A STUDY OF THE PERCEPTUALLY WEIGHTED PEAK SIGNAL-TO-NOISE RATIO (WPSNR) FOR IMAGE COMPRESSION

Johannes Erfurt[1], *Christian R. Helmrich*[1], *Sebastian Bosse*[1], *Heiko Schwarz*[1,2], *Detlev Marpe*[1], *and Thomas Wiegand*[1,3]

[1]Video Coding and Analytics Department, Fraunhofer Heinrich Hertz Institute (HHI), Berlin, Germany
[2]Institute for Computer Science, Free University of Berlin, Germany
[3]Image Communication Chair, Technical University of Berlin, Germany
[johannes.erfurt, christian.helmrich, heiko.schwarz, detlev.marpe, thomas.wiegand]@hhi.fraunhofer.de

ABSTRACT

The peak signal-to-noise ratio (PSNR) is the most used objective measure for assessing perceptual image quality when it comes to image and video compression tasks, despite the fact that it exhibits weak performance in reflecting human perception. To address this problem, many image quality assessment (IQA) methods were proposed, e. g. the structural similarity quality measure (SSIM) and its extension, the multiscale SSIM (MS-SSIM). In this paper we revisit and evaluate a block-based perceptually weighted PSNR (WPSNR) which calculates weighting factors to capture visual sensitivity of local image regions. We further introduce a sample-based version of WPSNR which determines those sensitivity weights with higher spatial accuracy. These methods are computationally inexpensive compared to other similarity measures and are shown to outperform PSNR, SSIM and similar perceptual quality measures when it comes to approximate subjective ratings of JPEG or JPEG2000 compressed images.

Index Terms— IQA, PSNR, SSIM, image compression

1. INTRODUCTION

The network traffic due to image and video content is constantly increasing [1]. In order to meet today's transmission or storage constraints, most signals are highly compressed and therefore suffer from coding artifacts such as blocking, ringing or other distortions that are visible to human observers. To evaluate or optimize coding engines, e. g. a video encoder, these distortions have to be measured. The most reliable way to measure the quality of such data is to perform subjective evaluation. In practice this is usually not feasible as such tests are very time-consuming, expensive and typically not practical for quality assurance. Therefore, quantitative measures need to be developed to predict such perceived quality experiences by human viewers which are referred to as mean opinion scores (MOS). In the past two decades several models for image quality assessment (IQA) have been proposed [2]-[12]. These can be classified as full reference (FR), no-reference (NR) or reduced-reference (RR) methods, depending on the (full, non or partial) availability of an original image which acts as the "ground truth" and entity a distorted image is compared with. For image and video compression applications the reference image is usually available and considered as the optimum. The conventional model for such tasks is the peak signal-to-noise ratio (PSNR). As it possesses a very low complexity and high simplicity it is widely used and always compared with. However it only poorly correlates with perceived quality by humans [13]-[15]. Two IQAs which try to overcome this problem and influenced many researchers in image processing are the structural similarity measure (SSIM) and the multiscale SSIM (MS-SSIM) [2],[3], which assume that the human visual system is highly adapted to extract structural information and compares luminance, contrast and structural similarity of the distorted with the reference image. Even though SSIM is a good assessor for subjective quality, its acquisition is still time consuming compared to PSNR, especially if it has to be calculated several times in optimization processes of encoding engines for compression tasks.

In [10] a weighted PSNR (WPSNR) was introduced which was motivated by the fact that in a compressed image or video occurring coding artifacts are often only perceivable in some specific regions. For example, it can often be observed that the subjective quality of low-contrast image regions (i. e. regions with low visual activity) is clearly reduced through better visibility of compression artifacts [16],[17] while high-contrast image regions often effectively obscure these artifacts such that a degradation is hardly visible. This effect motivates to evaluate local regions differently in terms of their contribution to subjective quality.

In this paper the WPSNR metric is further investigated. In its original configuration the image is divided into equal-sized blocks. Based on the properties of the original image a weighting factor is calculated for each block, which can be interpreted as a visual sensitivity weight [9]-[12]. For each block the sum of squared errors (SSE) is determined, but in-

stead of directly using the SSE for establishing a distortion measure for the whole image, a weighted SSE is calculated and each SSE of the local block scaled with a corresponding weighting factor. The WPSNR is then computed by using the local weighted SSE instead of the conventional SSE. Furthermore we introduce a more localized sample-based version of WPSNR where each sample is assigned its own sensitivity weight. These two IQA methods satisfy both requirements of relatively low complexity and good approximation of subjective scores. For verification the two WPSNR methods, which we call bWPSNR for the block-based version and sWPSNR for the sample-based version, are evaluated on four different image databases, i. e. LIVE, TID2008, TID2013 and CSIQ [18]-[21] specifically for JPEG and JPEG2000 compressed images and compared with different quality assessment methods. Here the scores of the WPSNR metric serve as a predictor for the MOS values set by human test subjects and their correlation is measured by Pearson linear correlation coefficient (PLCC) and by Spearman rank-order correlation coefficient (SROCC) [22].

The rest of the paper is structured as follows. In Sections 2 and 3 the concept of bWPSNR and sWPSNR is explained in detail. Section 4 evaluates the performance of this method on the given image databases and Section 5 concludes the paper. Visual demonstrations are provided at [23]. Unless noted otherwise, the WPSNR parameters were chosen using other test images and videos than those employed in this study.

2. BLOCK-BASED PERCEPTUALLY WEIGHTED PSNR (BWPSNR)

This section is based on the description in [10] and [26].

2.1. Block-based Weighted MSE Distortion Measure

Let x and y be the luminance values for the reference and the distorted image. The image is divided into equal-sized blocks B_k of size N and MSE_k denotes the mean square error for the block k given by

$$MSE_k = \frac{1}{N} \sum_{(i,j)\in B_k} \big(x(i,j) - y(i,j)\big)^2. \tag{1}$$

Next a generalization through a weighting factor w_k is given which better reflects the subjective relevance of the block B_k:

$$MSE_k^w = w_k \cdot MSE_k. \tag{2}$$

The weighting factor w_k can be calculated in an arbitrary way but only by incorporating features of the reference image x. It can be interpreted as a visual subjective sensitivity measure for the corresponding local block. If B_k possesses a large sensitive measure w_k the local MSE distortion has a higher impact on perceptual quality than a block with a small w_k.

The overall MSE distortion for an image is composed of the sum of all block distortions, i. e.

$$MSE = \sum_k MSE_k = \frac{1}{W \cdot H} \sum_{i,j} \big(x(i,j) - y(i,j)\big)^2 \tag{3}$$

with W and H being the width and height, respectively of the input images x and y. Hence, we can define the overall weighted MSE distortion for an image as

$$\begin{aligned} MSE^w &= \sum_k MSE_k^w = \sum_k w_k \cdot MSE_k \\ &= \frac{1}{W \cdot H} \sum_k w_k \sum_{(i,j)\in B_k} \big(x(i,j) - y(i,j)\big)^2. \end{aligned} \tag{4}$$

Note if $w_k = 1$ for all k the weighted MSE becomes identical to the ordinary MSE.

In case x and y are color images we calculate the weighted MSE only for the luminance part of x and y for simplicity. An extension to chromatic image channels is described in [10].

2.2. WPSNR Calculation

For given MSE of an image, bit depth BD, width W and height H the corresponding PSNR is defined by

$$PSNR = 10 \cdot \log_{10} \left(\frac{(2^{BD} - 1)^2}{MSE} \right). \tag{5}$$

Intuitively the weighted PSNR is defined through the weighted MSE, therefore

$$PSNR^w = 10 \cdot \log_{10} \left(\frac{(2^{BD} - 1)^2}{MSE^w} \right). \tag{6}$$

2.3. Perceptually Relevant Block Weighting Factors

The remaining and most important question is how to calculate the sensitivity measures w_k for the local blocks B_k. It is often observed that in regions governed by low frequencies, i. e. smooth regions with low visual activity, coding artifacts are becoming more visible than in regions with dominant high frequencies, i. e. regions with high activity. Therefore, we derive the local sensitivity weights w_k based on a local energy measure of the high-pass filtered reference image x. A 9-tap high-pass filter is applied to x with the filter kernel

$$F = \frac{1}{4} \cdot \begin{bmatrix} -1 & -2 & -1 \\ -2 & 12 & -2 \\ -1 & -2 & -1 \end{bmatrix}, \tag{7}$$

and the filtered samples of x are obtained by

$$h_x = x * F. \tag{8}$$

Then for each block B_k the local activity a_k is obtained by

$$a_k = \max\left(a_{min}^2, \left(\frac{1}{N^2} \sum_{(i,j)\in B_k} |h_x(i,j)| \right)^2 \right), \tag{9}$$

where [10] defines $a_{min} = 2^{BD-6}$ empirically in the context of perceptual bit-allocation and N^2 is the number of samples in B_k. The clipping via the max-function is required to avoid very small a_k and, thus, divisions by very small values.

In order to obtain sensitivity weights w_k with mean value close to 1, we normalize a_k by a factor a_{pic}. Now instead of trying to average over all local activities in the present image we average over a whole set of images. Using common high-resolution (HD, UHD) images, a_{pic} can be approximated by

$$a_{pic} = \frac{1}{K_{set}} \sum_{k \in K_{set}} a_k \approx 2^{BD} \cdot \sqrt{\frac{3840 \cdot 2160}{W \cdot H}}, \qquad (10)$$

where K_{set} specifies the total number of considered blocks. Note that it is not desired to obtain sensitivity weights which are on a picture level on average close to 1, because it could occur that most parts of the image are subjectively amplified due to the existing smooth structure of the image or on the opposite if most parts of the image are rich in contrast. Then in the former case most of the sensitivity weights and the average should be larger and in the latter case smaller than 1.

The final visual sensitivity weight for block B_k can now be calculated using the reciprocal of the normalized a_k:

$$w_k = \left(\frac{a_{pic}}{a_k} \right)^{\beta}, \qquad (11)$$

where β controls the impact of the deviating block sensitivity values and should be chosen between 0 and 1. Note that if $\beta = 0$ all scaling factors are equal 1 and WPSNR reduces to the traditional PSNR measure. We found based on our visual experiments [10] $\beta = \frac{1}{2}$ to be a good fit. In this case and by shifting N^2, the sensitivity weight for block B_k simplifies to

$$w_k = N^2 \cdot \frac{\sqrt{a_{pic}}}{a'_k} \qquad \text{and} \qquad (12)$$

$$a'_k = \max\left(N^2 \cdot a_{min}, \sum_{(i,j) \in B_k} |h_x(i,j)| \right), \qquad (13)$$

which avoids the division by N^2 and the exponentials in (9) and (11) and, thus, reduces the complexity. It is worth noting that $N^2 \cdot \sqrt{a_{pic}}$ and $N^2 \cdot a_{min}$ are constants and calculated just once for the whole image or video.

At last we chose the size $N \times N$ of the local square-sized blocks B_k to be dependent on the resolution of the input image, i. e. width W and height H, and define

$$N = \text{round}\left(128 \cdot \sqrt{\frac{W \cdot H}{3840 \cdot 2160}} \right). \qquad (14)$$

3. SAMPLE-BASED PERCEPTUALLY WEIGHTED PSNR (SWPSNR)

The block-based approach to calculate the visual sensitivity weights is very efficient as only one weight is determined for a whole block. Depending on its application this approach might be desirable since it requires little algorithmic complexity. On the other hand it might be useful to avoid unnecessary blocking artifacts in the resulting "weighting map". Moreover, a more accurate way of representing subjective perception of varying regions (by more closely modelling the "continuous" cell-wise operation of the human retina [24]) might be beneficial. This can be achieved if the sensitivity weights are calculated on a sample-level, i. e. each individual sample is assigned its own weight. For that purpose, some adaptations have to be made. Instead of determining the local activities a_k and subsequently the sensitivity weight w_k for each block B_k, they are calculated for each sample location (i, j), hence

$$w_{i,j} = \left(\frac{a_{pic}}{a_{i,j}} \right)^{\beta} \qquad \text{and} \qquad (15)$$

$$a_{i,j} = \max\left(a_{min}^2, \left(\frac{1}{M^2} \sum_{(k,l) \in N_{i,j}} |h_x(k,l)| \right)^2 \right). \qquad (16)$$

Here $N_{i,j}$ is the local neighborhood of (i, j), a local window of size $M \times M$ with center pixel location (i, j) with replicated boundary samples at the image borders. The window size defined through M, like the definition of N in (14), is related to the resolution of the input image and empirically defined as

$$M = 2 \cdot \text{round}\left(14 \cdot \sqrt{\frac{W \cdot H}{3840 \cdot 2160}} \right) + 1. \qquad (17)$$

Note that, given the sample-wise definition of $w_{i,j}$, it is reasonable to choose M smaller than N.

For $\beta = \frac{1}{2}$ equations (15) and (16) simplify similarly to (12) and (13). The weighted MSE and WPSNR is similarly obtained by replacing w_k for each block with the scaling factor $w_{i,j}$ for each sample location, hence

$$MSE^w = \frac{1}{W \cdot H} \sum_{i,j} w_{i,j} \big(x(i,j) - y(i,j) \big)^2. \qquad (18)$$

4. EXPERIMENTS

We evaluated the proposed block-based and sample-based WPSNR metrics on the LIVE, TID2008, TID2013 and CSIQ image databases [18]-[21] for the contained JPEG2000 and JPEG compressed images. The methods serve as a predictor for the MOS values collected in subjective tests. The prediction accuracy is measured, in terms of linear-model fit, using Pearson linear correlation coefficient (PLCC) values and, in terms of prediction monotonicity, using Spearman rank-order correlation coefficient (SROCC) values. The correlation results of PSNR, SSIM and MS-SSIM serve as comparative values. For this study we set $a_{min} = 2^{BD-8}$ in (9) and (16), which shows a slightly better fit on the tested databases than the value chosen in [10] (which, as noted, was defined specifically for bit-allocation purposes in a HEVC-based codec).

Tables 1 and 2 show the performance of WPSNR in comparison to PSNR and the SSIM approaches. The bold printed numbers represent the highest correlation for the corresponding database and compression type. Figure 1 illustrates the IQA model performances on the JPEG and JPEG2000 distortion types of the TID2013 database, including second-order polynomial curve fitting to visualize the level of correlation.

It can be concluded that the WPSNR measures outperform PSNR by far. Even in comparison with SSIM and MS-SSIM the WPSNR measures reach higher correlation values on average. Only in a few cases SSIM and MS-SSIM perform better. It is worthwhile to note that bWPSNR and sWPSNR perform very well on all databases (except for CSIQ for JPEG compression) if Pearson correlation is calculated. Here both methods are about 0.02 better than SSIM and 0.03 better than MS-SSIM on average. For Spearman's correlation, the advantage of the WPSNR methods is smaller. The block-based approach is slightly better than the sample-based approach for SROCC and vice versa for PLCC, although the difference is not statistically significant.

As can be seen from Figure 1 both WPSNR methods deal very well with compression artifacts. They correlate much better with the corresponding MOS values than PSNR and SSIM (MS-SSIM is omitted as the results look very similar to SSIM). As a welcome side-effect, both JPEG and JPEG2000 scores can be well fitted with only one curve (the red and blue curves are almost identical). This indicates that the WPSNR, either in its block-based or sample-based configuration, can serve as a codec-agnostic predictor of visual coding quality.

5. DISCUSSION AND CONCLUSION

We studied the perceptually weighted peak signal-to-noise ratio (WPSNR), a novel approach for image quality assessment by perceptually weighting the well-known PSNR measure. By calculating the mean activity for local regions and incorporating these into the weight calculation we showed that dealing with compression distortion the WPSNR is a good predictor for subjective perception of images. It outperforms in this regard the PSNR measure, which is still the most used objective metrics for optimizing tasks in image and video compression. Both presented methods achieve similar, and sometimes better, results compared to SSIM/MS-SSIM with bWPSNR having a better performance complexity trade-off.

There are clearly more IQA models to compare which achieve even higher correlation with subjective scores. These methods, including [4]-[8], usually require much higher computing power [6],[8] and are not suitable for optimization tasks, where the given quality measure has to be calculated repeatedly. Here WPSNR has a crucial advantage: the sensitivity weights need to be calculated only once per image, even if the corrupted image changes during an optimization process. In this case only the MSE has to be recalculated. The evaluation of the WPSNR metric on high-resolution image or video coding databases remains a subject for future research.

SROCC	PSNR	SSIM	MS-SSIM	bWPSNR	sWPSNR
LIVE	0.8809	0.9764	**0.9815**	0.9604	0.9598
	0.8954	0.9614	**0.9627**	0.9513	0.9472
TID2008	0.8717	0.9252	0.9322	**0.9473**	0.9460
	0.8132	0.9625	0.9700	**0.9751**	0.9734
TID2013	0.9189	0.9200	0.9265	0.9499	**0.9549**
	0.8840	0.9468	0.9504	**0.9701**	**0.9701**
CSIQ	0.8881	0.9546	**0.9634**	0.9583	0.9567
	0.9362	0.9606	0.9683	0.9706	**0.9711**
Overall	0.8861	0.9509	0.9569	**0.9604**	0.9599

Table 1: Spearman rank-order correlation coefficient on four different image databases specifically for JPEG (top row) and JPEG2000 (bottom row) compressed images.

PLCC	PSNR	SSIM	MS-SSIM	bWPSNR	sWPSNR
LIVE	0.8650	0.9279	0.9184	0.9502	**0.9535**
	0.8747	0.8925	0.8697	**0.9275**	0.9231
TID2008	0.8597	0.9319	0.9279	**0.9630**	0.9613
	0.8629	0.9492	0.9365	**0.9685**	0.9664
TID2013	0.8972	0.9278	0.9207	0.9585	**0.9633**
	0.9078	0.9424	0.9183	0.9642	**0.9665**
CSIQ	0.7898	**0.9165**	0.9064	0.8423	0.8460
	0.9270	0.8967	0.8843	0.9522	**0.9573**
Overall	0.8730	0.9231	0.9103	0.9408	**0.9422**

Table 2: Pearson linear correlation coefficient on four different image databases specifically for JPEG (top row) and JPEG2000 (bottom row) compressed images.

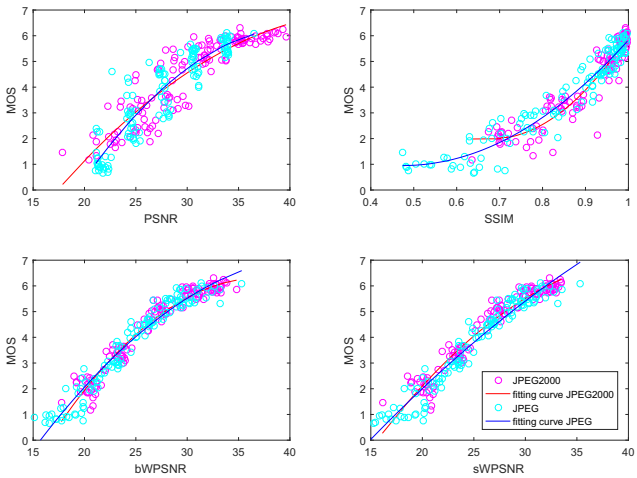

Fig. 1: PSNR (top left), SSIM (top right), bWPSNR (bottom left), sWPSNR (bottom right) correlation with MOS for JPEG and JPEG2000 distorted images of the TID2013 database.

6. REFERENCES

[1] Cisco, *Cisco visual networking index: Forecast and trends*, 2017-2022, White paper, 2018.

[2] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, *Image quality assessment: From error visibility to structural similarity*, IEEE Trans. Image Process. **13**, no. 4, pp. 600–612, 2004.

[3] Z. Wang, E. P. Simoncelli, and A. C. Bovik, *Multiscale structural similarity for image quality assessment*, in Proc. IEEE 37th Asilomar Conf. on Signals, Systems, and Computers, 2003.

[4] H. R. Sheikh and A. C. Bovik, *Image Information and Visual Quality*, IEEE Trans. Image Process. **15**, no. 2, pp. 430–444, 2006.

[5] L. Zhang, L. Zhang, X. Mou, and D. Zhang, *FSIM: A Feature Similarity Index for Image Quality Assessment*, IEEE Trans. Image Process. **20**, no. 8, pp. 2378–2386, 2011.

[6] L. Zhang and H. Li, *SR-SIM: A fast and high performance IQA index based on spectral residual*, in Proc. 19th IEEE Int. Conf. Image Process., pp. 1473–1476, 2012.

[7] A. Liu, W. Lin, and M. Narwaria, *Image quality assessment based on gradient similarity*, IEEE Trans. Image Process. 21 **4**, pp. 1500–1512, 2012.

[8] R. Reisenhofer, S. Bosse, G. Kutyniok, and T. Wiegand, *A Haar Wavelet-Based Perceptual Similarity Index for Image Quality Assessment*, Signal Processing: Image Communication, **61**, pp. 33–43, 2018.

[9] S. Bosse, M. Siekmann, W. Samek, and T. Wiegand, *A Perceptually Relevant Shearlet-Based Adaptation of the PSNR*, in Proc. IEEE Int. Conf. on Image Process., 2017.

[10] C. R. Helmrich, S. Bosse, M. Siekmann, H. Schwarz, D. Marpe, and T. Wiegand, *Perceptually Optimized Bit-Allocation and Associated Distortion Measure for Block-Based Image or Video Coding*, in Proc. IEEE Data Comp. Conf., 2019.

[11] S. Bosse, S. Becker, Z. V. Fisches, W. Samek, and T. Wiegand, *Neural Network-based Estimation of Distortion Sensitivity for Image Quality Prediction*, in Proc. 25th IEEE Int. Conf. Image Process., 2018.

[12] S. Bosse, S. Becker, K.-R. Müller, W. Samek, and T. Wiegand, *Estimation of distortion sensitivity for visual quality prediction using a convolutional neural network*, Digital Signal Process., In Press, 2018. https://doi.org/10.1016/j.dsp.2018.12.005

[13] Z. Wang and A. C. Bovik, *Mean squared error: Love it or leave it? A new look at Signal Fidelity Measures*, IEEE Signal Processing Magazine **26**, no. 4, 2009.

[14] B. Girod, *What's wrong with mean-squared error?*, Digital Images and Human Vision, A. B. Watson, Ed. Cambridge, MA: MIT Press, pp. 207–220, 1993.

[15] Z. Wang, A. C. Bovik, and L. Lu, *Why is image quality assessment so difficult*, in Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Process. **4**, Orlando, FL, pp. 3313–3316, 2002.

[16] S. J. Daly, *Application of a noise-adaptive contrast sensitivity function to image data compression*, Opt. Eng. 29 **8**, pp. 977–987, 1990.

[17] Y. Jia, W. Lin, and A. A. Kassim, *Estimating just-noticeable distortion for video*, IEEE Trans. Circuits Syst. Video Technol. 16 **7**, pp. 820–829, 2006.

[18] H. R. Sheikh, Z. Wang, L. Cormack, and A. C. Bovik. *LIVE Image Quality Assessment Database*, release 2, 2005. http://live.ece.utexas.edu/research/Quality/subjective.htm

[19] N. Ponomarenko, V. Lukin, A. Zelensky, K. Egiazarian, M. Carli, and F. Battisti, *TID2008 - A Database for Evaluation of Full-Reference Visual Quality Assessment Metrics*, Advances of Modern Radioelectronics, **10**, pp. 30–45, 2009.

[20] N. Ponomarenko, L. Jin, O. Ieremeiev, V. Lukin, K. Egiazarian, J. Astola, B. Vozel, K. Chehdi, M. Carli, F. Battisti, and C.-C. Jay Kuo, *Image database TID2013: Peculiarities, results and perspectives*, Signal Processing: Image Communication **30**, pp. 57–77, 2015.

[21] E. C. Larson and D. M. Chandler, *Most apparent distortion: a dual strategy for full-reference image quality assessment*, Proc. SPIE **7242**, 72420S, 2009.

[22] Z. Wang and Q. Li, *Information content weighting for perceptual image quality assessment*, IEEE Trans. Image Process. **20**, no. 5, pp. 1185–1198, 2011.

[23] C. Helmrich, J. Erfurt, "ecodis WPSNR Demonstration Page", 2019, http://www.ecodis.de/wpsnr.htm

[24] A. Valberg, *Light Vision Color*, Wiley, Mar. 2005.

[25] G. Bjøntegaard, *Calculation of average PSNR differences between RD-curves*, VCEG-M33, Austin, USA, Mar. 2001.

[26] C. Helmrich, H. Schwarz, D. Marpe, and T. Wiegand, *Improved Perceptually Optimized QP Adaptation and Associated Distortion Measure*, JVET-K0206, Ljubljana, Slovenia, July 2018.