

PerSIM: MULTI-RESOLUTION IMAGE QUALITY ASSESSMENT IN THE PERCEPTUALLY UNIFORM COLOR DOMAIN

Dogancan Temel and Ghassan AlRegib

Center for Signal and Information Processing (CSIP)
School of Electrical and Computer Engineering
Georgia Institute of Technology, Atlanta, GA, 30332-0250 USA
{cantemel,alregib}@gatech.edu

ABSTRACT

An average observer perceives the world in color instead of black and white. Moreover, the visual system focuses on structures and segments instead of individual pixels. Based on these observations, we propose a full reference objective image quality metric modeling visual system characteristics and chroma similarity in the perceptually uniform color domain (Lab). Laplacian of Gaussian features are obtained in the L channel to model the retinal ganglion cells in human visual system and color similarity is calculated over the a and b channels. In the proposed perceptual similarity index (PerSIM), a multi-resolution approach is followed to mimic the hierarchical nature of human visual system. LIVE and TID2013 databases are used in the validation and PerSIM outperforms all the compared metrics in the overall databases in terms of ranking, monotonic behavior and linearity.

Index Terms— image quality analysis, human visual system, perception, LoG features, similarity index

1. INTRODUCTION

Image quality metrics are designed to estimate the perceived quality. A full reference objective image quality metric tries to quantify the differences between the original and the distorted image if both the images are available. Root mean-squared error (RMSE) is calculated by obtaining the pixel-wise difference between the two images, taking the square root of the difference and calculating the mean. RMSE is scaled with the bit depth of the image and mapped using a monotonic logarithmic function to obtain the peak signal-to-noise ratio (PSNR). The authors in [1] introduce PSNR-HVS by stretching the contrast block-wise, quantizing the DCT coefficients with the JPEG compression table and removing mean shift. PSNR-HVS is further modified by multiplying the difference between the DCT coefficients with contrast masking metric (PSNR-HVS-M) [2]. The authors in [3] add mean shift sensitivity and contrast change to the pixel-wise metrics and the modified versions are denoted as PSNR-HA and PSNR-HMA. Signal-to-noise ratio (SNR) is weighted by the authors in [4] using contrast sensitivity function and the authors in [5] use wavelet-based models of visual masking.

In addition to pixel-wise metrics, structural metrics are also commonly used to estimate the quality of images. The authors in [6] compare the reference and distorted images in terms of luminance, contrast and structure similarity in the spatial domain to estimate the image quality. These structure-based metrics are also extended to multi-scale (MS-SSIM) [7], complex domain (CW-SSIM) [8] and information-weighted (IW-SSIM) [9] versions. Instead of directly using the pixel values, phase and magnitude of the images can also be used separately to estimate the quality. FSIM is a feature similarity index introduced by the authors in [10] which consists of phase congruency (PC) and gradient magnitude (GM). PC consists of log-Gabor filter and Gaussian spread function and GM is based on gradient operators. The feature similarity index is further extended as FSIM-c which also includes the color similarity in the YIQ domain. GM is utilized along with the LoG features to obtain joint statistics that can be used for blind image quality assessment by the authors in [11]. LoG features are also used in [12] to directly assess the image quality but they overlook the hierarchical procedure and color perception in the visual system.

In this paper, we operate in the perceptually uniform Lab color space where luma and chroma information are separated. Retinal ganglion cells in the visual system are modeled using the LoG features in the L channel. Chroma similarity is calculated over the a and b channels. We obtain the similarity maps at different resolutions and calculate the geometric mean of these maps to obtain the multi-resolution similarity maps. LoG and chroma similarities are tuned using the ratios in the 4:2:2 chroma sub-sampling format. After sensitivity tuning, the minimum similarity is selected pixel-wise over the quality maps. Mean pooling is performed over the full map to calculate a single quality value. The resulting value is monotonically mapped by taking the power to obtain the perceptual similarity index PerSIM.

2. PERSIM

2.1. Log Features

Instead of using the pixel values as raw data, image features are extracted to represent the images in a more compact and distinctive way. Difference of Gaussian and Laplacian

of Gaussian are among the most commonly used operators in the image processing literature and the computer vision literature. Difference of Gaussian operators can be used to model the retinal Ganglion cells of the cat as discussed in [13]. Moreover, the authors in [14] discuss that Gaussian derivative-like approaches can model neural mechanisms in the human foveal retinal vision. These Gaussian derivative-like approaches also outperform the Gabor filter-based models according to model-free Wiener filter analysis as explained in [14]. In the difference of Gaussian models, standard deviation and the scale of the difference need to be tuned to obtain distinctive features. In case of using various scales, fusion of these models also becomes an issue. Difference of Gaussian operator can be used as an approximation to the second derivative of Gaussian when the scale is adjusted. And the second derivative of Gaussian corresponds to the Laplacian of Gaussian operator. In order to avoid the tuning of the scale and simplify the problem, we use Laplacian of Gaussian as formulated in Eq. (1).

$$\hat{LoG} = \frac{1}{\sqrt{2\pi\sigma^2}} \frac{m^2 + n^2 - 2\sigma^2}{\sigma^4} e^{-(m^2+n^2)/(2\sigma^2)} \quad (1)$$

The standard deviation of the LoG operator is represented with σ and m and n are the respective pixel locations. Reference (f_1) and distorted (f_2) images are convolved with the LoG operator as formulated in Eq. (2) where i corresponds to the image index.

$$LoG_i = f_i[m, n] * \hat{LoG}[m, n] \quad (2)$$

The similarity between LoG maps is calculated using the familiar similarity formulation that has been part of most of the structural and pixel-wise comparison metrics as expressed in Eq. (3).

$$LoGSIM[m, n] = \frac{2 \cdot LoG_1[m, n] \cdot LoG_2[m, n] + c_1}{(LoG_1[m, n])^2 + (LoG_2[m, n])^2 + c_2} \quad (3)$$

Similarity metric becomes 1.0 when the images are same and it gets closer to 0.0 as the differences between images become very large. We set constants c_1 and c_2 to 0.001 to avoid the issues when the denominator converges to 0.0.

2.2. Color Similarity

Color similarity is directly calculated over the a and the b channels separately. We use the similarity formulation as expressed in Eq. (4) and Eq. (5).

$$aSIM[m, n] = \frac{2 \cdot a_1[m, n] \cdot a_2[m, n] + c_3}{(a_1[m, n])^2 + (a_2[m, n])^2 + c_4} \quad (4)$$

$$bSIM[m, n] = \frac{2 \cdot b_1[m, n] \cdot b_2[m, n] + c_5}{(b_1[m, n])^2 + (b_2[m, n])^2 + c_6}, \quad (5)$$

where $a_1[m, n]$ and $b_1[m, n]$ are the chroma channels in the reference image and $a_2[m, n]$ and $b_2[m, n]$ are the chroma channels in the distorted image. c_3 , c_4 , c_5 and c_6 are the constants set to 0.001.

2.3. Fusion

The human visual system is more sensitive to structural information compared to color. Based on this observation, chroma sub-sampling is introduced in image and video coding to assign less resolution to chroma information. 4:2:2 is one of the most commonly used chroma sub-sampling format where chroma channels get half the resolution of luma channels. In the proposed work, we follow a similar approach and tune the significance of the intensity and the color-based components. The power of LoG similarity is set to 4.0 and the powers of similarities for chroma channels are set to 2.0. After this sensitivity adjustment, we choose the minimum among the similarity indexes as formulated in Eq. (6) because the perceived quality is dominated by the most significant degradation.

$$LabSIM_{SR}[m, n] = \min((LoGSIM[m, n])^4, (aSIM[m, n])^2, (bSIM[m, n])^2) \quad (6)$$

We perform mean pooling to obtain a single quality value corresponding to the distortion map. Similarity is calculated over the full feature map so the pixels that are slightly distorted would bias the metric to be close to 1.0. In order to increase the variation of the metric and spread the range of the estimations, we monotonically scale the resulting value with a power function as given in Eq. (7).

$$PerSIM_{SR} = \left(\sum_{m=1}^M \sum_{n=1}^N \frac{LabSIM_{SR}[m, n]}{M \cdot N} \right)^{c_7} \quad (7)$$

Mean pooling is performed over the whole image where M is number of rows, N is the number of columns and c_7 is the power index. SR refers to single resolution since we use the reference and distorted images at the original resolution. Power index is set to 25 so that the metric similarity index goes down to 0.0 under severe degradation. Power indexes less than 25 does not use the full metric range and indexes more than 25 become extra sensitive to even slight degradation. This monotonic scaling does not bias the results since ranking-based validation metrics are insensitive to the monotonic mapping and the regression step before the linear correlation calculation perform monotonic mapping automatically.

2.4. Multi-Resolution

Perception in the visual system is hierarchical. At first, the raw data is acquired with the sensor-like structures. Then, the data is processed and transferred into different abstraction layers with varying resolutions. Different features and regions of interest can be more distinctive at different resolutions. Therefore, we calculate the perceptual similarity maps at different resolutions and fuse them together. We start by calculating $LoGSIM$, $aSIM$ and $bSIM$ over three different resolutions. The first set is calculated over the full resolution while the second and third are calculated at 3/5 and 2/5 times

the full resolution, respectively. We refer to all the maps as $LoGSIM$, $aSIM$ and $bSIM$ and the scales of the resolution are shown with a subscript. LoG features and chroma similarities are extracted over the scaled maps and then interpolated to the original resolution using the bicubic approach. Since the average value and range of the metrics are not known, we directly calculate the geometric mean of the interpolated maps pixel-wise to obtain the multi-resolution perceptual similarity map as formulated in Eqs. (8)-(10).

$$LoGSIM_{MR}[m, n] = \sqrt[3]{LoGSIM_{1.0} \cdot LoGSIM_{0.6} \cdot LoGSIM_{0.4}} \quad (8)$$

$$aSIM_{MR}[m, n] = \sqrt[3]{aSIM_{1.0} \cdot aSIM_{0.6} \cdot aSIM_{0.4}} \quad (9)$$

$$bSIM_{MR}[m, n] = \sqrt[3]{bSIM_{1.0} \cdot bSIM_{0.6} \cdot bSIM_{0.4}} \quad (10)$$

Multi-resolution indexes are combined in the same way as the single resolution given in Eq. (11).

$$LabSIM_{MR}[m, n] = \min((LoGSIM_{MR}[m, n])^4, (aSIM_{MR}[m, n])^2, (bSIM_{MR}[m, n])^2) \quad (11)$$

Finally, multi-resolution perceptual quality map is mean pooled and monotonically mapped as formulated in Eq. (12).

$$PerSIM = \left(\sum_{m=1}^M \sum_{n=1}^N \frac{LabSIM_{MR}[m, n]}{M \cdot N} \right)^{25} \quad (12)$$

As the resolution gets lower, it becomes more challenging to detect distinctive features. Therefore, we decrease the block-size and standard deviation accordingly as tabulated in Table 1. The scale values, standard deviation and block size are selected by visually assessing the distinctiveness of randomly selected feature maps.

Table 1. Multi-resolution PerSIM parameters

Scaling Ratio	Standard Deviation	Block Size
1.0	10.0	13x13
0.6	8.0	4x4
0.4	7.0	2x2

3. VALIDATION

LIVE and TID2013 image databases are used in the validation of PerSIM. LIVE database includes 29 reference images and 779 degraded images under the distortion of JPEG, JPEG2000 (Jp2k), White Noise (Wn), Gaussian blur (Gblur) and Fast Fading Rayleigh channel errors (FF). TID2013 consists of 25 reference images that are originally from Kodak Lossless True Color Image Suite [15]. Reference images are degraded with 24 different types of distortions that fall into the categories of Noise, Actual, Simple, Exotic, New and Color. TID2013 database is introduced in [16] where ranking based metrics Spearman and Kendall correlation coefficients are used for the validation. Therefore, we follow the same approach as in [16].

Objective image quality metrics are defined in different numerical ranges and monotonic regression is necessary for a

Table 2. LIVE Results

Sequence	Jp2k	Jpeg	Wn	Gblur	FF	All
	Pearson (PLCC)					
PSNR	0.923	0.913	0.945	0.843	0.887	0.898
SSIM	0.963	0.957	0.976	0.940	0.956	0.945
MS-SSIM	0.962	0.961	0.977	0.943	0.948	0.946
IW-SSIM	0.959	0.959	0.981	0.957	0.953	0.951
FSIMc	0.960	0.953	0.977	0.955	0.953	0.950
PSNR-HA	0.976	0.971	0.980	0.935	0.953	0.953
CW-SSIM	0.926	0.927	0.949	0.768	0.835	0.872
LogSIM	0.956	0.952	0.987	0.943	0.942	0.943
PerSIM	0.976	0.959	0.968	0.967	0.946	0.955
	RMSE					
PSNR	9.92	10.10	8.34	11.80	10.22	10.12
SSIM	7.11	7.74	8.65	7.54	6.45	7.52
MS-SSIM	7.12	7.30	8.38	7.38	7.04	7.43
IW-SSIM	7.38	7.64	6.95	6.37	6.86	7.11
FSIMc	7.55	7.73	6.97	6.71	6.68	7.20
PSNRHA	6.59	6.89	5.98	8.27	6.83	6.93
CW-SSIM	9.75	9.30	9.24	14.45	13.62	10.87
LogSIM	7.93	8.36	6.06	7.70	7.71	7.66
PerSIM	6.22	7.30	7.25	5.68	7.36	6.80

fair comparison if validation includes metrics that are based on linearity and accuracy including but not limited to Pearson linear correlation coefficient (PLCC) and root mean-squared error (RMSE). In the literature, validation of the metrics in the LIVE database are mostly based on PLCC and RMSE and the function formulated in Eq. (13) is used for monotonic regression. Therefore, we also calculate the PLCC and the RMSE after the monotonic regression as tabulated in Table 2.

$$S = \beta_1 \left(\frac{1}{1 - \frac{1}{2 + \exp(\beta_2(S_0 - \beta_3))}} \right) + \beta_4 S_0 + \beta_5 \quad (13)$$

We use the TID2013 benchmark data to compare the proposed metric with the state of the art. In the LIVE database, we use the common structural metrics and the best performing ones in the TID benchmark. In order to show the effect of color similarity in the overall metric, we replace the $LabSIM_{MR}[m, n]$ in Eq. 12 with $LoGSIM_{MR}[m, n]$ and report the results as $LogSIM$. Top three performance values are highlighted in the results tables to indicate best performing metrics. In the LIVE database, PerSIM is among the top metrics in the compression-related degradation JPEG2000 and JPEG and also in Gaussian blur. However, PerSIM is not as good as structure and phase conjugacy-based metrics in case of White noise and Fastfading artifacts. White noise artifacts are captured by the LoG features but color similarity is less sensitive to these artifacts. In case of the Fastfading, communication channel errors can lead to local errors that are perceptually very disturbing but they would be overlooked by PerSIM since relative size of the errors can be negligible compared to the rest of the sharp transitions in the image. In the overall LIVE database, PerSIM still performs better than the compared metrics.

The performance of PerSIM in the TID2013 database is tabulated in Table 3. According to the validation results, PerSIM is among the best performing metrics in all the cat-

Table 3. TID2013 Results

Sequence	Spearman (SROCC)							Kendall (KCC)						
	Noise	Actual	Simple	Exotic	New	Color	Full	Noise	Actual	Simple	Exotic	New	Color	Full
FSIM-c	0.902	0.915	0.947	0.841	0.788	0.775	0.851	0.722	0.742	0.792	0.651	0.611	0.592	0.666
PSNR-HA	0.923	0.938	0.953	0.825	0.701	0.632	0.819	0.760	0.787	0.818	0.624	0.541	0.477	0.643
PSNR-HMA	0.915	0.934	0.937	0.814	0.738	0.675	0.813	0.745	0.777	0.785	0.610	0.572	0.507	0.631
FSIM	0.897	0.911	0.949	0.844	0.649	0.565	0.801	0.715	0.736	0.795	0.655	0.518	0.447	0.629
MS-SSIM	0.873	0.887	0.905	0.841	0.631	0.566	0.787	0.679	0.697	0.720	0.647	0.490	0.450	0.607
IW-SSIM	0.871	0.887	0.911	0.840	0.619	0.549	0.778	0.678	0.701	0.730	0.644	0.475	0.424	0.597
PSNRc	0.769	0.803	0.876	0.562	0.777	0.734	0.687	0.562	0.596	0.689	0.392	0.576	0.536	0.496
VSNR	0.869	0.882	0.912	0.706	0.589	0.512	0.681	0.676	0.690	0.731	0.519	0.437	0.378	0.508
PSNR-HVS	0.917	0.926	0.951	0.601	0.646	0.555	0.654	0.754	0.766	0.809	0.435	0.512	0.441	0.507
PSNR	0.822	0.825	0.913	0.597	0.618	0.535	0.640	0.623	0.624	0.745	0.425	0.468	0.408	0.470
SSIM	0.757	0.788	0.837	0.632	0.579	0.505	0.637	0.551	0.577	0.628	0.455	0.418	0.378	0.463
NQM	0.836	0.857	0.875	0.589	0.625	0.538	0.635	0.641	0.666	0.681	0.412	0.478	0.401	0.466
PSNR-HVS-M	0.906	0.917	0.938	0.564	0.646	0.553	0.625	0.733	0.749	0.780	0.403	0.513	0.433	0.481
VIFP	0.784	0.815	0.897	0.557	0.589	0.506	0.608	0.587	0.621	0.714	0.406	0.445	0.385	0.456
WSNR	0.880	0.897	0.933	0.423	0.646	0.555	0.580	0.696	0.718	0.772	0.297	0.510	0.429	0.446
LogSIM	0.910	0.923	0.947	0.806	0.662	0.604	0.787	0.736	0.756	0.799	0.615	0.521	0.473	0.618
PerSIM	0.925	0.936	0.950	0.799	0.863	0.856	0.854	0.760	0.778	0.807	0.606	0.681	0.674	0.677

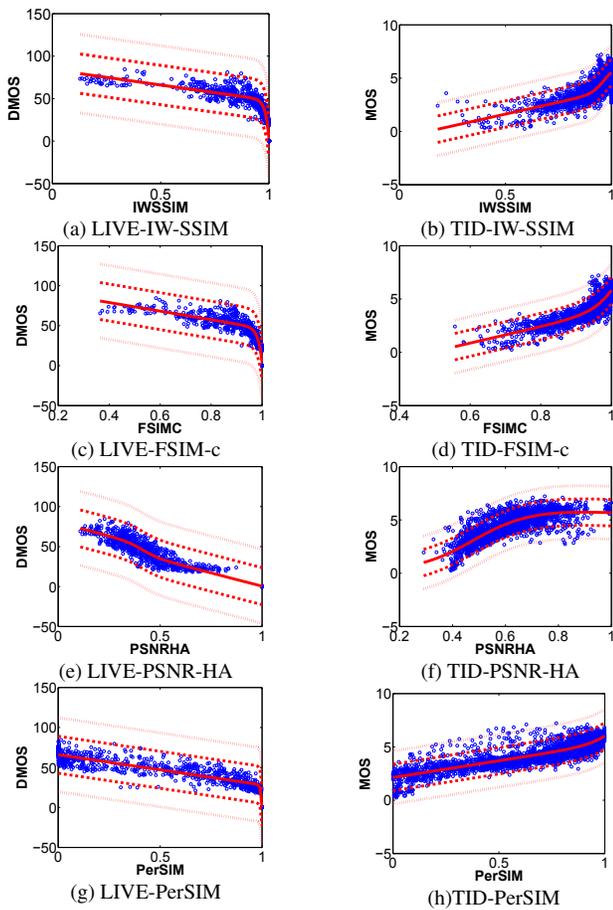


Fig. 1. Scatter plots of objective image quality metrics values

egories except Exotic. Exotic category includes local block-wise distortions and JPEG2000 transmission errors that can lead to local and slice-based distortions that are overlooked by the LoG features and color similarity. In the overall TID2013, PerSIM is still the best performing objective quality metric among the compared ones. As it can be seen in Table 2 and

Table 3, the performance of the quality metric over the full image set degrades without the color similarity in both LIVE and TID2013.

If we compare the metrics in both LIVE and TID2013 databases, the best performing metrics are IW-SSIM, FSIMc, PSNR-HA and PerSIM. The scatter plots of the best performing metrics are given in Fig. 1 to observe the distributional characteristics. Information theoretic-weighting based structural metric (IW-SSIM) scores mostly cluster around the high quality region and the same observation is valid for the metric based on phase-conjugacy (FSIMc). PSNR-HA has an outlier problem when we include identical images and the range of the metric is not bounded. In order to obtain the scatter plot in Fig. 1(f), we exclude the outliers. PSNR-HA estimates are mostly centered in the metric range and it has a higher linearity compared to the structural and phase-conjugacy-based metrics. PerSIM scores are distributed in the full metric range and show a high linearity. Almost all the estimates are in the one standard deviation range in the LIVE database. However, in the TID2013 database, some of the estimates are located between one and two standard deviation and only a minority is located outside of two standard deviation range. Most of the outliers in the TID2013 database correspond to the Exotic class since PerSIM has difficulty in capturing local degradations.

4. CONCLUSION

We proposed a full reference multi-resolution image quality metric based on LoG features and chroma similarity in the perceptually uniform Lab domain. LoG features are used to model the retinal ganglion cells in the human visual system and the color similarity complements the structural similarity. The results in the LIVE and TID2013 database show that PerSIM outperforms state of the art metrics in terms of monotonicity, accuracy and linearity. Even PerSIM detects majority of the distortions accurately, it overlooks local distortions. As an ongoing work, we are working on a smarter pooling strategy to make the metric sensitive to local distortions.

5. REFERENCES

- [1] K. Egiazarian, J. Astola, N. Ponomarenko, V. Lukin, F. Battisti, and M. Carli, "A New Full-reference Quality Metrics based on HVS," in *Proceedings of the Second International Workshop on Video Processing and Quality Metrics*, 2006.
- [2] N. Ponomarenko, F. Silvestri, K. Egiazarian, M. Carli, J. Astola, and V. Lukin, "On between-coefficient contrast masking of dct basis functions," in *Proceedings of the Second International Workshop on Video Processing and Quality Metrics*, 2007, pp. 1–4.
- [3] N. Ponomarenko, O. Ieremeiev, V. Lukin, K. Egiazarian, and M. Carli, "Modified Image Visual Quality Metrics for Contrast Change and Mean Shift Accounting," *Proceedings of CADSM*, 2011.
- [4] T. Mitsa and K. L. Varkur, "Evaluation of Contrast Sensitivity Functions for the Formulation of Quality Measures Incorporated in Halftoning Algorithms," in *ICASSP*, 1993.
- [5] D. M. Chandler and S. S. Hemami, "VSNR: a wavelet-based visual signal-to-noise ratio for natural images.," *IEEE Transactions on Image Processing*, vol. 16, no. 9, pp. 2284–98, Sept. 2007.
- [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 : a publication of the IEEE Signal Processing Society*, vol. 13, no. 4, pp. 600–12, Apr. 2004.
- [7] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multi-Scale Structural Similarity For Image Quality Assessment (Invited Paper)," *the Thirty-Seventh Asilomar Conference on Signals, Systems and Computers*, vol. 2, pp. 9–13, 2004.
- [8] Z. Wang and E. P. Simoncelli, "Translation Insensitive Image Similarity In Complex Wavelet Domain Zhou Wang and Eero P . Simoncelli," vol. II, no. March, pp. 573–576, 2005.
- [9] Z. Wang and Q. Li, "Information Content Weighting for Perceptual Image Quality Assessment.," *IEEE Transactions on Image Processing*, vol. 20, no. 5, pp. 1185–98, May 2011.
- [10] L. Zhang, L. Zhang, X. Mou, and D. Zhang, "FSIM: A Feature Similarity Index for Image Quality Assessment.," *IEEE Transactions on Image Processing*, vol. 20, no. 8, pp. 2378–86, Aug. 2011.
- [11] W. Xue, X. Mou, L. Zhang, A.C. Bovik, and X. Feng, "Blind Image Quality Assessment Using Joint Statistics of Gradient Magnitude and Laplacian Features," *Image Processing, IEEE Transactions on*, vol. 23, no. 11, pp. 4850–4862, Nov 2014.
- [12] X. Mou, W. Xue, C. Chen, and L. Zhang, "LoG Acts as a Good Feature in the Task of Image Quality Assessment.," *Proc. SPIE*, vol. 9023, pp. 902313–902313–7, 2014.
- [13] C. Enroth-Cugell and J. G. Robson, "The Contrast Sensitivity of Retinal Ganglion Cells of the Cat," *The Journal of Physiology*, 1966.
- [14] R. A. Young, "The Gaussian Derivative Model for Spatial Vision: I. Retinal Mechanisms," *Spatial Vision*, 1987.
- [15] Eastman Kodak Company, "Lossless True Color Image Suite," <http://r0k.us/graphics/kodak/>, [Online].
- [16] 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," pp. 106–111, June 2013.