A PERCEPTUALLY RELEVANT SHEARLET-BASED ADAPTATION OF THE PSNR

Sebastian Bosse[1], *Mischa Siekmann*[1],
Wojciech Samek[1], Member, IEEE, and *Thomas Wiegand*[1,2], Fellow, IEEE

[1] Fraunhofer Institute for Telecommunications, Heinrich Hertz Institute, Berlin, Germany
[2] Department of Electrical Engineering, Technical University of Berlin, Germany.

ABSTRACT

Although being one of the simplest and most widely used image quality metrics (IQMs) the peak signal-to-noise ratio (PSNR) correlates only poorly with visual quality as perceived by humans. Based on an analysis of the non-linear mapping from PSNR to mean opinion scores (MOS) we identify a functional mapping parameter to adapt the PSNR perceptually meaningful. Neurophysiologically motivated, a shearlet-based correction is proposed for controlling this perceptual PSNR adaption. The performance of the proposed perceptually adapted PSNR is evaluated on the LIVE and TID2013 databases and shows to be superior or comparable to benchmark IQMs.

Index Terms— Image quality assessment, shearlets, human visual system, quality perception, image coding

1. INTRODUCTION

Robust and reliable assessment of visual quality in a for humans perceptually meaningful way is crucial for every image and video communication system and enables the optimization, monitoring and evaluation of these systems.

A whole variety of models for image quality assessment (IQA) has been proposed and is commonly categorized as full reference (FR), reduced reference (RR) or no reference (NR) IQMs, depending on how much information from the reference image is available to the IQA algorithm. However, a precise model for the subjective perception of distortion is not available for any of these categories. Although NR IQA targets at a very general solution to the IQA problem, it may not be feasible for all applications. A popular example is video encoder optimization, where e.g. distortions, such as film grain, deliberately introduced by the director should not be 'corrected' by the encoding algorithm.

The simplest and most widely used FR IQMs are the mean square error (MSE), computed as the average energy of the samplewise error between the reference and the distorted image and the PSNR, a logarithmic approximation of the human distortion perception based on the application of the Weber-Fechner law [1] to the MSE. However, although being simple to compute and convenient to optimize on, MSE and PSNR do not correlate satisfyingly with human perception [2]. Thus, researchers and engineers have proposed a large number of computational FR IQMs. Typically, these are classified in *bottom-up* and *top-down* approaches, depending on if the human visual system (HVS) or its subsystem is modeled explicitly, or if hypothesized general functionalities of the HVS are mimicked in a rather abstract sense and from the perspective of signal processing, commonly based on statistical image features [3]. Popular examples of the first category incorporate the contrast sensitivity function of the HVS [2, 4] or noise detection thresholds [2, 5]. Most prominent approaches of the second category exploit changes in statistical image properties introduced by distortions, as the structural similarity index (SSIM) [6], based on the local mean and local (co-)variance of the reference and the distorted image, and its multiscale extension (MS-SSIM) [7].

Generally, FR IQMs can benefit from adaptations to the specific content of the images or videos to be tested [8]. This may be done for a whole video sequence, a single image or locally, e.g. by considering HVS models such as saliency [9] or scalewise divisive normalization [10], information content [11], conditional probability [12] or learned features [13, 14].

To predict the visual quality as perceived by humans the output of a IQA algorithm is typically mapped by a non-linear function [15]. In this paper, we show that this mapping provides insight into the shortcomings of the PSNR. We identify a parameter of the mapping function driving source image dependent distortion sensitivity. In order to correct for distortion sensitivity and in order to improve prediction performance we propose a perceptually motivated local weighting scheme for the PSNR based on the spatially local distribution of shearlet coefficients of the source reference image without any preprocessing of the image data. We evaluate the proposed approach on the LIVE [16] and the TID2013 [17] databases.

2. METHODS

2.1. Distortion Sensitivity

Due to saturation effects in the two extreme cases of imperceptible quality loss and very low quality, subjective image quality ratings do not relate linearly to computational qual-

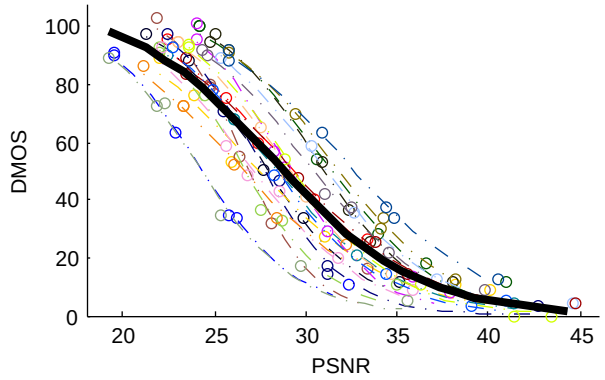


Fig. 1: PSNR vs. DMOS for the JPEG subset of the LIVE database [16]. Colored dashed curves and circles indicate regressed and measured DMOS values for individual reference images. The thick black curve shows regressed DMOS values for the whole ensemble.

ity models. When evaluating quality metrics, typically, this is accounted for by mapping a computational quality value Q_c to the prediction of a perceptual quality value Q_p (commonly MOS or DMOS) by a non-linear regression between Q_c and Q_p [15].

A commonly used regression function is the 4-parameter sigmoid function

$$\widehat{Q}_p = a + \frac{b-a}{1+e^{-c(Q_c-d)}}, \tag{1}$$

where $\widehat{Q}_p$ is the perceived quality estimated from the measurement of Q_c. In this function, a and b are the lower and upper bound of the range of perceived quality, c controls the slope of the regression curve and d shifts the regression with respect to Q_c. Parameters a and b are solely driven by the range of the chosen quality rating scale used in the according psychometric tests and thus in general known a-priori. Parameters c and d are not depending on the range of the quality scale but on the relation between the quality metric and the ground-truth quality scores for the ensemble of images under test. This is shown in Fig. 1 for the JPEG subset of the LIVE database [16] for example of Q_c being measured as PSNR and Q_p as DMOS. The colored circles show the PSNR vs. the reported DMOS, (Q_c vs. Q_p), the dashed colored curves show the regression estimated DMOS (Q_c vs. $\widehat{Q}_p$) for individual source images. The thick black curve shows the regression for the whole ensemble of source images. As Fig. 1 shows, the different source images exhibit distinct sensitivity of the MOS with regard to the PSNR, such that for some images a relatively high PSNR leads to relatively low perceived quality and vice versa. We refer to this behavior as *distortion sensitivity*, given as a property of a source reference image with regard to a specific quality metric.

Obviously, finding a suitable parameter c and d for each reference image would increase the performance of the image quality metric at hand. In this paper we focus on the influence of the shifting parameter d. However, distortion sensitivity might not only be diverging for different images, but also for different regions within one image.

2.2. Distortion Sensitivity and the PSNR

The PSNR is defined via the MSE by

$$\mathrm{PSNR} = 10 \cdot \log_{10} \frac{C^2}{\mathrm{MSE}} \tag{2}$$

with C being the maximum (peak) sample value of the given signal. Imagewise consideration of the shifting parameter d_I from Eq. 1 in an imagewise perceptually adapted PSNR (paPSNR_I) leads to

$$\begin{aligned} \mathrm{paPSNR}_I &= \mathrm{PSNR} - d_I & (3)\\ &= 10 \cdot \log_{10} \frac{C^2}{\mathrm{MSE}} - d_I \\ &= 10 \cdot \log_{10} \frac{C^2}{10^{\frac{d_I}{10}}\mathrm{MSE}}. & (4) \end{aligned}$$

From here we can integrate the offset parameter d_I to an imagewise perceptually adapted MSE (paMSE_I)

$$\mathrm{paMSE}_I = 10^{\frac{d_I}{10}}\mathrm{MSE} \tag{5}$$

As stated before, the concept of distortion sensitivity does not only apply to full images (or videos), but can be refined to regions within an image. By assigning distortion sensitivity $d(x,y)$ to every sample position (x,y) in an image with $s(x,y)$ being the reference and $\tilde{s}(x,y)$ being the distorted image we find a perceptually adapted MSE (paMSE) as

$$\mathrm{paMSE} = \frac{1}{M \cdot N} \sum_{x=0}^{M} \sum_{y=0}^{N} 10^{\frac{d(x,y)}{10}} (s(x,y) - \tilde{s}(x,y))^2 \tag{6}$$

leading directly to a perceptually adapted PSNR (paPSNR)

$$\mathrm{paPSNR} = 10 \cdot \log_{10} \frac{C^2}{\mathrm{paMSE}}. \tag{7}$$

2.3. Shearlet-Based Estimation of Distortion Sensitivity

Shearlets are a directional multiscale decomposition system providing an unified treatment of continuous as well as discrete models. For a certain class of natural images, so-called cartoon-like images (images consisting of smooth areas separated by piecewise smooth boundary curves), shearlet-based transforms were shown to provide optimally sparse approximations. This simplified model of natural images emphasizing anisotropic features, most notably edges, is found to be consistent with models of early stages of the HVS [18].

Shearlet coefficients $c_{l,k}(x,y)$ of an image can be pooled for each scale l by max-pooling the absolute values along the orientational shearings k to a scalewise feature map

$$c_l(x,y) = \max_{k\in[0,\dots,K-1]} |c_{l,k}(x,y)| \,. \tag{8}$$

This kind of pooling is also neurophysiologically plausible for modeling visual processing in the human brain [19]. It was shown that the pooled coefficients $c_l(x,y)$ of an image can be modeled by an inverse Gaussian Distribution and that the position parameters μ_l are feasible features for RR IQA applications [20].

In a neighborhood of suitable size the coefficients $c_l(x,y)$ can also be modeled locally, leading to local position parameters $\mu_l(x,y)$ for each scale l. As such, $\mu_l(x,y)$ is an indicator of the local activity in an image on a certain scale l. Based on [20], a local visual activity measure over all scales can then be calculated as the harmonic mean $\mu_l(x,y)$ along the scales leaving out the lowpass band $l = 0$

$$a(x,y) = \frac{L}{\sum_{l=1}^{L} \frac{1}{\mu_l(x,y)}}. \tag{9}$$

Local visual activity $a(x,y)$ translates into negative distortion sensitivity of an image [20] (a shift to the left in Fig. 1). Thus, and in order to account for potential scaling effects we estimate the samplewise distortion sensitivity $d(x,y)$ as

$$d(x,y) = -\beta \cdot a(x,y) \tag{10}$$

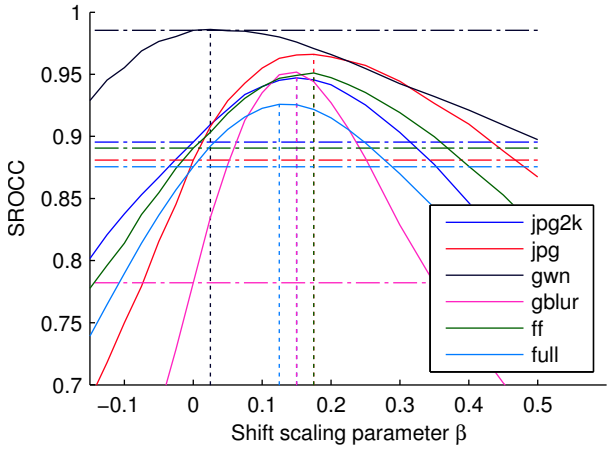
with β being a scaling parameter essentially providing an additional global offset to the shift of the PSNR in Eq. 3.

3. EXPERIMENTS

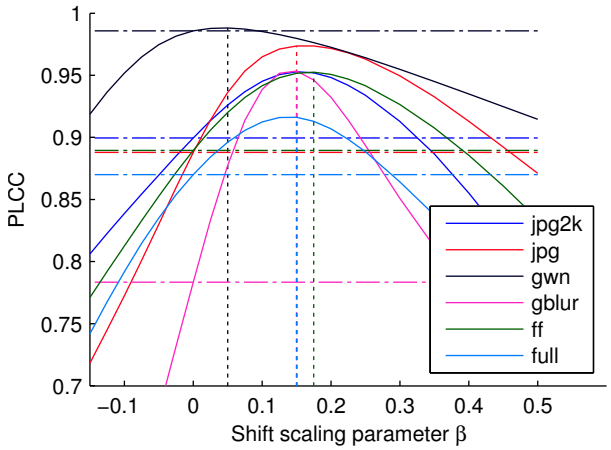
We evaluate the perceptually adapted PSNR (paPSNR) for a neighborhood size of 17 pixel.

3.1. Influence of shift parameter β

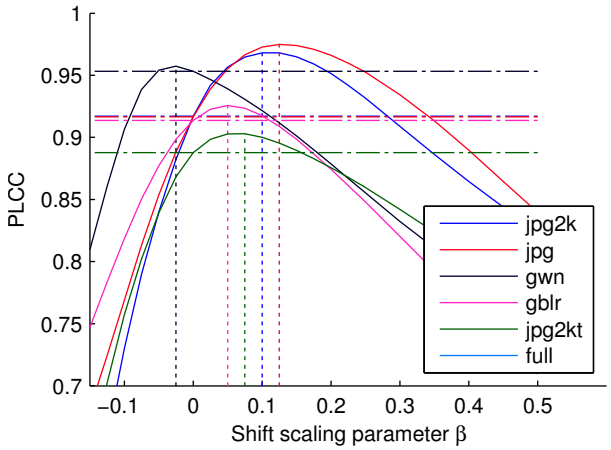
Fig. 2 shows the performance of the proposed method in dependence of the shift parameter β in terms of Pearson Linear Correlation Coefficient (PLCC) and Spearman Rank Order Correlation Coefficients (SROCC) with (D)MOS for different distortion categories of the LIVE [16] and TID2013 [17] databases (solid lines). Distortionwise PSNRs are indicated by horizontal dash-dotted lines, optimal parameters for β are indicated by vertical dotted lines, For $\beta = 0$, the paPSNR is identical to the PSNR. Although the optimal setting of β differs for different distortion types, except for additive white Gaussian noise, optimal parameters are relatively close. Interestingly, PLCC and SROCC are optimized for slightly different β (Figs. 2b and 2a). Comparing Fig. 2b and Fig. 2a shows that the optimal β also differs between the two databases.



(a) Spearman rank order correlation on LIVE database.



(b) Pearson correlation on LIVE database.



(c) Pearson correlation on TID2013 database.

Fig. 2: Influence of parameter β for different distortion categories of the LIVE and TID2013 databases. Solid lines represent correlations of paPSNR and (D)MOS for specific distortion classes. Vertical dotted lines indicate the corresponding optimal values of β, horizontal dash-dotted lines represent correlations of PSNR and (D)MOS for specific distortion classes.

		PSNR	SSIM	MS-SSIM	paPSNR $\beta = 0.1$	paPSNR $\beta = \beta_{opt}^{dt}$
LIVE	jpg2k	0.90	0.96	0.96	0.95	0.95
	jpg	0.88	0.98	0.98	0.97	0.97
	gwn	0.99	0.97	0.97	0.98	0.99
	gblur	0.78	0.95	0.95	0.95	0.95
	ff	0.89	0.96	0.95	0.95	0.95
TID2013	gwn	0.93	0.87	0.86	0.91	0.95
	scn	0.92	0.85	0.85	0.92	0.95
	mn	0.83	0.78	0.81	0.82	0.84
	hfn	0.91	0.86	0.86	0.89	0.97
	in	0.90	0.75	0.76	0.82	0.91
	gblr	0.91	0.97	0.97	0.92	0.93
	den	0.95	0.93	0.93	0.94	0.95
	jpg	0.92	0.92	0.93	0.95	0.95
	jpg2k	0.88	0.95	0.95	0.96	0.96
	mgn	0.89	0.78	0.78	0.85	0.94
	lcni	0.91	0.91	0.91	0.95	0.95

Table 1: SROCC of the proposed method for the different distortion types of LIVE database and the *actual* subset of TID2013 database with $\beta = 0.1$ and $\beta = \beta_{opt}^{dt}$ per distortion type in comparison to PSNR, SSIM and MS-SSIM.

	PSNR	SSIM	MS-SSIM	paPSNR $\beta = 0.1$	paPSNR $\beta = \beta_{opt}^{dt}$
LIVE	0.88	0.95	0.95	0.93	0.88
TID2013 (*actual*)	0.82	0.88	0.88	0.84	0.76
TID2013 (*full*)	0.64	0.74	0.79	0.65	0.58

Table 2: SROCC of the proposed method for the full LIVE and the full TID2013 databases and for the *actual* subset of the TID2013 database with $\beta = 0.1$ and $\beta = \beta_{opt}^{dt}$ per distortion type in comparison to PSNR, SSIM and MS-SSIM.

3.2. Performance of the paPSNR

The performance of the proposed method is evaluated on the LIVE and the TID2013 databases and different subgroups for two different settings of β and compared to SSIM [6] and MS-SSIM [7]. In one setting we set $\beta = 0.1$ for all calculations, in the other we set $\beta = \beta_{opt}^{dt}$ to the distortion type specific optimal parameter with regard to the SROCC.

The results in terms of SROCC are summarized in Tab. 1 for the LIVE database and TID2013 distortions of the *actual* subset of TID2013, containing distortion types of practical relevance. With β being selected distortion specific, the proposed method shows superior performance compared to PSNR, SSIM and MS-SSIM. With β being fixed to $\beta = 0.1$ paPSNR performs comparable or superior for most of the distortion types. When evaluated on the full datasets, paPSNR with $\beta = 0.1$ perform a little inferior but comparable on LIVE, inferior on the actual subset of TID2013 and significantly worse on the full dataset of TID2013 compared to SSIM and MS-SSIM. Interestingly, the performance decreases even further, if β is adapted distortion wise with $\beta = \beta_{opt}^{dt}$. The reason is exemplified in Fig. 3: Although setting $\beta = \beta_{opt}^{dt}$ reduces the variance within a distortion type, the quality estimates of different distortion types become more separated leading to a reduction of the SROCC.

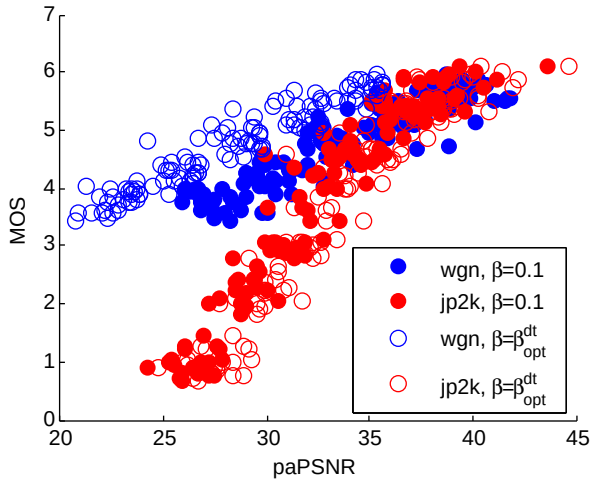


Fig. 3: paPSNR vs. MOS for the JPEG2K (red) and the additive white Gaussian noise (blue) subset of TID2013 for $\beta = 0.1$ (filled) and $\beta = \beta_{opt}^{dt}$ (unfilled).

4. CONCLUSION

Motivated by the non-linear mapping of computational quality values Q_c to subjective quality scores Q_p and a discussion of the parameters of the mapping function we derived the concept of distortion sensitivity and argued that a shift of the parameter d in the mapping function can serve as an approximation of distortion sensitivity. We showed that the harmonic mean of the scalewise distribution of shearlet coefficients is a feasible estimator of d and can be used for improving the performance of the a computational quality metric. This was exemplify by adapting the PSNR. Although the performance of the PSNR can be increased consistently for specific distortion types the approach method is not able to estimate the quality of image subject to an unknown distortion type. This might be corrected by identifying the distortion type in a first step and adjusting β to the specific distortion in a second step. The proposed approach may be improved further by analyzing the influence of neighborhood sizes other than 17 pixel for estimating the local location parameter of the shearlet coefficients and by testing other scalewise pooling methods than the harmonic mean. The non-linear mapping function also suggests an influence of the slope parameter c and potentially an interdependency of c and d. Modeling this may further increase the performance and improve generalization. In principle, the presented approach could be used for improving any other IQM as well. The PSNR has the convenient property of being computationally low-complex, which makes the proposed method a feasible approach for real-time systems. This is especially interesting, e.g. for controlling the mode decision in a video encoder, as β only has to be computed once for each block and only based on the reference image.

5. REFERENCES

[1] S. E. Palmer, *Vision science: Photons to phenomenology*, vol. 1, MIT press Cambridge, MA, 1999.

[2] B. Girod, "What's Wrong with Mean-squared Error?," in *Digital Images and Human Vision*, pp. 207–220. 1993.

[3] W. Lin and C.-C. J. Kuo, "Perceptual visual quality metrics: A survey," *Journal of Visual Communication and Image Representation*, vol. 22, no. 4, pp. 297–312, 2011.

[4] S. J. Daly, "Application of a noise-adaptive contrast sensitivity function to image data compression," *Optical Engineering*, vol. 29, no. 8, pp. 977–987, 1990.

[5] J. Lubin, "A human vision system model for objective picture quality measurements," *International Broadcasting Convention*, pp. 498–503, 1997.

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

[7] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multiscale structural similarity for image quality assessment," *IEEE Asilomar Conference on Signals, Systems and Computers*, vol. 2, no. 1, pp. 1398–1402, 2003.

[8] B. Ortiz-Jaramillo, J. Niño-Castañeda, L. Platiša, and W. Philips, "Content-aware video quality assessment: predicting human perception of quality using peak signal to noise ratio and spatial/temporal activity," in *SPIE/IS&T Electronic Imaging*. International Society for Optics and Photonics, 2015, vol. 9399, p. 939917.

[9] W. Zhang, A. Borji, Z. Wang, P. Le Callet, and H. Liu, "The application of visual saliency models in objective image quality assessment: A statistical evaluation," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 27, no. 6, pp. 1266–1278, 2016.

[10] V. Laparra, J. Ballé, A. Berardino, and E. P. Simoncelli, "Perceptual image quality assessment using a normalized Laplacian pyramid," *Electronic Imaging*, vol. 2016, no. 16, pp. 1–6, 2016.

[11] Z. Wang and Q. Li, "Information content weighting for perceptual image quality assessment," *IEEE Transactions on Image Processing*, vol. 20, no. 5, pp. 1185–1198, 2011.

[12] S. Hu, L. Jin, H. Wang, Y. Zhang, S. Kwong, and C.-C. J. Kuo, "Compressed Image Quality Metric Based on Perceptually Weighted Distortion," *IEEE Transactions on Image Processing*, vol. 24, no. 12, pp. 5594–5608, 2015.

[13] S. Bosse, D. Maniry, K.-R. Müller, T. Wiegand, and W. Samek, "Neural Network-Based Full-Reference Image Quality Assessment," in *Proceedings of the Picture Coding Symposium (PCS)*, 2016.

[14] S. Bosse, D. Maniry, T. Wiegand, and W. Samek, "A deep neural network for image quality assessment," in *2016 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2016, pp. 3773–3777.

[15] VQEG, "Objective perceptual assessment of video quality: full reference television," *ITU-T Telecommunication Standardization Bureau*, 2004.

[16] H. R. Sheikh, M. F. Sabir, and A. C. Bovik, "A statistical evaluation of recent full reference image quality assessment algorithms," *IEEE Transactions on image processing*, vol. 15, no. 11, pp. 3440–51, nov 2006.

[17] N. Ponomarenko, O. Ieremeiev, V. Lukin, K. Egiazarian, L. Jin, J. Astola, B. Vozel, K. Chehdi, M. Carli, F. Battisti, and C.-C. J. Kuo, "Color Image Database TID2013: Peculiarities and Preliminary Results," *4th European Workshop on Visual Information Processing (EUVIP)*, pp. 106–111, 2013.

[18] G. Kutyniok and D. Labate, *Shearlets: Multiscale Analysis for Multivariate Data*, Birkhäuser Basel, 2012.

[19] I. Lampl, D. Ferster, T. Poggio, and M. Riesenhuber, "Intracellular Measurements of Spatial Integration and the MAX Operation in Complex Cells of the Cat Primary Visual Cortex," *Journal of Neurophysiology*, vol. 92, pp. 2704–2713, 2004.

[20] S. Bosse, Q. Chen, M. Siekmann, W. Samek, and T. Wiegand, "Shearlet-based reduced reference image quality assessment," in *2016 IEEE International Conference on Image Processing (ICIP)*, 2016, pp. 2052–2056.