J. Fleureau, F. Thudor, R. Doré, B. Salahieh, M. Dmytrychenko, J. Boyce, “Technicolor-Intel Response to 3DoF+ CfP”, ISO/IEC JTC1/SC29/WG11 MPEG/M47445, Mar. 2019, Geneva, Switzerland.
[6] Y. Sun, A. Lu, and L. Yu. "Weighted-to-Spherically-Uniform Quality Evaluation for Omnidirectional Video." *IEEE Signal Processing Letters* 24.9(2017):1408-1412.