|  |  |
| --- | --- |
| **Joint Collaborative Team on Video Coding (JCT-VC)**  **of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11**  25th Meeting: Chengdu, CN, 14–21 October 2016 | Document: JCTVC-Y0037 |

|  |  |  |  |
| --- | --- | --- | --- |
| *Title:* | **Cross-Component Peak Signal to Noise Ratio (xPSNR) support  in the HM and JEM software** | | |
| *Status:* | Input Document to JCT-VC | | |
| *Purpose:* | Proposal | | |
| *Author(s) or Contact(s):* | Y.T. Peng, A.M. Tourapis,  Y. Su, D. Singer  1 Infinite Loop Cupertino, CA 95014 USA | Tel: Email: | +1-408-228-7983 atourapis@apple.com |
| *Source:* | Apple Inc. | | |

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

# Abstract

This contribution presents an implementation of the Cross-Component Peak Signal to Noise Ratio (xPSNR), which is essentially an extension of PSNR that accounts for distortion in all colour components simultaneously, in both the HEVC HM and JVET JEM reference software.

# Introduction

Peak Signal to Noise Ratio (PSNR) has been traditionally the goto distortion metric for video standardization activities. It is relatively simple to compute, provides a somewhat meaningful evaluation of the distortion[[1]](#footnote-1) introduced on an image, and *under certain conditions* can be used to evaluate and compare the performance of two or more different coding schemes. To compute the PSNR value for a test image K of size m x n vs a reference image I of equal size, we can first compute the Mean Squared Error (MSE) as:

Then PSNR can be computed as:

where MAXi corresponds to the maximum possible pixel value for that image (e.g. commonly for 8 bit data that value is equal to 255, while for 10 bit data that value is commonly equal to 1023 or 1020). Commonly for video coding standardization purposes, this quantity is computed independently for color component, i.e. Y’, Cb, and Cr, and for every picture *i*. An overall sequence score per color component can then be computed by computing a combined measurement either using the overall average of PSNRs as:

or using the overall sequence MSE average, computed as:

Using these computations, basically a PSNRY, PSNRU, and PSNRV metric were provided per sequence. There has been considerable debate on how to combine these metrics and provide a single distortion measurement for the entire sequence that would account for the distortion in all colour components. Commonly a weighted average was suggested, with a higher weight given to the Y component than the two color components, i.e.

There is, unfortunately, one major drawback with this kind of computation. Distortion is accumulated at the colour plane level and ignores the influence of the distortion of each individual colour component to each pixel. Furthermore, the Y, Cb, and Cr components do not in reality hold any physical meaning. They are quantities that were defined primarily for decorrelating the actual signal of interest (commonly in an RGB representation or a CIE 1936 XYZ). Ideally, a metric that would account for distortion for each individual pixel, not colour component, would be highly desirable.

# Cross-Component PSNR (xPSNR)

Instead of computing distortion individually per colour component and averaging the overall per picture colour distortion, it is suggested that a better computation would be to consider the distortion of all colour components to a pixel simultaneously. This is a practice that is considered in several other metrics, such as DE2000 [1][2] and was recommended for use for HDR content evaluation in the CIE 1936 XYZ color space in [3]. In the case of YCbCr, and assuming that we are dealing with 4:4:4 content, this can be done by computing the value of *d* as follows:

where m,n are the width and height of the image, wY, wU, wV are weighting control parameters, and Y, U, and V are the three color components for the reference (index r) and test (index t) images. One can essentially argue that such a computation computes the magnitude of error for the three dimensional vector that corresponds to the pixel that we are accounting for.

The above computation can be extended also to 4:2:2 and 4:2:0. For 4:2:0 in particular, we can use sample replication to compute the value of d. In this way a distortion for all pixels can be properly computed.

The final xPSNR metric, given *d* can then be computed as:

# Simulation results

The xPSNR metric was implemented in both the HM and JEM software and tested using class C, D, and F from the common test conditions. We have evaluated performance using both the Open (Table 3‑1)and Closed GOP (Table 3‑2) configurations. For all of these configurations, weights of 1, 4, and 4, for the per component weighting parameters for Y, U/Cb, and V/Cr respectively were used.

Table 3‑1. Performance of the 1s Open GOP HM vs 1s Open GOP JEM reference software

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Piecewise Cubic | | | | | Cubic | | | | |
| C | Y | U | V | YUV | xPSNR | Y | U | V | YUV | xPSNR |
| C | -36.12% | -26.80% | -33.22% | -26.60% | -33.55% | -35.83% | -26.80% | -33.18% | -36.12% | -26.80% |
| D | -34.70% | -26.70% | -32.23% | -26.56% | -32.42% | -34.42% | -26.72% | -32.17% | -34.70% | -26.70% |
| F | -30.11% | -21.40% | -27.81% | -21.17% | -29.73% | -29.75% | -21.43% | -27.73% | -30.11% | -21.40% |

Table 3‑2. Performance of the 2s Closed GOP HM vs 2s Closed GOP JEM reference software

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Piecewise Cubic | | | | | Cubic | | | | |
| C | Y | U | V | YUV | xPSNR | Y | U | V | YUV | xPSNR |
| C | -32.52% | -27.87% | -30.55% | -26.27% | -29.72% | -32.26% | -27.87% | -30.51% | -32.52% | -27.87% |
| D | -30.20% | -26.32% | -29.48% | -26.52% | -27.86% | -29.86% | -26.34% | -29.42% | -30.20% | -26.32% |
| F | -28.31% | -22.79% | -26.39% | -21.25% | -27.81% | -28.03% | -22.82% | -26.37% | -28.31% | -22.79% |

It should be noted that we do not claim that xPSNR is a reliable metric or that it can match subjective quality. It does not still capture temporal or regional quality aspects but it is still a PSNR like metric that now accounts, at the pixel level, for the quality of all colour components simultaneously. It is very simple to compute and we would suggest its use for any future experiments as a better indicator of overall quality than per component PSNR.

# Conclusion/Recommendations

The cross-component PSNR (xPSNR) metric was presented in this contribution. This metric is essentially a simple extension of PSNR that accounts for the distortion of all colour components at the pixel, instead of the colour plane level. Implementations in both the HM and JEM, as well as in HDRTools are available.

# References

|  |
| --- |
| 1. Wikipedia article: Color Difference, <https://en.wikipedia.org/wiki/Color_difference> 2. G. Sharma, W. Wu, and E.N. Dalal, "The CIEDE2000 color-difference formula: Implementation notes, supplementary test data, and mathematical observations", Color Research & Applications (Wiley Interscience) 30 (1): 21–30, doi:10.1002/col.20070, 2005 3. A.M.Tourapis, Y. Su, D. Singer, C. Fogg, “Distortion Metrics for HDR/WCG: Peak Overall Signal to Noise Ratio”, m36769, Geneva, CH, June, 2015 |

# Patent rights declaration(s)

**Apple Inc. may have current or pending patent rights relating to the technology described in this contribution and, conditioned on reciprocity, is prepared to grant licenses under reasonable and non-discriminatory terms as necessary for implementation of the resulting ITU-T Recommendation | ISO/IEC International Standard (per box 2 of the ITU-T/ITU-R/ISO/IEC patent statement and licensing declaration form).**

1. Even though such distortion does not always correlate to [↑](#footnote-ref-1)