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

Title: [CVQM] Corrigendum for XPSNR Metric Statistics Provided in m74357
Status: Input document
Source: Fraunhofer HHI, Zhejiang University
Author: Christian Lehmann, Xiaoqi Zhuang, Christian Helmrich, Jiaqi Zhang, Adam Wieckowski, Chaojun Hu, Benjamin Bross, Lu Yu

Abstract

MPEG document m74357 provides very useful correlation-with-MOS results on five external, non-CVQM datasets for the PSNR, XPSNR, SSIM, MS-SSIM, VMAF, and FDIM visual quality assessment (VQA) metrics. However, it was recently observed that, due to a misalignment of the FFmpeg command line for XPSNR and the other metrics like VMAF and PSNR, the XPSNR results are partially incorrect. This errata document provides the corrected correlation-with-MOS data for the XPSNR metric, which are mostly higher than previous results.

1 Introduction

MPEG document m74357 [1] thankfully provides results regarding the correlation between objective visual quality assessment (VQA) metrics and mean opinion score (MOS) data, measured on external, non-CVQM datasets. The objective of this contribution is to facilitate subsequent discussions within the MPEG AG5 CVQM AhG regarding the potential role of external datasets in evaluating the generalization and robustness of metrics. Aside from the traditional VQA metrics (PSNR, SSIM, MS-SSIM, and VMAF), results for the extended perceptually weighted peak signal-to-noise ratio (XPSNR) [2] and the hybrid framework metric FDIM [3] are tabulated for 5 VQA datasets, 3 at UHD and 2 at full-HD resolution. For XPSNR, in particular, the weighted mean of the per-sequence averaged Y,U,V component outputs was computed using the formula

$$YUV = \frac{6 * Y + U + V}{8}$$

to determine an overall luma-chroma VQA score, as done for the PSNR and (MS-)SSIM; VMAF is Y-only.

Table 6 in [1] summarizes the across-sequence and across-codec (i.e., *all-sequence*) performances of above VQA metrics in terms of Pearson linear correlation (PLCC) and Spearman rank-order correlation (SROCC) with the ground-truth MOS data for each of the 5 datasets. Arguably, this table may be the most interesting to MPEG AG5, so its correctness appears crucial to the authors. However, unfortunately, the XPSNR results were recently identified to be partially incorrect, namely, the values are lower than expected. Sec. 2 provides further explanations regarding the cause of the issue, and Sec. 3 presents a corrected version of Table 6.

2 Reason for Error in XPSNR Assessment

Many full-reference VQA methods, including PSNR and VMAF, accept as command line arguments **first** the (usually distorted) signal under test, called somewhat vaguely *main* in FFmpeg's *psnr* filter, and **second** the (usually undistorted, i.e., original) reference signal. The *xpsnr* filter in FFmpeg [4], however, assumes the reversed order of the signal related command line arguments. Although this has been documented in [4] ("The first input is considered a ... reference source ..., whereas the second input is a (distorted) test signal"), we conclude that this difference in order of video under test and reference video may be confusing to users of XPSNR. Hence, the behavior of the xpsnr filter in FFmpeg may be corrected in a future FFmpeg version.

Currently, the correct command line for XPSNR measurements on two video files, each previously decoded to (preferably 10-bit) YUV format, is as follows in FFmpeg, where *dec.yuv* and *ref.yuv* are the distorted (or compressed here) and reference video sequences, respectively, of same length, resolution, and color format:

ffmpeg -s ... -i ref.yuv   -s ... -i dec.yuv   -lavfi xpsnr="stats_file=xpsnr.log" -f null -

3 Corrected Correlation-with-MOS Table

A corrected version of Table 6 is provided below. As noted in [1], for each of the metrics, performance was evaluated by correlating the objective VQA scores with the corresponding subjective MOS data on all five datasets. Both a Pearson Linear Correlation Coefficient (PLCC) and a Spearman Rank-Order Correlation Coefficient (SROCC) were calculated for each metric and dataset using the raw metric outputs, without applying any post non-linear mapping, under the *all-sequence* evaluation setting – see [1] for further details.

The best and second-best performing metrics are highlighted in red and blue, respectively.

Table 6. The correlation of metrics and MOS across datasets (**All-sequence** evaluation).

	MCML 4K		AVT-VQDB-UHD-1		WaterlooIVC 4K		CVQAC		BVI-HD	
	PLCC	SROCC	PLCC	SROCC	PLCC	SROCC	PLCC	SROCC	PLCC	SROCC
VMAF	0.9520	0.9161	0.7527	0.7979	0.7030	0.7113	0.5937	0.5552	0.7723	0.7724
XPSNR	0.9277	0.9324	0.7004	0.7377	0.7464	0.7807	0.5414	0.5135	0.7189	0.7574
PSNR	0.8181	0.8495	0.5942	0.6656	0.4816	0.4706	0.2749	0.3375	0.5539	0.5725
SSIM	0.6869	0.8872	0.3906	0.6236	0.4409	0.4978	0.3697	0.4307	0.5643	0.6711
MS-SSIM	0.6724	0.8696	0.5684	0.6426	0.4351	0.5198	0.3511	0.4268	0.5458	0.6623
FDIM	0.9715	0.9612	0.8561	0.8794	0.8048	0.8398	0.5585	0.5437	0.8141	0.7990

Corrected versions of Tab. 7–8 in [1] could not be created in time for the meeting and will be provided later.

4 Conclusion

Errors in the XPSNR measurements in MPEG input document m74357 were identified and, since the errors are significant, tables with corrected PLCC and SROCC correlation-with-MOS values are provided herein.

5 References

[1]   X. Zhuang, J. Zhang, C. Hu, and L. Yu, "[AHG CVQM] Public Datasets and Metrics Performance," *MPEG input doc. m74357*, Oct. 2025.

[2]   C. Helmrich, S. Bosse, H. Schwarz, D. Marpe, and T. Wiegand, "A Study of the Extended Perceptually Weighted Peak Signal-to-Noise Ratio (XPSNR) for Video Compression with Different Resolutions and Bit Depths," *ITU Journal: ICT Discoveries – Special Issue: The Future of Video and Immersive Media*, vol. 3, no. 1, online, May 2020. Online: http://handle.itu.int/11.1002/pub/8153d78b-en.

[3]   X. Zhuang, J. Zhang, and L. Yu, "A Full-Reference Objective Quality Assessment Method for Compressed Video," *MPEG input doc. m73451*, June 2025.

[4]   FFmpeg, "FFmpeg Filters Documentation." Online: https://ffmpeg.org/ffmpeg-filters.html#xpsnr-1.