## RESEARCH

# Multi-level contrast filtering in image difference metrics

Gabriele Simone[1]*, Marius Pedersen[1], Ivar Farup[1] and Claudio Oleari[2]

## Abstract

In this paper, we present a new metric to estimate the perceived difference in contrast between an original image and a reproduction. This metric, named weighted-level framework $\Delta E_E$ (WLF-DEE), implements a multilevel filtering based on the difference of Gaussians model proposed by Tadmor and Tolhurst (2000) and the new Euclidean color difference formula in log-compressed OSA-UCS space proposed by Oleari et al. (2009). Extensive tests and analysis are presented on four different categories belonging to the well-known Tampere Image Database and on two databases developed at our institution, providing different distortions directly related to color and contrast. Comparisons in performance with other state-of-the-art metrics are also pointed out. Results promote WLF-DEE as a new stable metric for estimating the perceived magnitude of contrast between an original and a reproduction.

## 1 Introduction

The quality of image reproduction depends on many factors that cannot be completely distinguished since they are mutually dependent. It is generally considered that the quality of color in an image reproduction may be quantified by measuring how accurate the reproduction of color is to the original when viewed by a human observer. Color discrimination is determined by many factors, including the spatial pattern of the image and the visual processing, which starts with cone activation and is followed by adaptation to the illumination. Images are constituted by spatial color patterns, which are different from the uniform color patches considered in the colorant factories. Color discrimination and appearance in images are a function of spatial pattern, and color differences are harder to detect at higher frequencies, where contrast plays an important role. Moreover, color opponencies have a role in color discrimination, especially differences along the blue-yellow color direction [1]. As a consequence, for measuring the perceptual difference between an original image and its reproduction, a perceptual image difference metric is needed. Generally, this metric is obtained by extending the color difference formulae to the color of a complex image. The history of the color metrics is more than one century long.

Over the years, many image difference metrics have been proposed [2], some for measuring general image quality and image difference and some for detecting specific distortions. However, at the moment, no universal image difference metric exists. Image difference metrics are based on a number of different ideas but are usually following a general framework. The image and its reproduction are first transformed into a suitable color space, preferably a perceptually uniform one. Then, a simulation of the human visual system (HVS) is carried out, from simplistic methods as smoothing of the image by a local neighborhood to more complex methods, e.g. using contrast sensitivity functions (CSFs). Finally, the difference is calculated usually by a color difference formula. Thus, an efficient image difference metric needs a filter which is able to mimic the HVS and a suitable color difference formula. This work will mainly focus on the filtering side, particularly taking into account that contrast is an important image attribute playing an important role in image quality discrimination [3]. However, recent improvements of color difference formula and image processing methods will be considered for the development of our metrics.

After introducing the improvements of color difference formulae in Section 2 and image difference metrics in Section 3, we present our proposal in Section 4. We propose a new image difference metrics based on multi-level contrast filtering using the difference of Gaussians

*Correspondence: gabriele.simone@hig.no
[1] The Norwegian Colour and Visual Computing Laboratory, Faculty of Computer Science and Media Technology, Gjøvik University College, P.O. Box 191, Gjøvik 2802, Norway
Full list of author information is available at the end of the article

Springer

(DOG) model proposed by Tadmor and Tolhurst [4]. This metric will use also the new Euclidean color difference formula in log-compressed OSA-UCS color space proposed by Oleari et al. [5] instead of the traditional $\Delta E^*_{ab}$ that most of the state-of-the-art metrics generally use. Section 5 presents the description of three state-of-the-art databases used to evaluate the performance of the new metric with the experimental results and discusses how our proposal reflects perceptual image difference estimation performance. Finally, in Section 6 conclusions are drawn.

## 2 Color difference metrics

The first studies on color difference metrics date back to the end of the nineteenth century [6]. In color metrics history, a milestone is represented by the CIELAB color system [7,8] proposed by the International Commission on Illumination (CIE) in 1976 as a uniform color space, in which the color difference between two colors is represented by the Euclidean distance $\Delta E^*_{ab}$.

Since 1976, the main effort was to improve the CIELAB color difference formula, and almost all the proposals are based on the CIELAB space. Then, the structure of such a space has strongly conditioned the shape and the quality of almost all other proposals. The empirical color difference data used for evaluating and improving the color difference formulae were the supra-threshold Bradford University (BFD) ellipses [9], here represented on the $a^*_{10}b^*_{10}$ coordinates of the CIELAB space (Figure 1).

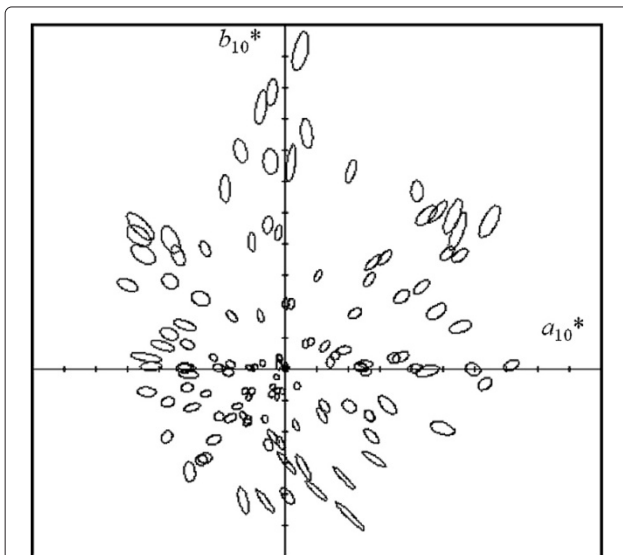In 1984, the British 'Colour-Measurement Committee' (CMC) of the 'Society of Dyers and Colourist - UK'

recommended a color difference formula, $\Delta E_{CMC}$, that has been integrated into some ISO standards [10]. The $\Delta E_{CMC}$ formula introduces a local metric tensor in the CIELAB space with the consequence that the color tolerances in the CIELAB space are represented by ellipsoids with semi-axis lengths depending on the point in the space and with one axis oriented as the lightness, one as the chroma and one as the hue, i.e. as a function of the differences of $\Delta L^*$, $\Delta C^*$ and $\Delta H^*$. Moreover, the weighting factors of the formula are hue-dependent. The CMC formula has acceptance in industrial color control applications.

In 1987, Luo and Rigg [11,12] proposed the BFD color difference formula providing a correction of the CMC formula in the blue region. In 1994, CIE proposed the non-Euclidean formula $\Delta E_{94}$ [13] by introducing a local metric tensor in the CIELAB space in analogy with the CMC formula. In 2000, CIE proposed its last color difference formula, named CIEDE2000 and denoted by $\Delta E_{00}$ [14], which is based on an enlarged dataset of empirical color differences, known as the COM dataset [14], which is added to the original BFD color difference data. This formula is only apparently based on the CIELAB space because a coordinate transformation is made in order to correct hue distortions typical of this space. The CIEDE2000 formula represents distorted ellipsoids. Although enlarged, the empirical dataset shows an evident noise, and the data related to different laboratories are not completely in agreement. The quality of these datasets combined with the large number of parameters used in their fitting induces us to suppose a risk of over-fitting.

These color difference formulae have been also applied to natural images, and evaluation of some of them for the measurement of color image reproduction quality can be found in [15-18].

In 2009, a color difference formula for small-medium color differences not based on CIELAB was published. This formula, termed $\Delta E_E$, is Euclidean [5]. The empirical color difference data plotted in the OSA-UCS space show a higher regularity than what has been seen before (Figures 2 and 3), inducing the authors to state a hue-independence hypothesis. On the basis on this hypothesis, the OSA-UCS space has been chroma log-compressed, and in this new space, the color difference formula assumes a Euclidean shape.

Since this formula is used in this work, let us recall the formula completely from Oleari et al. [5]. The transformation from the tristimulus space $(X_{10}, Y_{10}, Z_{10})$, where $Y_{10}$ is the percentage luminance factor, and the OSA-UCS space $(L_{OSA}, G, J)$ is the following: the lightness is the same as defined by the OSA-UCS committee:



**Figure 1 BFD ellipses represented on the $a^*_{10}b^*_{10}$.** The ellipses show only approximate regularity, with exclusion of the blue-magenta hues, where the ellipses are distorted. The CIEDE2000 formula is particularly complicated for regularizing such a distortion.

$$L_{OSA} \equiv 5.9 \left\{ \left[ Y_0^{1/3} - \frac{2}{3} + 0.042(Y_0 - 30)^{1/3} \right] - 14.4 \right\} \frac{1}{\sqrt{2}},$$
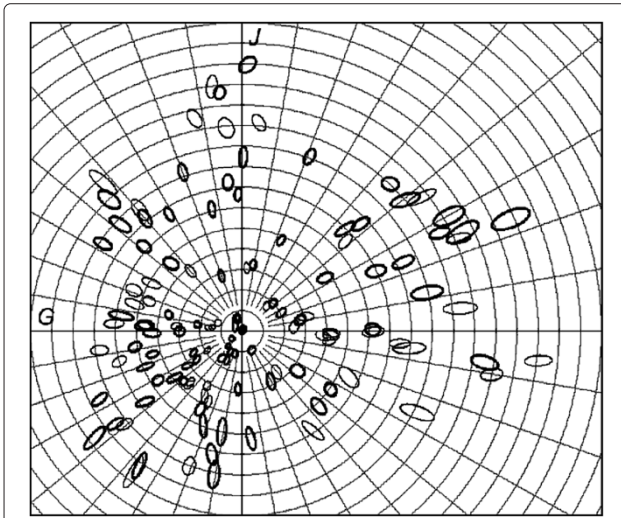
(1)

**Figure 2 BFD ellipses on the coordinates *JG* of the OSA-UCS space.** The ellipses show a general regularity that suggests the hue-independence hypothesis and consequently a simple ellipsoidal color difference formula [5,21], which is based on all the existing data (COM dataset [14]). The next step towards the Euclidean formula [21] is the logarithmic compression of the chroma (Figure 3).

with

$$Y_0 \equiv Y_{10}(4.4934x_{10}^2 + 4.3034y_{10}^2 - 4.276x_{10}y_{10} - 1.3744x_{10} - 2.5643y_{10} + 1.8103), \quad (2)$$

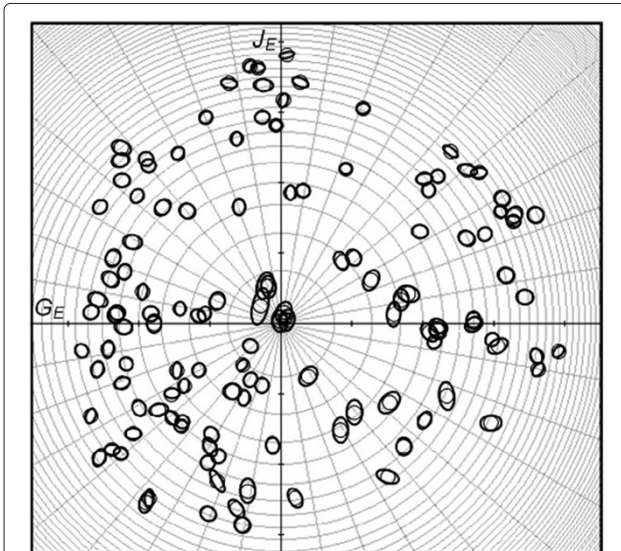the tristimulus values in the main reference frame [19,20] are



**Figure 3 OSA-UCS constant lightness plane with log-compression of the chroma (*cfr.* Figure 2).** The BFD ellipses are compared with circles with unitary radius, showing the goodness of the Euclidean color difference formula.

$$\begin{bmatrix} A \\ B \\ C \end{bmatrix} = \begin{bmatrix} 0.6597 & 0.4492 & -0.1089 \\ -0.3053 & 1.2126 & 0.0927 \\ -0.0374 & 0.4795 & 0.5579 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \quad (3)$$

(this is an active transformation that takes into account the adaptation to the visual situation used to define the OSA-UCS system), and the OSA-UCS coordinates $(g, j)$ obtained by computation and denoted by $(G, J)$ are

$$\begin{pmatrix} J \\ G \end{pmatrix} = \begin{bmatrix} 2(0.5735L_{\text{OSA}} + 7.0892) & 0 \\ 0 & -2(0.7640L_{\text{OSA}} + 9.2521) \end{bmatrix}$$
$$\begin{bmatrix} 0.1792 & 0.9837 \\ 0.9482 & -0.3175 \end{bmatrix} \begin{pmatrix} \ln\left(\frac{A/B}{0.9366}\right) \\ \ln\left(\frac{B/C}{0.9807}\right) \end{pmatrix} \quad (4)$$

$$h = \arctan\left(-\frac{J}{G}\right) \text{ and the chroma } C_{\text{OSA}} = \sqrt{G^2 + J^2}. \quad (5)$$

Then, the logarithmic compression that transforms $(L_{\text{OSA}}, C_{\text{OSA}}, h)$ into $(L_E, G_E, J_E)$ is

$$L_E = \frac{1}{b_L} \ln\left[1 + \frac{b_L}{a_L}(10L_{\text{OSA}})\right] \text{ with} \quad (6a)$$

$$a_L = 2.890, \quad (6b)$$

$$b_L = 0.015, \quad (6c)$$

$$\begin{cases} G_E = -C_E \cos(h), \\ J_E = C_E \sin(h) \end{cases} \text{ with} \quad (7a)$$

$$C_E = \frac{1}{b_C} \ln\left[1 + \frac{b_C}{a_C}(10C_{\text{OSA}})\right], \quad (8a)$$

$$a_C = 1.256, \quad (8b)$$

$$b_C = 0.050. \quad (8c)$$

Finally, the Euclidean color difference formula is

$$\Delta E_E = \sqrt{(\Delta L_E)^2 + (\Delta G_E)^2 + (\Delta J_E)^2}. \quad (9)$$

This formula is as equally good as CIEDE2000 in the prediction of many available empirical datasets, but with higher simplicity [21]. The transformation between the tristimulus space and the OSA-UCS psychometric space is simple and shows a clear relationship with visual processing. No hue distortion, as in CIELAB, exists. The chroma log compression is required in the passage from large to small color differences. Because the image quality is evaluated in the visual situation of the small-medium color differences, we think this formula is a good candidate for the color image metrics [22,23].

## 3 Image difference metrics

Nowadays, more than 100 image difference metrics are available in the literature [2], some for general image quality and image difference and some for specific distortion detection, and it would be impossible to report all of

them. For this reason, in this paper we discuss a small selection of important image difference metrics, focusing on those based on contrast filtering or which have been shown to have high correlation with observer perceived difference [2,24]. In this work, we will consider only full-reference image difference metrics, where the original and the reproduction are both available.

In 1997, Zhang and Wandell [25] proposed a spatial extension to the CIELAB color metric. This metric, named Spatial-CIELAB (S-CIELAB), should fulfill two goals: a spatial filtering to simulate the blurring of the HVS and a consistency with the basic CIELAB calculation for large uniform areas. The image is separated into an opponent color space, and each opponent color image is spatially convolved with a kernel defined on the spatial visual sensitivity for the color opponencies. Finally, the filtered image is transformed into the CIELAB system, and $\Delta E_{ab}^*$ is used for calculation of the color differences in each pixel; these are averaged to obtain a single quality value for the whole image.

The structural similarity (SSIM) index, proposed by Wang et al. in 2004 [26], attempts to quantify the visible difference between a distorted image and a reference image. This index is based on the UIQ index of Wang and Bovik [27]. The algorithm defines the structural information in an image as those attributes that represent the structure of the objects in the scene, independent of the average luminance and contrast. The index is based on a combination of luminance, contrast and structure comparison. The comparisons are done for local windows in the image, and the overall image quality is the mean of all these local windows.

In 2006, Egiazarian et al. [28] proposed the peak signal-to-noise ratio (PSNR)-HVS metric based on the HVS and PSNR, which is a measure of the peak error between the original and the reproduction. The metric uses a scanning window to remove mean shift and contrast stretching similar to UIQ [27]. PSNR-HVS is then calculated on the scanned images by using PSNR, where MSE is calculated as described by Nill [29]. An extension of this metric that incorporates contrast masking, PSNR-HVS-M, was given by Ponomarenko et al. [30].

In 2007, Chandler and Hemami [31] proposed a new metric, termed visual signal-to-noise ratio (VSNR), based on near-threshold and supra-threshold properties of the human visual system, incorporating both low-level and mid-level features. Low-level features are related to contrast sensitivity and mid-level features to global precedence, which states that the HVS visually integrates image edges in a coarse-to-fine-scale (global-to-local) fashion. The metric consists of two stages: First, contrast thresholds are used to detect visible distortions in the image, which is done in the wavelet domain by computing the contrast signal-to-noise ratio (CSNR). Then, the contrast

detection threshold is computed based on the CSNR, which is done for each octave band. The contrast is then compared with the detection threshold, and if above, the distortion is assumed supra-threshold (visible). In this case, a second stage is carried out, where a model of global precedence is proposed to account for mid-level properties of the HVS. The global precedence takes into account that contrasts of distortions should be proportioned across spatial frequency. The final metric is computed as the combination of perceived contrast of the distortion and disruption of global precedence. The VSNR is an interesting metric since it is based on contrast thresholds and does not modulate the entire image as the CSFs in, for example, S-CIELAB.

Spatial-DEE (S-DEE) developed in 2009 is the first metric using the Euclidean color difference formula in log-compressed OSA-UCS space (see Section 2), and it has been proposed by Simone et al. [32]. This metric is based on the modified S-CIELAB workflow developed by Johnson and Fairchild [33] in 2001. The metric is obtained by substituting $\Delta E_{ab}^*$ with $\Delta E_E$ in the last step of the workflow.

## 4 From contrast to image difference: WLF-DEE

The history of contrast is over a century long, and a lot of effort has been concerned on developing accurate contrast measures which are able to predict observer perceived contrast in natural images [34]. Furthermore, recent studies have shown that contrast is an important image quality attribute that falls under the umbrella of image quality [3].

In this work, we will propose a new image difference metric based on the work on contrast from Tadmor and Tolhurst [4] with two key features:

1. It uses a multi-level approach (or the so-called pyramidal image structure) as first suggested by Frankle and McCann in 1983 [35] and then Adelson et al. in 1984 [36] in order to perform a full investigation of all the frequencies in the image.
2. It uses the new Euclidean color difference formula in log-compressed OSA-UCS space proposed by Oleari et al. [5] and described in detail in Section 2.

In 2000, Tadmor and Tolhurst [4] based their analysis of contrast on the DOG model, which is modified and adapted to natural images. Since this model is used in this work, let us recall it completely:

In the conventional model, the spatial sensitivity in the center of receptive fields is described by a bi-dimensional Gaussian function with unit amplitude:

$$\text{Center}(x, y) = \exp\left[-\left(\frac{x}{r_c}\right)^2 - \left(\frac{y}{r_c}\right)^2\right], \quad (10)$$

where the radius $r_c$ represents the distance at which the sensitivity decreases to $1/e$ and $(x,y)$ are the spatial coordinates of the receptive field. The surround component is represented by another Gaussian curve, with a larger radius, $r_s$:

$$\text{Surround}(x,y) = \rho \left(\frac{r_c}{r_s}\right)^2 \exp\left[-\left(\frac{x}{r_s}\right)^2 - \left(\frac{y}{r_s}\right)^2\right], \tag{11}$$

where the scaling factor $\rho = 0.85$ sets the integrated sensitivity of the surround component to be 85% of that of the center. This scaling factor used by Tadmor and Tolhurst [4] is representative of the values reported for retinal ganglion cells and lateral geniculate nucleus (LGN) neurons of cat and monkey [37,38]. When the central point of the receptive field is placed at the location $(x,y)$, the output of the central component is calculated as

$$R_c(x,y) = \sum_{i=x-3r_c}^{i=x+3r_c} \sum_{j=y-3r_c}^{j=y+3r_c} \text{Center}(i-x, j-y) I(i,j), \tag{12}$$

while the output of the surround component is

$$R_s(x,y) = \sum_{i=x-3r_s}^{i=x+3r_s} \sum_{j=y-3r_s}^{j=y+3r_s} \text{Surround}(i-x, j-y) I(i,j), \tag{13}$$

where in both cases $I(i,j)$ is the image pixel value at position $(i,j)$.

The result of the DOG model is given by

$$\text{DOG}(x,y) = R_c(x,y) - R_s(x,y). \tag{14}$$

The conventional DOG model [39] assumes that the response of a neuron depends uniquely on the local luminance difference ($\Delta I$) between the center and the surround. After the light adaptation process, the gains of the ganglion cells of the retina and the neurons of the LGN depend on the average local luminance $I$. Thus, the model response depends on the contrast stimulus. After testing three different outputs for contrast
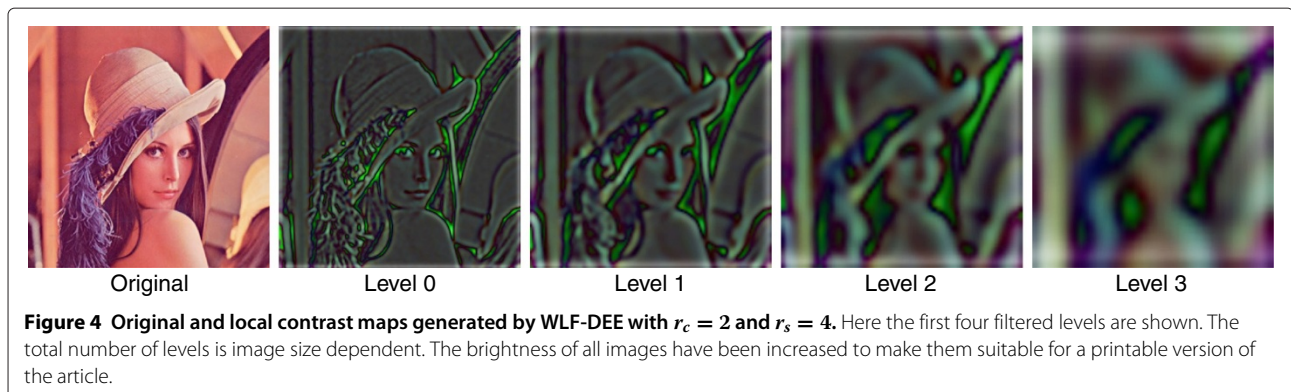
$$C^{TT}(x,y) = \frac{R_c(x,y) - R_s(x,y)}{R_c(x,y)} \text{(center-only scheme)}, \tag{15a}$$

$$C^{TT}(x,y) = \frac{R_c(x,y) - R_s(x,y)}{R_s(x,y)} \text{(surround-only scheme)}, \tag{15b}$$

$$C^{TT}(x,y) = \frac{R_c(x,y) - R_s(x,y)}{R_c(x,y) + R_s(x,y)} \text{(center-plus-surround scheme)}, \tag{15c}$$

they propose the criterion in Equation 15c for the contrast measuring, which is similar with the Michelson definition of grating contrast [40].

In this approach, a new metric is proposed and referred to as weighted-level framework $\Delta E_E$ (WLF-DEE). This metric is analogous to the S-CIELAB approach, but the original S-CIELAB spatial filtering is replaced with a multi-level DOG calculation, while $\Delta E_{ab}^*$ is replaced with the $\Delta E_E$ color difference formula. The general workflow of the metric is the following: The images are subsampled to various levels. The undersampling is simple since the images are reduced to half, and antialiasing filtering avoids artifacts at low resolutions. A pixelwise neighborhood contrast calculation is executed at each level by using the DOG on the three channels separately. Thus, local contrast maps for each level and each channel are obtained. An example is shown in Figure 4. Local contrast differences are computed by using the $\Delta E_E$, described in Section 2. A weighted recombination of the local contrast maps is computed and represented by a global image difference metric. Once local contrast maps are generated for each level, how to reduce the concept of contrast from local values at each pixel location to a single number representing the global image difference is still subject to debate. The simplest strategy is taking the mean of each level and averaging all together. This new metric performs a weighted recombination of the levels, given by the



**Figure 4 Original and local contrast maps generated by WLF-DEE with $r_c = 2$ and $r_s = 4$.** Here the first four filtered levels are shown. The total number of levels is image size dependent. The brightness of all images have been increased to make them suitable for a printable version of the article.

following equation, which leads to the final global image difference:

$$\text{WLF-DEE} = \frac{1}{N_l} \sum_{l=1}^{N_l} \lambda_l \cdot \overline{c}_l, \tag{16}$$

where $N_l$ is the number of levels, $\overline{c}_l$ is the mean contrast in the level $l$, and $\lambda_l$ is the weight assigned to each level $l$. So with the final measure, WLF-DEE aims at predicting the perceived magnitude of contrast between an original and a reproduction. This value can be used also as a quality indicator as contrast is one of the main quality attributes [3].

### 4.1 WLF-DEE characteristics
#### 4.1.1 Parameter tuning
As we can see from the previous Section, the DOG model is affected by the radius of the center component $r_c$, the radius of the surround component $r_s$ and the scaling factor of the surround component $\rho$, in addition to the three different schemes described in Equation 15. With the weighting parameters $\lambda$, WLF-DEE is subject to a total of four parameters. Since this metric comes out as a direct application of the research on perceptual contrast of Simone et al. [34] and performing the calculation of all possible values for these four parameters would be computationally exhausting, we have followed their rules of thumb for the choice. Table 1 reports our selection of tested configurations.

We have chosen four different configurations of $r_c$ and $r_s$, with the particular choice of $r_c = 2$ and $r_s = 4$, which

**Table 1 Selected values of parameters for WLF-DEE**

| Configuration | $r_c$ | $r_s$ | $\rho$ | $\lambda$ |
|---|---|---|---|---|
| A | 1 | 2 | 0.85 | Uniform |
| B | 2 | 3 | 0.85 | Uniform |
| C | 3 | 4 | 0.85 | Uniform |
| D | 2 | 4 | 0.85 | Uniform |
| E | 1 | 2 | 0.85 | Variance |
| F | 2 | 3 | 0.85 | Variance |
| G | 3 | 4 | 0.85 | Variance |
| H | 2 | 4 | 0.85 | Variance |
| I | 1 | 2 | 1.00 | Uniform |
| J | 2 | 3 | 1.00 | Uniform |
| K | 3 | 4 | 1.00 | Uniform |
| L | 2 | 4 | 1.00 | Uniform |
| M | 1 | 2 | 1.00 | Variance |
| N | 2 | 3 | 1.00 | Variance |
| O | 3 | 4 | 1.00 | Variance |
| P | 2 | 4 | 1.00 | Variance |

can be considered as the most effective in the identification of edges and blocks in relation to contrast [34]. Using larger radiuses would reduce the identification of local contrast. For the scaling factor of the surround component, in addition to the value $\rho = 0.85$ suggested by Tadmor and Tolhurst [4], we have chosen $\rho = 1.0$ justified by the fact that when the central and surround components are placed on a completely uniform area, the DOG model would give an output of contrast even though no contrast would be perceived by an observer. For the weighting level, we have chosen uniform weighting ($\lambda = 1$), which is the simplest strategy, and the variance of the level, which means that $\lambda$ assumes values taken from the image itself. Thus, $\lambda$ becomes level and image dependent. This strategy has shown benefits in correlation with observer perceived contrast [34]. For the three different schemes of the DOG model proposed, we have decided to test all of them. In conclusion, WLF-DEE will be tested for a total of 48 different configurations.

#### 4.1.2 Computational complexity
The computational complexity of WLF-DEE follows other state-of-the-art image difference metrics using multi-level approach, which is

$$\Theta\left(N \log N\right), \tag{17}$$

where $N$ is the number of pixels in the image.

## 5 Tests and results
### 5.1 Databases
Many different databases have been used for evaluating the image difference metrics, but in order to extensively test WLF-DEE, we have chosen three databases: the public Tampere Image Database 2008 (TID2008) and two databases developed at 'Norwegian Colour and Visual Computing Laboratory'.

The first database, the TID2008 database [24], contains a total of 1,700 images, with 25 reference images (Figure 5) and 17 types of distortions over four distortion levels. The mean opinion scores are the results of 654 observers attending the experiments. No viewing distance is stated in the TID database; therefore, we have used a standard viewing distance of 50 cm for the metrics requiring this setting. The authors have decided to include the images where distortions provide directly or indirectly a change in contrast, narrowing the TID2008 to a total of 400 images equally divided in the following four categories: masked noise, quantization noise, denoising and contrast change.

The second database, proposed by Pedersen et al. [41], contains four original images (Figure 6), three portraits and one illustration. The originals were altered in lightness, where each image had four versions with global lightness differences and four versions with local lightness changes. The lightness changes were 3 and 5 $\Delta E_{ab}^*$. Four

**Figure 5 TID2008 database [24].** The 25 reference images.

versions were brighter than the original, and four darker, for a total of 32 modified images. The psychophysical experiment was done on a calibrated CRT monitor, LaCie electron 22 blue II (LaCie, Basel, Switzerland), in a grey room . The observers were seated approximately 80 cm

from the screen. The light was measured to approximately 17 lux in front of the monitor. A total of 25 observers were recruited for the experiment, and they were asked in a pairwise comparison experiment to choose the image most similar to the original. This database is



**Figure 6 Pedersen database [41].** The four reference images changed globally and locally in lightness.

particularly of our interest because contrast is directly related to change in luminance [40], which is related to lightness [7,8].

The third database from Ajagamelle [42] contains a total of 10 original images covering a wide range of characteristics and scenes Figure 7. The images were modified using Adobe Photoshop software on a global scale with separate and simultaneous variations of contrast, lightness and saturation, resulting in a total number of 80 test images. The experiment was carried out as a category-judgment experiment with 14 observers. Each pair of images was displayed on an Eizo ColorEdge CG241W digital LCD display (Eizo Corporation, Ishikawa, Japan). The monitor was calibrated and profiled using Gretag-Macbeth Eye-One Match 3. The settings on the monitor were *sRGB* with a resolution of $1,600 \times 1,200$ pixels. The experiment took place in a windowless room with neutral grey walls, ceiling and floor. The ceiling lights in the room was set to provide a level of ambient illumination around 40 lux, which is below the upper threshold of 64 recommended by the CIE [43]. The white point was set to the D65, the gamma to 2.2 and the luminance level to 80 cd/m$^2$. The display was placed at a viewing distance of 70 cm. The images presented were $750 \times 499$ pixels or $499 \times 750$ pixels, which subtended roughly 20° of the visual angle when viewed at this distance.

### 5.2 Performance measures
Two types of correlation coefficients are computed [45] in order to evaluate the performance of WLF-DEE:

1. *The Pearson product-moment.* It assumes that the variables are ordinal, and it evaluates the linear relationship between two variables. This is a performance measure relating to the prediction accuracy of the metric [46].
2. *Spearman rank.* It is a non-parametric measure of correlation, and it is used as a measure of linear

relationship between two sets of ranked data, instead of the actual values. This describes the relationship between variables with no assumptions on the frequency distribution of the variables and on how tightly the ranked data clusters are around a straight line. This is a performance measure relating to the prediction monotonicity of the metric [46].

The relationships between the metrics and the observers are not necessarily linear. In order to remove any non-linearities due to the subjective experimental process and to facilitate comparison of the metrics in a common analysis space, we investigate the relationship between the metrics and observers by using non-linear regression [46]. In this work, we apply the same mapping function as that of Sheikh et al. [47]:

$$f(x) = \theta_1 \left( \frac{1}{2} - \frac{1}{1 + e^{\theta_2(x-\theta_3)}} \right) + \theta_4 X + \theta_5, \qquad (18)$$

where $\theta_i$, $i = 1, 2, 3, 4, 5$, are parameters to be be fitted. The 95% confidence intervals for the correlation values are calculated using Fisher's $Z$ transformation as described by the Video Quality Expert Group [48].

In order to have a complete analysis the following coefficients are also presented:

- *Root-mean-square error (RMSE)* [48]. It is a measure of the differences between the values predicted by the metric and the scores actually given by the observers.
- *Significance of the difference between the Pearson correlation coefficients (t-value)* [48]. This measure assumes that a good fit for observers' quality score is given by the normal distribution. It uses the $H_0$ hypothesis that assumes that there is no significant difference between correlation coefficients and the $H_1$ hypothesis, which considers that the difference is significant, although not specifying better or worse.



**Figure 7 Ajagamelle database [42].** Nine images were captured and provided by an independent photographer, and one image was selected from a standard natural image set provided by the CIE [44].

**Table 2 Pearson correlation for WLF-DEE and the selected state-of-the-art metrics on all databases**

| Metric | TID (masked noise) | | TID (quantization noise) | | TID (image denoising) | | TID (contrast change) | | Ajagamelle | | Pedersen | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Linear | Logistic | Linear | Logistic | Linear | Logistic | Linear | Logistic | Linear | Logistic | Linear | Logistic |
| PSNR | 0.57 | 0.59 | 0.50 | 0.51 | 0.55 | 0.55 | 0.47 | 0.49 | 0.73[a] | 0.80[a] | 0.66 | 0.66 |
| PSNR-HVS-M | 0.57 | 0.58 | 0.52 | 0.53 | 0.56 | 0.57 | 0.49 | 0.49 | 0.72 | 0.78 | 0.66 | 0.66 |
| S-CIELAB | 0.62 | 0.68 | 0.63 | 0.66 | 0.66 | 0.71 | 0.60 | 0.64 | 0.67 | 0.79 | 0.80[a] | 0.85[a] |
| SSIM | 0.71[a] | 0.75[a] | 0.83[a] | 0.85[a] | 0.79[a] | 0.81[a] | 0.73 | 0.76 | 0.64 | 0.64 | 0.22 | 0.22 |
| VSNR | 0.52 | 0.53 | 0.45 | 0.50 | 0.53 | 0.54 | 0.47 | 0.47 | 0.16 | 0.28 | 0.02 | 0.02 |
| WLF-DEE-C I | 0.60 | 0.64 | 0.72 | 0.75 | 0.62 | 0.62 | 0.65[b] | 0.69[b] | 0.39[b] | 0.65[b] | 0.27[b] | 0.49[b] |
| WLF-DEE-C J | 0.61[b] | 0.64 | 0.74 | 0.76 | 0.64 | 0.64 | 0.62 | 0.66 | 0.39 | 0.62 | 0.27[b] | 0.44 |
| WLF-DEE-C K | 0.61[b] | 0.68[b] | 0.75[b] | 0.77[b] | 0.66[b] | 0.71[b] | 0.63 | 0.67 | 0.38 | 0.60 | 0.27[b] | 0.44 |
| WLF-DEE-C L | 0.57 | 0.62 | 0.68 | 0.71 | 0.58 | 0.66 | 0.59 | 0.67 | 0.40 | 0.64 | 0.27[b] | 0.45 |
| WLF-DEE-C M | 0.35 | 0.38 | 0.25 | 0.55 | 0.22 | 0.22 | 0.25 | 0.53 | 0.21 | 0.26 | 0.16 | 0.17 |
| WLF-DEE-C N | 0.38 | 0.53 | 0.33 | 0.58 | 0.23 | 0.60 | 0.29 | 0.55 | 0.20 | 0.20 | 0.16 | 0.19 |
| WLF-DEE-C O | 0.39 | 0.56 | 0.38 | 0.62 | 0.24 | 0.62 | 0.34 | 0.57 | 0.20 | 0.34 | 0.17 | 0.40 |
| WLF-DEE-C P | 0.32 | 0.53 | 0.26 | 0.57 | 0.19 | 0.60 | 0.24 | 0.54 | 0.21 | 0.31 | 0.17 | 0.47 |

[a]The most performant metric. [b]The highest performance of WLF-DEE.

**Table 3 Spearman correlation for WLF-DEE and the selected state-of-the-art metrics on all databases**

| Metric | TID (masked noise) | TID (quantization noise) | TID (image denoising) | TID (contrast change) | Ajagamelle | Pedersen |
|---|---|---|---|---|---|---|
| | Linear | Linear | Linear | Linear | Linear | Linear |
| PSNR | 0.55 | 0.47 | 0.53 | 0.45 | 0.71 | 0.72 |
| PSNR-HVS-M | 0.53 | 0.46 | 0.53 | 0.45 | 0.70 | 0.72 |
| S-CIELAB | 0.66 | 0.63 | 0.69 | 0.62 | 0.72 | 0.77[a] |
| SSIM | 0.74[a] | 0.83[a] | 0.78[a] | 0.74[a] | 0.72[a] | 0.49 |
| VSNR | 0.48 | 0.39 | 0.49 | 0.43 | 0.14 | 0.02 |
| WLF-DEE-C I | 0.60[b] | 0.71[b] | 0.66[b] | 0.63[b] | 0.61[b] | 0.43 |
| WLF-DEE-C J | 0.59 | 0.69 | 0.65 | 0.59 | 0.59 | 0.42 |
| WLF-DEE-C K | 0.59 | 0.70 | 0.66[b] | 0.60 | 0.57 | 0.44[b] |
| WLF-DEE-C L | 0.57 | 0.65 | 0.62 | 0.59 | 0.60 | 0.43 |
| WLF-DEE-C M | 0.43 | 0.48 | 0.50 | 0.46 | 0.56 | 0.37 |
| WLF-DEE-C N | 0.50 | 0.56 | 0.56 | 0.48 | 0.55 | 0.38 |
| WLF-DEE-C O | 0.53 | 0.59 | 0.58 | 0.48 | 0.53 | 0.38 |
| WLF-DEE-C P | 0.48 | 0.51 | 0.53 | 0.47 | 0.56 | 0.37 |

[a]The most performant metric. [b]The highest performance of WLF-DEE.

**Table 4 RMSE for WLF-DEE and the selected state-of-the-art metrics on all databases**

| Metric | TID (masked noise) | TID (quantization noise) | TID (image denoising) | TID (contrast change) | Ajagamelle | Pedersen |
|---|---|---|---|---|---|---|
| PSNR | 14.73 | 15.38 | 15.72 | 15.91 | 1.49 | 0.40 |
| PSNR-HVS-M | 14.84 | 15.15 | 15.54 | 15.85 | 1.54 | 0.40 |
| S-CIELAB | 13.35 | 13.41 | 13.29 | 13.94 | 0.88 | 0.28[a] |
| SSIM | 11.99[a] | 9.45[a] | 10.99[a] | 11.80[a] | 1.75 | 0.53 |
| VSNR | 15.42 | 15.41 | 15.92 | 16.07 | 1.03 | 0.53 |
| WLF-DEE-C I | 13.99 | 11.88 | 14.79 | 13.11[b] | 1.01 | 0.46 |
| WLF-DEE-C J | 13.98 | 11.54 | 14.51 | 13.73 | 1.01 | 0.47 |
| WLF-DEE-C K | 13.33[b] | 11.36[b] | 13.33[b] | 13.53 | 1.01 | 0.48 |
| WLF-DEE-C L | 14.36 | 12.54 | 14.22 | 13.56 | 1.02 | 0.47 |
| WLF-DEE-C M | 16.88 | 14.93 | 18.39 | 15.49 | 0.94 | 0.52 |
| WLF-DEE-C N | 15.48 | 14.55 | 15.12 | 15.19 | 0.78[a,b] | 0.52 |
| WLF-DEE-C O | 15.12 | 14.03 | 14.74 | 14.98 | 1.06 | 0.49 |
| WLF-DEE-C P | 15.48 | 14.64 | 15.15 | 15.29 | 1.43 | 0.47 |

[a]The most performant metric. [b]The highest performance of WLF-DEE.

**Table 5 Significance of the difference between Pearson correlation coefficients on TID-masked noise database**

| Metric | PSNR | PSNR-HVS-M | S-CIELAB | SSIM | VSNR | WLF-DEE-C I | WLF-DEE-C J | WLF-DEE-C K | WLF-DEE-C L | WLF-DEE-C M | WLF-DEE-C N | WLF-DEE-C O | WLF-DEE-C P |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| PSNR | - | | | | | | | | | | | | |
| PSNR-HVS-M | 0.08 | - | | | | | | | | | | | |
| S-CIELAB | −1.08 | −1.16 | - | | | | | | | | | | |
| SSIM | −2.13[a] | −2.21[a] | −1.04 | - | | | | | | | | | |
| VSNR | 0.57 | 0.49 | 1.65 | 2.69[a] | - | | | | | | | | |
| WLF-DEE-C I | −0.59 | −0.67 | 0.50 | 1.54 | −1.16 | - | | | | | | | |
| WLF-DEE-C J | −0.60 | −0.68 | 0.48 | 1.53 | −1.17 | −0.01 | - | | | | | | |
| WLF-DEE-C K | −1.10 | −1.19 | −0.02 | 1.02 | −1.67 | −0.52 | −0.50 | - | | | | | |
| WLF-DEE-C L | −0.30 | −0.38 | 0.78 | 1.82 | −0.87 | 0.28 | 0.30 | 0.80 | - | | | | |
| WLF-DEE-C M | 1.94 | 1.86 | 3.02[a] | 4.07[a] | 1.37 | 2.53[a] | 2.54[a] | 3.05[a] | 2.24[a] | - | | | |
| WLF-DEE-C N | 0.62 | 0.54 | 1.70 | 2.75[a] | 0.05 | 1.21 | 1.22 | 1.72 | 0.92 | −1.32 | - | | |
| WLF-DEE-C O | 0.32 | 0.23 | 1.40 | 2.44[a] | −0.25 | 0.90 | 0.92 | 1.42 | 0.62 | −1.63 | −0.30 | - | |
| WLF-DEE-C P | 0.62 | 0.53 | 1.70 | 2.74[a] | 0.05 | 1.20 | 1.21 | 1.72 | 0.92 | −1.33 | 0.00 | 0.30 | - |

[a]Significantly different at 5% significance level (1.96).

**Table 6 Significance of the difference between Pearson correlation coefficients on TID-quantization noise database**

| Metric | PSNR | PSNR-HVS-M | S-CIELAB | SSIM | VSNR | WLF-DEE-C I | WLF-DEE-C J | WLF-DEE-C K | WLF-DEE-C L | WLF-DEE-C M | WLF-DEE-C N | WLF-DEE-C O | WLF-DEE-C P |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| PSNR | - | | | | | | | | | | | | |
| PSNR-HVS-M | −0.20 | - | | | | | | | | | | | |
| S-CIELAB | −1.63 | −1.43 | - | | | | | | | | | | |
| SSIM | −4.81[a] | −4.62[a] | −3.19[a] | - | | | | | | | | | |
| VSNR | 0.03 | 0.22 | 1.66 | 4.84[a] | - | | | | | | | | |
| WLF-DEE-C I | −2.82[a] | −2.63[a] | −1.19 | 1.99[a] | −2.85[a] | - | | | | | | | |
| WLF-DEE-C J | −3.10[a] | −2.90[a] | −1.47 | 1.72 | −3.12[a] | −0.27 | - | | | | | | |
| WLF-DEE-C K | −3.23[a] | −3.04[a] | −1.61 | 1.58 | −3.26[a] | −0.41 | −0.14 | - | | | | | |
| WLF-DEE-C L | −2.30[a] | −2.11[a] | −0.67 | 2.51[a] | −2.33[a] | 0.52 | 0.79 | 0.93 | - | | | | |
| WLF-DEE-C M | −0.39 | −0.19 | 1.24 | 4.43[a] | −0.42 | 2.43[a] | 2.71[a] | 2.85[a] | 1.91 | - | | | |
| WLF-DEE-C N | −0.71 | −0.51 | 0.92 | 4.11[a] | −0.73 | 2.12[a] | 2.39[a] | 2.53[a] | 1.60 | −0.32 | - | | |
| WLF-DEE-C O | −1.13 | −0.94 | 0.49 | 3.68[a] | −1.16 | 1.69 | 1.96 | 2.10[a] | 1.17 | −0.75 | −0.43 | - | |
| WLF-DEE-C P | −0.64 | −0.44 | 0.99 | 4.18[a] | −0.66 | 2.19[a] | 2.46[a] | 2.60[a] | 1.67 | −0.25 | 0.07 | 0.50 | - |

[a] Significantly different at 5% significance level (1.96).

**Table 7 Significance of the difference between Pearson correlation coefficients on TID-image denoising database**

| Metric | PSNR | PSNR-HVS-M | S-CIELAB | SSIM | VSNR | WLF-DEE-C I | WLF-DEE-C J | WLF-DEE-C K | WLF-DEE-C L | WLF-DEE-C M | WLF-DEE-C N | WLF-DEE-C O | WLF-DEE-C P |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| PSNR | - | | | | | | | | | | | | |
| PSNR-HVS-M | −0.14 | - | | | | | | | | | | | |
| S-CIELAB | −1.84 | −1.70 | - | | | | | | | | | | |
| SSIM | −3.57[a] | −3.43[a] | −1.73 | - | | | | | | | | | |
| VSNR | 0.16 | 0.30 | 2.00[a] | 3.73[a] | - | | | | | | | | |
| WLF-DEE-C I | −0.73 | −0.58 | 1.12 | 2.85[a] | −0.89 | - | | | | | | | |
| WLF-DEE-C J | −0.94 | −0.79 | 0.91 | 2.64[a] | −1.10 | −0.21 | - | | | | | | |
| WLF-DEE-C K | −1.81 | −1.66 | 0.04 | 1.76 | −1.97 | −1.08 | −0.87 | - | | | | | |
| WLF-DEE-C L | −1.15 | −1.01 | 0.69 | 2.42[a] | −1.31 | −0.43 | −0.22 | 0.66 | - | | | | |
| WLF-DEE-C M | 2.77[a] | 2.91[a] | 4.61[a] | 6.34[a] | 2.61[a] | 3.49[a] | 3.70[a] | 4.57[a] | 3.92[a] | - | | | |
| WLF-DEE-C N | −0.47 | −0.33 | 1.37 | 3.10[a] | −0.63 | 0.25 | 0.46 | 1.34 | 0.68 | −3.24[a] | - | | |
| WLF-DEE-C O | −0.76 | −0.62 | 1.08 | 2.81[a] | −0.92 | −0.04 | 0.17 | 1.05 | 0.39 | −3.53[a] | −0.29 | - | |
| WLF-DEE-C P | −0.45 | −0.30 | 1.39 | 3.12[a] | −0.61 | 0.28 | 0.49 | 1.36 | 0.70 | −3.21[a] | 0.02 | 0.31 | - |

[a]Significantly different at 5% significance level (1.96).

**Table 8 Significance of the difference between Pearson correlation coefficients on TID-contrast change database**

| Metric | PSNR | PSNR-HVS-M | S-CIELAB | SSIM | VSNR | WLF-DEE-C I | WLF-DEE-C J | WLF-DEE-C K | WLF-DEE-C L | WLF-DEE-C M | WLF-DEE-C N | WLF-DEE-C O | WLF-DEE-C P |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| PSNR | - | | | | | | | | | | | | |
| PSNR-HVS-M | −0.06 | - | | | | | | | | | | | |
| S-CIELAB | −1.62 | −1.56 | - | | | | | | | | | | |
| SSIM | −3.27[a] | −3.21[a] | −1.65 | - | | | | | | | | | |
| VSNR | 0.15 | 0.20 | 1.77 | 3.41[a] | - | | | | | | | | |
| WLF-DEE-C I | −2.26[a] | −2.21[a] | −0.64 | 1.01 | −2.41[a] | - | | | | | | | |
| WLF-DEE-C J | −1.78 | −1.73 | −0.16 | 1.49 | −1.93 | 0.48 | - | | | | | | |
| WLF-DEE-C K | −1.94 | −1.88 | −0.32 | 1.33 | −2.08[a] | 0.32 | −0.16 | - | | | | | |
| WLF-DEE-C L | −1.92 | −1.86 | −0.30 | 1.35 | −2.06[a] | 0.34 | −0.14 | 0.02 | - | | | | |
| WLF-DEE-C M | −0.37 | −0.31 | 1.25 | 2.90[a] | −0.51 | 1.89 | 1.41 | 1.57 | 1.55 | - | | | |
| WLF-DEE-C N | −0.62 | −0.57 | 1.00 | 2.65[a] | −0.77 | 1.64 | 1.16 | 1.32 | 1.30 | −0.25 | - | | |
| WLF-DEE-C O | −0.79 | −0.74 | 0.83 | 2.48[a] | −0.94 | 1.47 | 0.99 | 1.15 | 1.13 | −0.42 | −0.17 | - | |
| WLF-DEE-C P | −0.54 | −0.48 | 1.08 | 2.73[a] | −0.68 | 1.72 | 1.24 | 1.40 | 1.38 | −0.17 | 0.08 | 0.25 | - |

[a]Significantly different at 5% significance level (1.96).

**Table 9 Significance of the difference between Pearson correlation coefficients on Ajagamelle database**

| Metric | PSNR | PSNR-HVS-M | S-CIELAB | SSIM | VSNR | WLF-DEE-C I | WLF-DEE-C J | WLF-DEE-C K | WLF-DEE-C L | WLF-DEE-C M | WLF-DEE-C N | WLF-DEE-C O | WLF-DEE-C P |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| PSNR | - | | | | | | | | | | | | |
| PSNR-HVS-M | 0.29 | - | | | | | | | | | | | |
| S-CIELAB | 0.09 | −0.20 | - | | | | | | | | | | |
| SSIM | 2.08[a] | 1.79 | 1.99[a] | - | | | | | | | | | |
| VSNR | 4.96[a] | 4.67[a] | 4.87[a] | 2.88[a] | - | | | | | | | | |
| WLF-DEE-C I | 1.96 | 1.67 | 1.87 | −0.12 | −3.00[a] | - | | | | | | | |
| WLF-DEE-C J | 2.28[a] | 1.99[a] | 2.19[a] | 0.20 | −2.68[a] | 0.32 | - | | | | | | |
| WLF-DEE-C K | 2.47[a] | 2.18[a] | 2.38[a] | 0.39 | −2.49[a] | 0.51 | 0.20 | - | | | | | |
| WLF-DEE-C L | 2.01[a] | 1.73 | 1.93 | −0.07 | −2.95[a] | 0.06 | −0.26 | −0.46 | - | | | | |
| WLF-DEE-C M | 5.11[a] | 4.83[a] | 5.02[a] | 3.03[a] | 0.15 | 3.16[a] | 2.84[a] | 2.64[a] | 3.10[a] | - | | | |
| WLF-DEE-C N | 5.46[a] | 5.17[a] | 5.37[a] | 3.38[a] | 0.50 | 3.50[a] | 3.18[a] | 2.99[a] | 3.44[a] | 0.35 | - | | |
| WLF-DEE-C O | 4.51[a] | 4.23[a] | 4.42[a] | 2.43[a] | −0.45 | 2.56[a] | 2.24[a] | 2.04[a] | 2.50[a] | −0.60 | −0.95 | - | |
| WLF-DEE-C P | 4.76[a] | 4.48[a] | 4.67[a] | 2.68[a] | −0.20 | 2.81[a] | 2.49[a] | 2.29[a] | 2.75[a] | −0.35 | −0.70 | 0.25 | - |

[a]Significantly different at 5% significance level (1.96).

**Table 10 Significance of the difference between Pearson correlation coefficients on Pedersen database**

| Metric | PSNR | PSNR-HVS-M | S-CIELAB | SSIM | VSNR | WLF-DEE-C I | WLF-DEE-C J | WLF-DEE-C K | WLF-DEE-C L | WLF-DEE-C M | WLF-DEE-C N | WLF-DEE-C O | WLF-DEE-C P |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| PSNR | - | | | | | | | | | | | | |
| PSNR-HVS-M | −0.03 | - | | | | | | | | | | | |
| S-CIELAB | −1.83 | −1.80 | - | | | | | | | | | | |
| SSIM | 2.15[a] | 2.18[a] | 3.98[a] | - | | | | | | | | | |
| VSNR | 2.90[a] | 2.93[a] | 4.73[a] | 0.75 | - | | | | | | | | |
| WLF-DEE-C I | 0.93 | 0.96 | 2.76[a] | −1.22 | −1.97[a] | - | | | | | | | |
| WLF-DEE-C J | 1.17 | 1.20 | 3.00[a] | −0.98 | −1.73 | 0.24 | - | | | | | | |
| WLF-DEE-C K | 1.18 | 1.21 | 3.01[a] | −0.97 | −1.72 | 0.25 | 0.01 | - | | | | | |
| WLF-DEE-C L | 1.16 | 1.19 | 2.99[a] | −0.99 | −1.74 | 0.23 | −0.01 | −0.02 | - | | | | |
| WLF-DEE-C M | 2.33[a] | 2.35[a] | 4.15[a] | 0.18 | −0.58 | 1.39 | 1.16 | 1.14 | 1.17 | - | | | |
| WLF-DEE-C N | 2.27[a] | 2.30[a] | 4.10[a] | 0.12 | −0.63 | 1.34 | 1.10 | 1.09 | 1.11 | −0.06 | - | | |
| WLF-DEE-C O | 1.39 | 1.42 | 3.22[a] | −0.76 | −1.51 | 0.46 | 0.22 | 0.21 | 0.23 | −0.93 | −0.87 | - | |
| WLF-DEE-C P | 1.06 | 1.09 | 2.89[a] | −1.09 | −1.84 | 0.13 | −0.11 | −0.12 | −0.10 | −1.26 | −1.21 | −0.33 | - |

[a] Significantly different at 5% significance level (1.96).

## 5.3  Results

As mentioned in Section 4.1.1, WLF-DEE has been tested for a total of 48 configurations, but in order to give a more readable and understandable presentation of the results, we will present only a selection of them. As WLF-DEE using the two DOG schemes in Equations 15a and 15b have lower performance in correlation with respect to WLF-DEE using the DOG scheme in Equation 15(c), these results will be excluded. This confirms also the statement of Tadmor and Tolhurst that the DOG model in analogy with the Michelson formula has better performance for contrast assessment [4]. On the same way, all configurations with $\rho = 0.85$ in Equation 13 will be excluded, as they show lower performance in correlation with respect to those configurations with $\rho = 1.00$. This will end in a presentation of a total of only eight results of WLF-DEE shown in Table 2 for Pearson correlation, Table 3 for Spearman correlation and Table 4 for RMSE. Significance of the difference between the Pearson correlation coefficients are presented for each database in Tables 5, 6, 7, 8, 9 and 10.

Considering the TID database, SSIM has the higher Pearson correlation in all four categories. In the masked noise category, SSIM is followed by WLF-DEE K, while in the quantization noise, it is followed by S-CIELAB and then WLF-DEE K. In the category denoising and contrast change, SSIM instead is followed by WLF-DEE I and then WLF-DEE K. For all the four categories, it is possible to notice that all the metrics give higher correlation with perceived observer difference using the logistic fitting. Furthermore, as the confidence intervals (Figures 8, 9, 10 and 11) of WLF-DEE K overlap with the confidence intervals of SSIM, we can claim to have the same performance. Overall, for the four categories of the TID database, WLF-DEE K shows to be significantly better than PSNR and VSNR and to have the same performance with SSIM and S-CIELAB.

For the Ajagamelle database, PSNR shows the higher Pearson correlation, followed by PSNR-HVS-M, SCI-ELAB, SSIM and then four configurations (I, J, K, L) of WLF-DEE, which have very close results. Since in this case, the confidence intervals (Figure 12) of WLF-DEE K overlaps with the confidence interval of those metrics with slightly higher correlation, we can claim to have the same performance. For the Pedersen database, S-CIELAB shows the higher Pearson correlation, followed by PSNR-HVS-M, PSNR, four configurations (I, J, K, L) of WLF-DEE with very close results and then SSIM. In this database instead, confidence intervals (Figure 13) show that WLF-DEE-K has a slightly lower performance than S-CIELAB, but not with respect to SSIM. Also for these two databases, it holds true

that all the metrics give higher correlation using logistic fitting.

Considering all the six database sets examined, WLF-DEE gives higher correlation using configurations I, J, K, L with respect to configurations M, N, O, P, and in particular, WLF-DEE K most agrees with observer perceived difference, indicating that large radiuses of the Gaussians and uniform weighting of the levels should be used for the estimation of perceived difference. Furthermore, it is possible to notice that WLF-DEE K with logistic fitting shows a stable trend among the six datasets having an average performance in correlation of 0.65. This holds true also for other tested metrics such as S-CIELAB and PSNR-HVS-M, but not for SSIM and VSNR which show very high correlation in one dataset and very low in an another one.

Analysis with the Spearman correlation follows the same discussion with the Pearson correlation except for the Ajagamelle database, where the highest correlation is shown by S-CIELAB, but not outperforming most of the other metrics. The results are presented only with linear fitting as no improvements can be found in any metric using the logistic fitting. Also with the Spearman correlation, WLF-DEE K shows its stability with an average performance of 0.59.

Analysis with root-mean-square error shows that for all the four categories of the TID database, SSIM has the lowest RMSE. As the confidence intervals of SSIM overlap with two configurations of WLF-DEE-C (J, K) (Figures 14, 15, 16 and 17), it cannot be claimed that the two metrics are significantly different in performance.

For Ajagamelle database instead, WLF-DEE-C N shows the lowest RMSE, followed by S-CIELAB and then by other several configurations of WLF-DEE-C (M, I, J, K, L) and VSNR. Confidence intervals (Figure 18) shows that these three metrics are not significantly different among each other but they outperform other tested metrics such as SSIM, PSNR and PSNR-HVS-M. For Pedersen database, S-CIELAB shows the lowest RMSE followed by PSNR, PSNR-HVS-M and then all the configurations of WLF-DEE-C. SSIM and VSNR have the highest RMSE. Confidence intervals (Figure 19) show that WLF-DEE-C (I, J, K, L, P) has no difference in performance with the other tested metrics but the overlap with S-CIELAB confidence interval is minimal.

Analysis with significance of the difference are presented with 5% significance level for Pearson correlation with logistic fitting only. Based on the definition in [48], two metrics can be significantly different if $-1.96 < t - value < 1.96$. This analysis confirms that WLF-DEE-C K is not significantly different in performance with respect to the other tested metrics for TID-masked noise and TID-image denoising databases. For TID-quantization noise instead, WLF-DEE-C K is
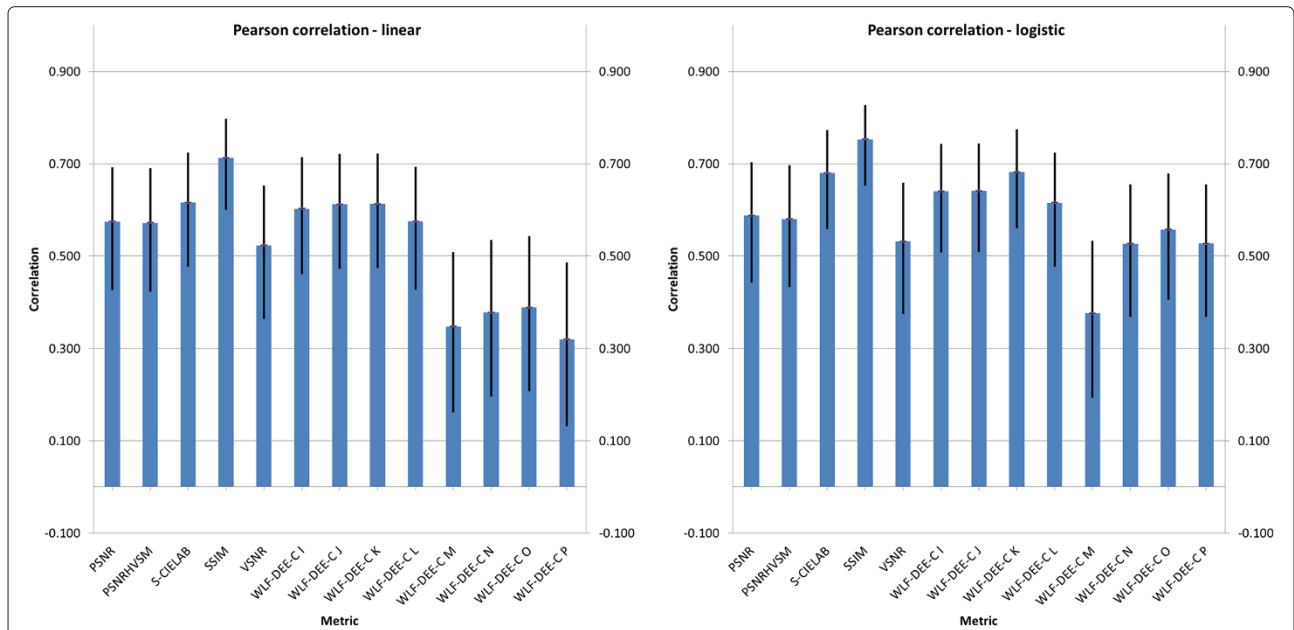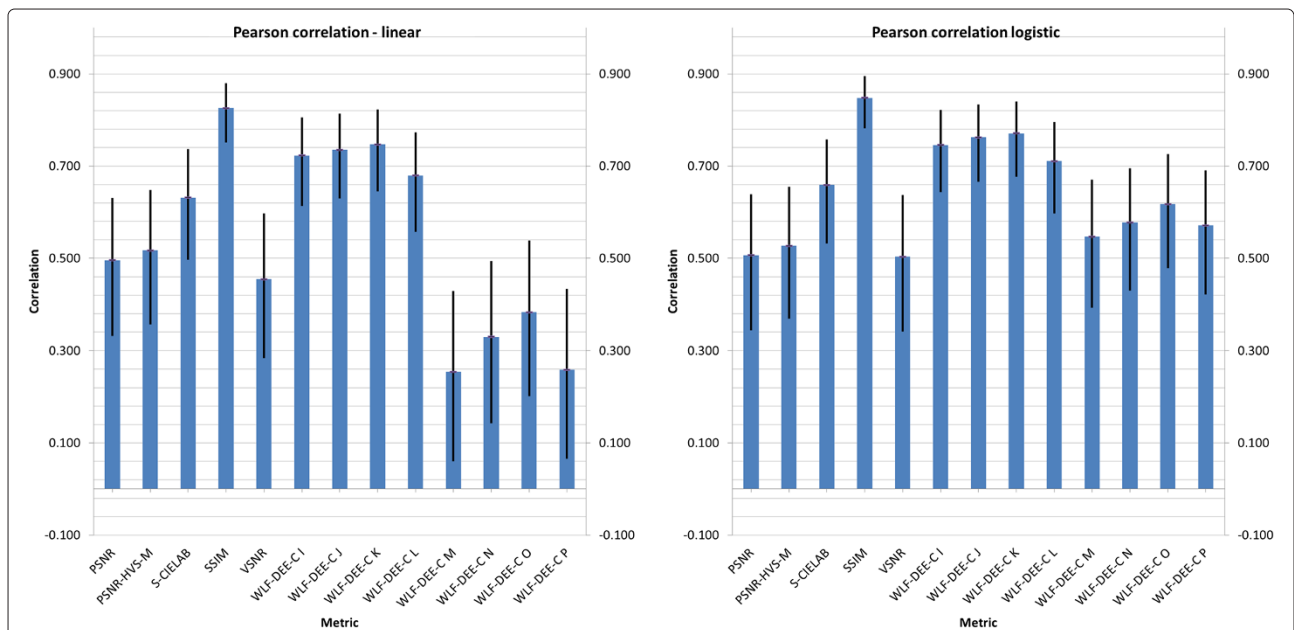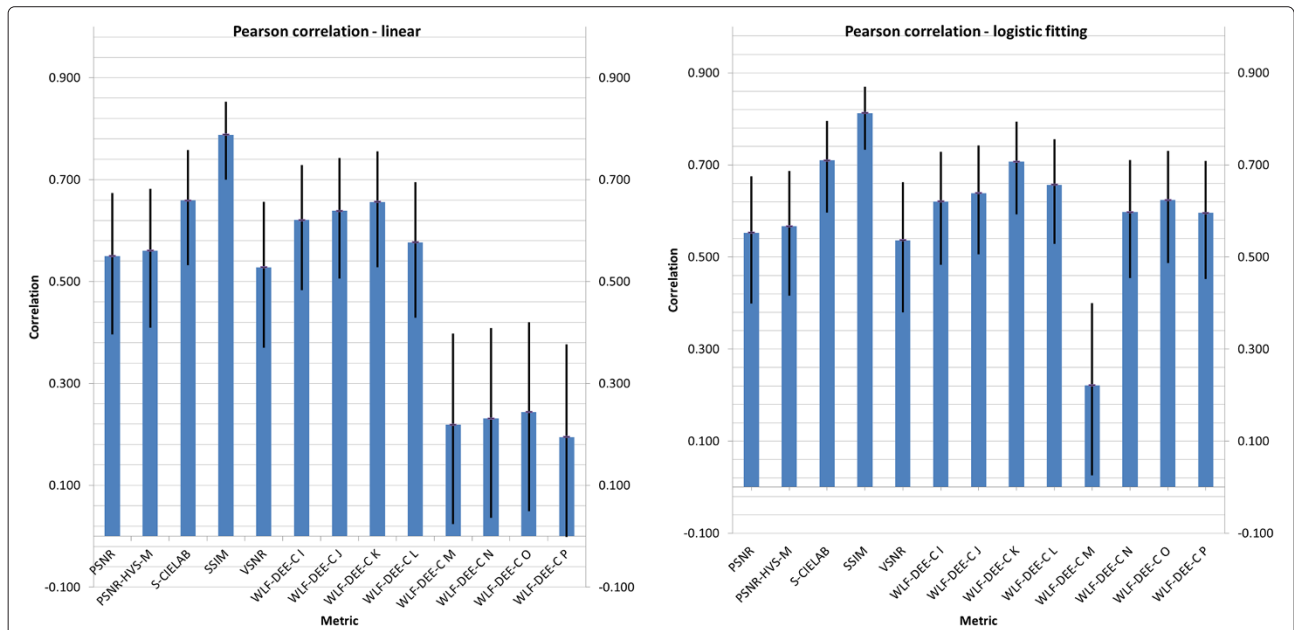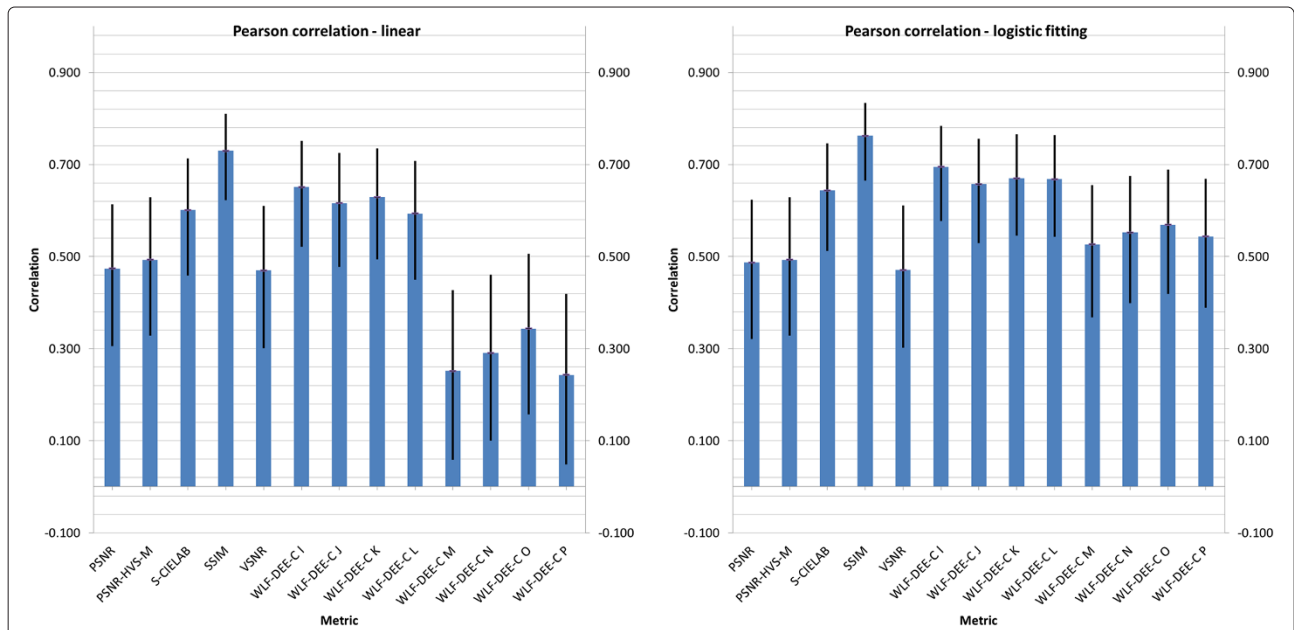
**Figure 8 Pearson correlation for WLF-DEE and the selected state-of-the-art metrics on TID-masked noise database.** Results using linear fitting on the left and logistic fitting on the right.

not significantly difference in performance with respect to SSIM and S-CIELAB. For TID-contrast change WLF-DEE-C K is significantly difference in performance only with respect to VSNR. For Ajagamelle database, WLF-DEE-C K is not significantly different in performance only

from SSIM while for Pedersen database is not significantly different in performance from the other tested metrics except S-CIELAB.

Overall, WLF-DEE-C K shows its particular strength on those databases where a change in contrast between the



**Figure 9 Pearson correlation for WLF-DEE and the selected state-of-the-art metrics on TID-quantization noise database.** Results using linear fitting on the left and logistic fitting on the right.

**Figure 10 Pearson correlation for WLF-DEE and the selected state-of-the-art metrics on TID-image denoising database.** Results using linear fitting on the left and logistic fitting on the right.

original and its reproduction is triggered by a change of color attributes and not particular distortions. In conclusion, WLF-DEE-C K promotes itself as a new metric for predicting the perceived magnitude of contrast between an original and a reproduction, fulfilling the purpose for which it was developed.

## 6 Conclusions

Recent studies have shown that contrast is an important image attribute that falls under the umbrella of image quality [3]. In this paper, we have developed a new metric based on recent work on contrast filtering. This metric called WLF-DEE consists of two key features: a multi-



**Figure 11 Pearson correlation for WLF-DEE and the selected state-of-the-art metrics on TID-contrast change database.** Results using linear fitting on the left and logistic fitting on the right.

**Figure 12 Pearson correlation for WLF-DEE and the selected state-of-the-art metrics on Ajagamelle database.** Results using linear fitting on the left and logistic fitting on the right.
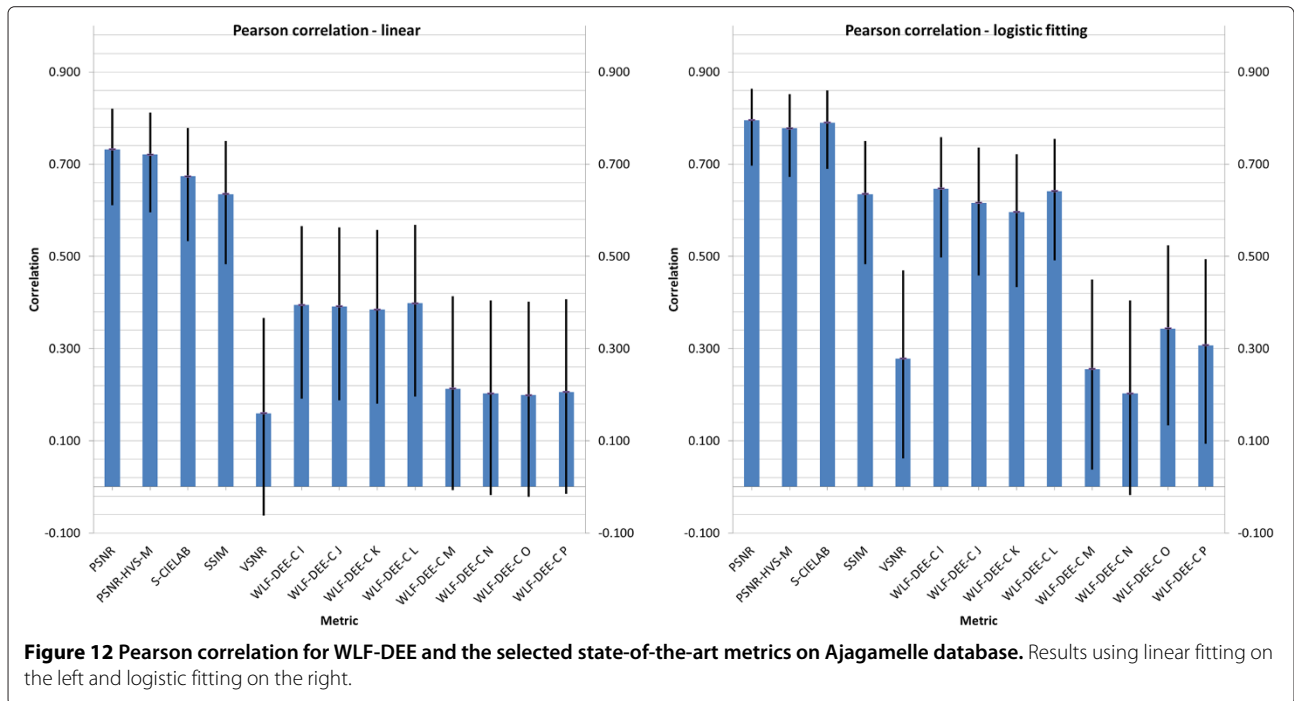
level filtering based on the work on contrast from Tadmor and Tolhurst [4] and the new Euclidean color difference formula in log-compressed OSA-UCS space proposed by Oleari et al. [5].

Extensive tests and analysis are carried out on four different categories of the well-known TID database and on two databases providing different distortions directly related to color and contrast. Comparisons with other state-of-the-art metrics are also presented. Results promote WLF-DEE as a new efficient metric for estimating the perceived magnitude of contrast between an original and a reproduction.



**Figure 13 Pearson correlation for WLF-DEE and the selected state-of-the-art metrics on Pedersen database.** Results using linear fitting on the left and logistic fitting on the right.
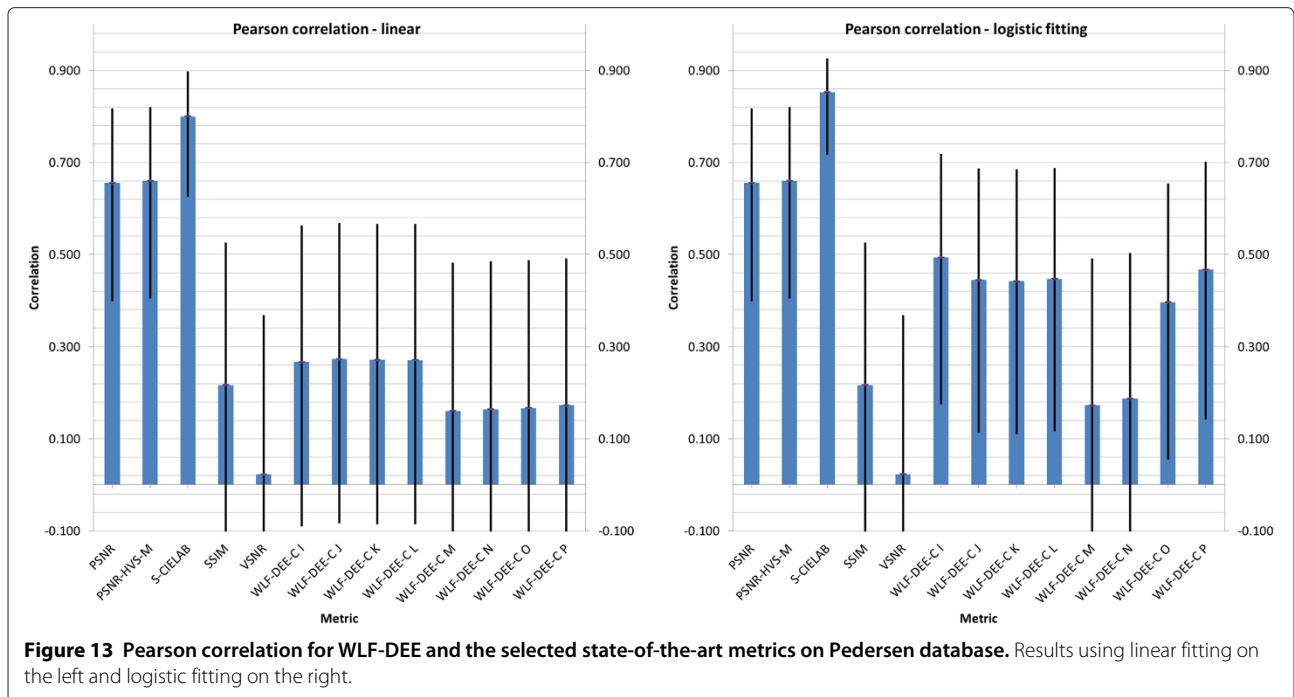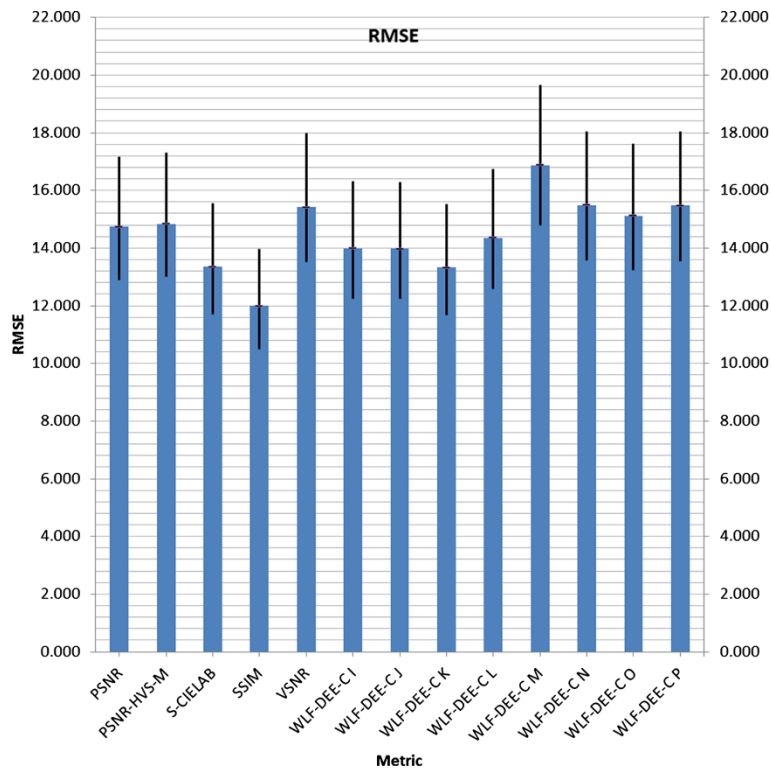
**Figure 14 RMSE for WLF-DEE and the selected state-of-the-art metrics on TID-masked noise database.**
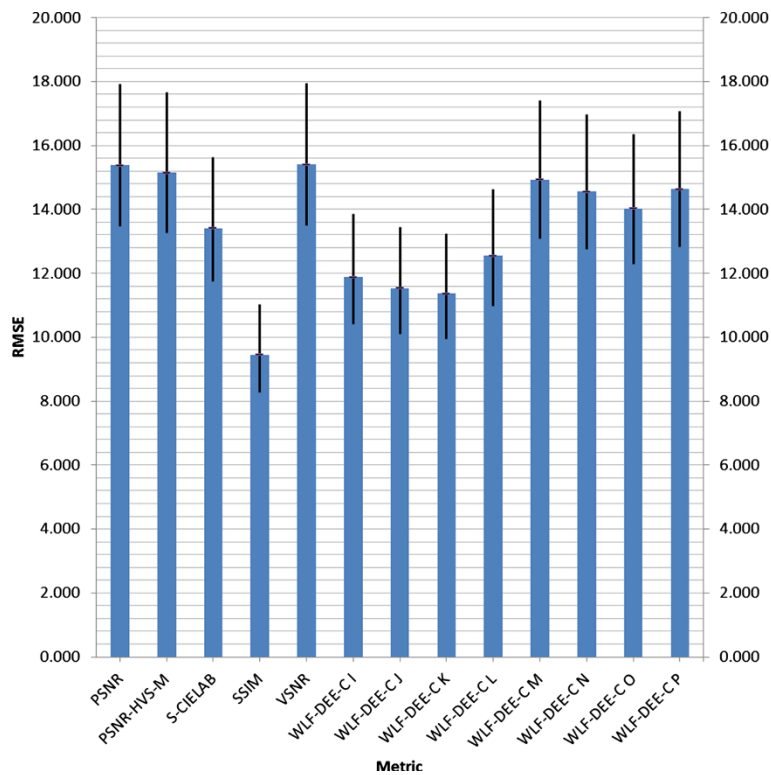


**Figure 15 RMSE for WLF-DEE and the selected state-of-the-art metrics on TID-quantization noise database.**
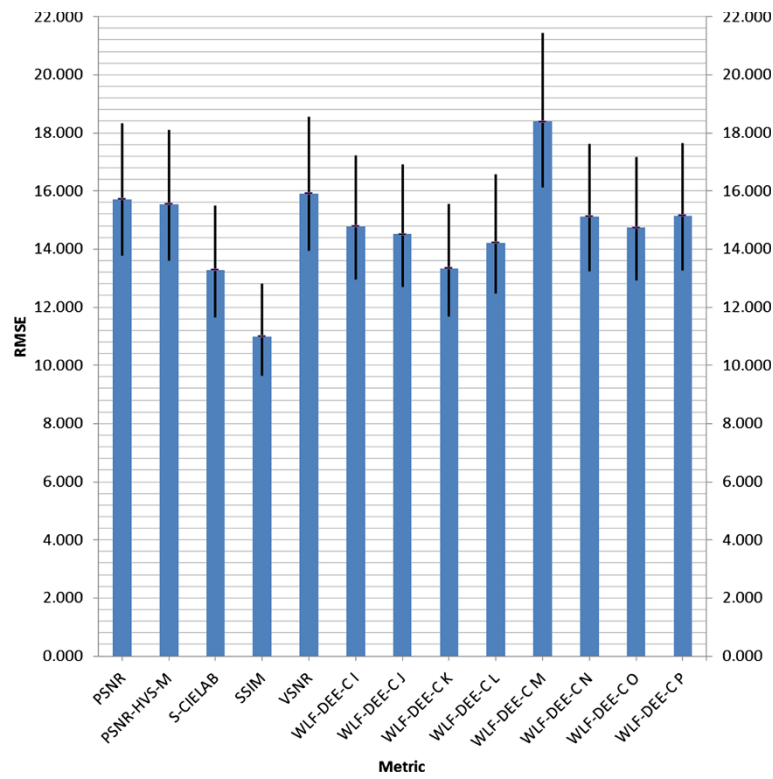
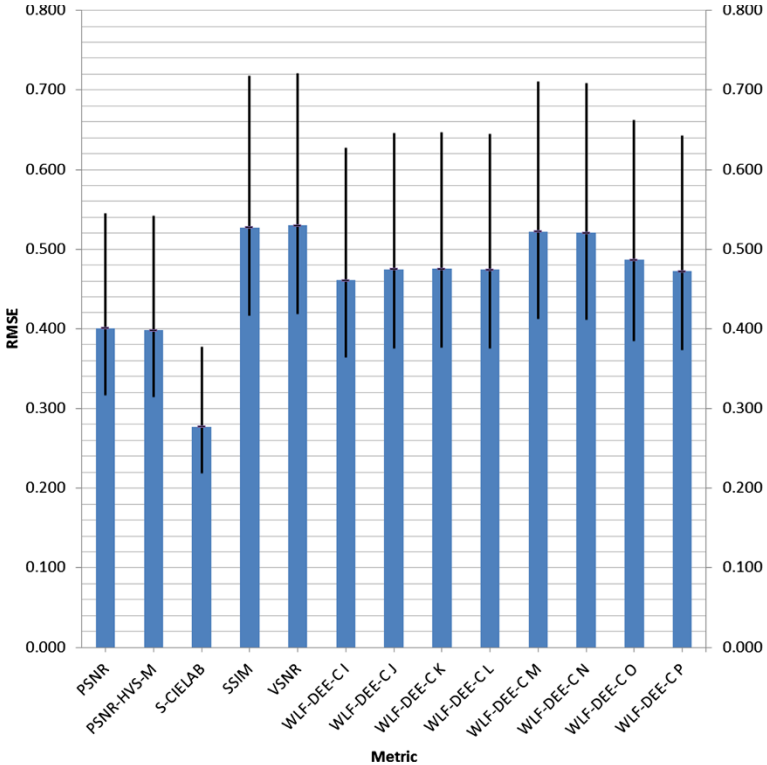**Figure 16 RMSE for WLF-DEE and the selected state-of-the-art metrics on TID-image denoising database.**
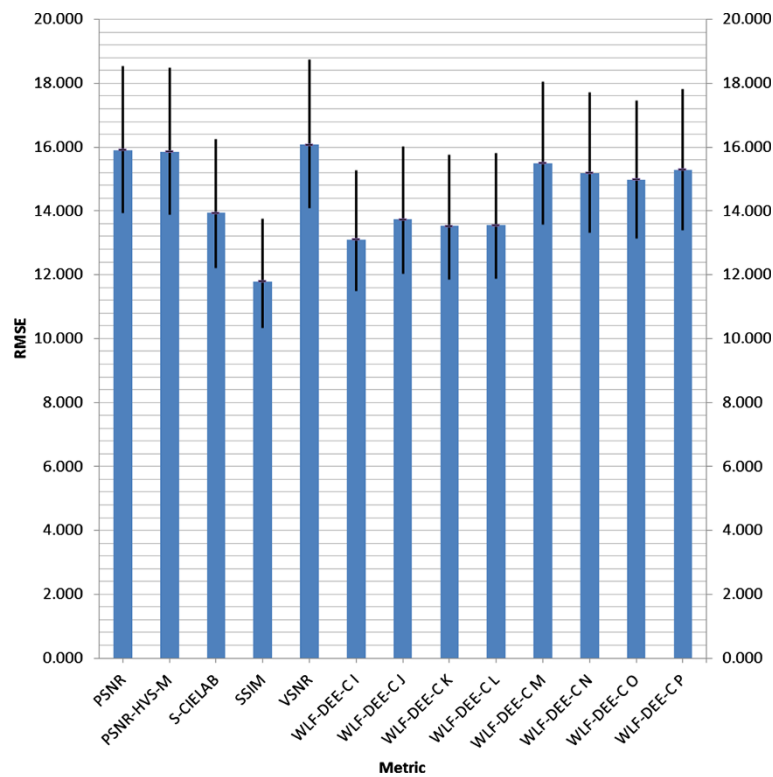


**Figure 17 RMSE for WLF-DEE and the selected state-of-the-art metrics on TID-contrast change database.**
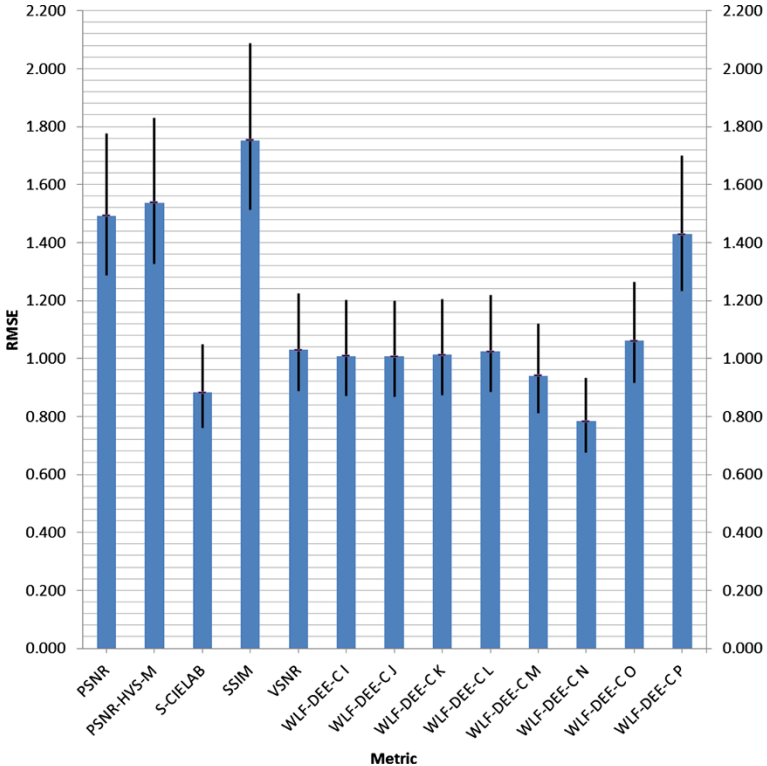
**Figure 18** RMSE for WLF-DEE and the selected state-of-the-art metrics on Ajagamelle database.



**Figure 19** RMSE for WLF-DEE and the selected state-of-the-art metrics on Pedersen database.

## Author details
[1]The Norwegian Colour and Visual Computing Laboratory, Faculty of Computer Science and Media Technology, Gjøvik University College, P.O. Box 191, Gjøvik 2802, Norway. [2]Dipartimento di Fisica, Universita' degli Studi di Parma, Parco Area delle Scienze 7/A, Parma 43100, Italy.

## References

1. A Poirson, B Wandell, Appearance of colored patterns: pattern-color separability. J. Opt. Soc. Am. A. **10**(12), 2458–2470 (1993)
2. M Pedersen, JY Hardeberg, Survey of full-reference image quality metrics: classification and evaluation. Foundations Trends. Comput. Graph. Vis. **7**, 1–80 (2012)
3. M Pedersen, N Bonnier, JY Hardeberg, F Albregtsen, Attributes of image quality for color prints. J. Electron. Imaging. **19**, 011016–1–011016–13 (2010)
4. Y Tadmor, D Tolhurst, Calculating the contrasts that retinal ganglion cells and LGN neurones encounter in natural scenes. Vis. Res. **40**, 3145–3157 (2000)
5. C Oleari, M Melgosa, R Huertas, Euclidean color-difference formula for small-medium color differences in log-compressed OSA-UCS space. J. Opt. Soc. Am. A **26**, 121–134 (2009)
6. H Von Helmholtz, Versuch das psychophysische Gesetz auf die Farbenunterschiede trichromatischer Augen anzuwenden. Z Psychol Physiol Sinnesorgane **3**, 1–20 (1892)
7. CIE, Colorimetry. Tech. Rep. 15. (CIE, Vienna, 2004)
8. Colorimetry – Part 4: CIE 1976 L*a*b* Colour space, ISO 11664-4:2008 (CIE S 014-4/E:2007). CIE, Vienna, 2006
9. MR Luo, B Rigg, Chromaticity-discrimination ellipses for surface colours. Color Res. Appli. **11**, 25–42 (1986)
10. FJJ Clarke, R McDonald, B Rigg, Modification to the JPC79 color-difference formula. J. Soc. Dyers Colorists. **100**, 128–132 (1984)
11. M Luo, B Rigg, BFD(l:c) colour-difference formula: Part I - development of the formula. J. Soc. Dyers Colorists. **103**, 86–94 (1987)
12. MR Luo, B Rigg, BFD(l:c) color-difference formula. Part II – performance of the formula. J. Soc. Dyers Colorists. **103**, 126–132 (1987)
13. CIE, Industrial colour-difference evaluation. CIE Technical Report 116, (Central Bureau of the CIE , Vienna, Austria, 1995)
14. M Luo, G Cui, B Rigg, The development of the CIE 2000 colour-difference formula: CIEDE2000. Color Res. Appli. **26**(5), 340–350 (2001)
15. E Bando, JY Hardeberg, D Connah, in *Human Vision and Electronic Imaging X, ed. B. Rogowitz, T. Pappas, S. Daly, Proc. of SPIE - IST Electronic Imaging,SPIE*, vol. 5666. Can gamut mapping quality be predicted by color image difference formulae (SPIE Bellingham, 2005), pp. 180–191
16. JY Hardeberg, E Bando, M Pedersen, Evaluating image difference metrics for gamut mapped images. Coloration Tech. **124**(4), 243–253 (2008)
17. H Kivinen, M Nuutinen, P Oittinen, in *CGIV 2010 – Fifth European Conference on Colour in Graphics, Imaging*. Comparison of colour difference methods for natural images (Joensuu Finland, 2010), pp. 510–515
18. G Hong, M Luo, in *Proceedings of SPIE: 9th Congress of the International Colour Association* vol. 4421, ed. by Chung R, Rodrigues A. Perceptually based colour difference for complex images (SPIE Bellingham, 2002), pp. 618–621
19. C Oleari, Color opponencies in the system of the uniform color scales of the Optical Society of America. J. Opt. Soc. Am. A. **21**, 677–682 (2004)
20. C Oleari, Hypotheses for chromatic opponency functions and their performance on classical psychophysical data. Color Res Appli. **30**, 31–41 (2005)
21. R Huertas, M Melgosa, C Oleari, Performance of a color-difference formula based on OSA-UCS space using small-medium color differences. J. Opt. Soc. Am. A. **23**(9), 2077–2084 (2006)
22. G Simone, C Oleari, I Farup, in *11th Congress of the International Colour Association (AIC)*. Performance of the Euclidean color-difference formula in log-compressed OSA-UCS space applied to modified image-difference metrics (Sydney Australia, 2009), p. 81
23. DR Pant, I Farup, Riemannian formulation, and comparison of color difference formulas. Color Res. Appl. **37**, 429–440 (2012)
24. N Ponomarenko, V Lukin, K Egiazarian, J Astola, M Carli, F Battisti, in *10th International Workshop on Multimedia Signal Processing*. Color image database for evaluation of image quality metrics (Cairns, Queensland Australia, 2008), pp. 403–408 [http://www.ponomarenko.info/tid2008.htm]. Accessed 20 Feb 2008
25. X Zhang, BA Wandell, A spatial extension of CIELAB for digital color image reproduction. SID J (1997) [21/05/05:http://white.stanford.edu/~brian/scielab/]. Accessed 20 Feb 2008
26. Z Wang, AC Bovik, HR Sheikh, EP Simoncelli, Image quality assessment: from error visibility to structural similarity. IEEE Trans. Image Proc. **13**(4), 600–612 (2004)
27. Z Wang, A Bovik, A universal image quality index. IEEE Signal Proc. Lett. **9**, 81–84 (2002)
28. K Egiazarian, J Astola, N Ponomarenko, V Lukin, F Battisti, M Carli, in *Proc. of the Second International Workshop on Video Proc. and Quality Metrics*. Two new full-reference quality metrics based on HVS. Scottsdale, Arizona, January 2006.
29. N Nill, A visual model weighted cosine transform for image compression and quality assessment. IEEE Trans. Commun. **33**, 551–557 (1985)
30. N Ponomarenko, F Silvestri, K Egiazarian, M Carli, J Astola, V Lukin, in *CD-ROM Proc. of the Third Int. Workshop on Video Proc. and Quality Metrics*. On between-coefficient contrast masking of DCT basis functions. Scottsdale, Arizona, 25–26 January, 2007 4 p
31. D Chandler, S Hemami, VSNR: A wavelet-based visual signal-to-noise ratio for natural images. IEEE Trans. Image Proc. **16**(9), 2284–2298 (2007)
32. G Simone, C Oleari, I Farup, in *Proceedings from Gjøvik Color Imaging Symposium 2009,* An alternative color difference formula for computing image difference. no. 4 in Høgskolen i Gjøviks rapportserie, Gjøvik Norway, 2009, pp. 8–11
33. GM Johnson, MD Fairchild, in *The 9th Color Imaging Conference: Color Science and Engineering: Systems, Technologies, Applications*. Darwinism of color image difference models. Scottsdale, Arizona, 2001, pp.108–112
34. G Simone, M Pedersen, JY Hardeberg, Measuring perceptual contrast in digital images. J. Vis. Commun. Image Representation. **23**, 491–506 (2012)
35. J Frankle, JJ McCann, Method and apparatus for lightness imaging. United States Patent No. 4,384,336 (1983)
36. EH Adelson, CH Anderson, JR Bergen, PJ Burt, JM Ogden, Pyramid methods in image processing. RCA Eng. **29**(6) (1984)
37. C Enroth-Cugell, JG Robson, The contrast sensitivity of retinal ganglion cells of the cat. J. Phys. **187**, 517552 (1966)
38. G Sclar, JH Maunsell, P Lennie, Coding of image contrast in central visual pathways of the macaque monkey. Vis. Res. **30**, 1–10 (1990)
39. R Rodieck, Quantitative analysis of cat retinal ganglion cell response to visual stimuli. Vis. Res. **5**(12), 583–601 (1965)
40. A Michelson, *Studies in Optics*. (University of Chicago Press, Chicago, 1927)
41. M Pedersen, JY Hardeberg, in *CGIV 2008 Fourth European Conference on Color in Graphics, Imaging and Vision*. Rank order and image difference metrics (Terrassa Spain: IS&T, 2008), pp. 120–125
42. S Ajagamelle, Analysis of the difference of Gaussians model in perceptual image difference metrics. *Master's thesis*, Gjøvik University College and Grenoble Institute of Technology (2009)
43. J Dijk, In search of an objective measure for the perceptual quality of printed images. *PhD thesis*. Technische Universiteit Delft, 2004
44. 12640-3, I: ISO 12640–3:2007 graphic technology - prepress digital data exchange - Part 3: CIELAB standard colour image data CIELAB/SCID, International Organization for Standardization 2007. [http://www.iso.org/iso/home/store/catalogue_tc/catalogue_detail.htm?csnumber=52497] Accessed 20 Feb 2008
45. MG Kendall, A Stuart, JK Ord, *Kendall's Advanced Theory of Statistics: Classical Inference and Relationship,* vol. 2, 5th edn. (Hodder Arnold Publisher, London, 1991)

46. QualityExperts Video, Objective perceptual assessment of video quality: full reference television. Tech. rep. (International Telecommunication Union,Geneva, 2004)
47. HR Sheikh, MF Sabir, AC Bovik, A statistical evaluation of recent full reference image quality assessment algorithms. IEEE Trans. Image Proc. **15**(11), 3440–3451 (2006)
48. Video Quality Experts Group, Final report from the video quality experts group: validation of reduced-reference and no-reference objective models for standard definition television. Phase I. Tech. rep. (International Telecommunication Union, Geneva, 2009)