Journal of Imaging Science and Technology® 56(5): 050501-1–050501-12, 2012.

A New Spatial Filtering Based Image Difference Metric Based on Hue Angle Weighting

Marius Pedersen▲ **and Jon Yngve Hardeberg**
Norwegian Color and Visual Computing Laboratory, Gjøvik University College, P.O. Box 191, N-2802 Gjøvik, Norway
E-mail: marius.pedersen@hig.no

Abstract. *Color image difference metrics have been proposed to find differences between an original image and a reproduction. One of these metrics is the hue angle algorithm proposed by Hong and Luo in 2002. This metric does not take into account the spatial properties of the human visual system, and it could therefore miscalculate the differences between the original and the reproduction. In this article we propose a new color image difference metric based on the hue angle algorithm that takes into account the spatial properties of the human visual system. The proposed metric, the Spatial Hue Angle Metric, has been subjected to extensive testing. The results show improvement in performance compared to the original metric proposed by Hong and Luo, and improvement over or similar performance to traditional metrics, such as the Structural Similarity Metric and Spatial-CIELAB.*

[DOI: 10.2352/J.ImagingSci.Technol.12.56.5.050501]

INTRODUCTION

Digital image processing systems are meant to create reproductions of visual information for observers. It is a wish that the image quality is not influenced by the processing, such as image compression, by the system. However, systems often have restrictions that influence image quality, creating a trade-off between the resources available and the perceptual quality. In these cases it is important to determine the quality loss of the system in order to produce high-quality reproductions.

There are basically two ways to judge image quality: subjectively or objectively. Subjective evaluation is carried out by observers, and is therefore influenced by the human visual system (HVS). Objective evaluation of image quality can be carried out using measurement devices gathering numerical values or using algorithms, commonly known as image quality metrics. Image quality metrics are usually developed to take into account properties of the HVS, thus with the goal of being well correlated with subjective evaluations.

Determining image quality is very complex and difficult since it is influenced by numerous criteria and the subjectivity of human observers. In order for observers to judge image quality, they must be able to perceive the differences between the images.[1] The perceived differences are used to determine the quality of the image. Being able to determine the differences between an original and a reproduction is considered to be easier than estimating the quality, and is a very important for obtaining an objective metric for image quality.

During the last two decades many different color image difference metrics have been proposed,[2–4] some for overall image difference and some for specific distortions. New and improved metrics are created every year, but so far no one has been able to create a universal color image difference metric. Most of the existing image difference metrics follow a similar framework, in which the images are filtered to simulate the HVS, before a calculation of the difference between them is carried out.

In this article we propose a new color image difference metric for perceptual image difference that takes into account the HVS. The proposed metric can be considered as a step towards predicting perceived image quality. The new metric is an extension of preliminary results.[5,6]

We start by giving a brief introduction to color difference formulas and image difference metrics, before we introduce the new image difference metric. Then the experimental evaluation of the proposed metric is presented together with a discussion of the results. Finally, we conclude and propose further research.

STATE OF THE ART

The CIE published the CIELAB ($L^*a^*b^*$) color space specification,[7] based on the idea of a perceptually uniform color space. By using the Euclidean distance it is straightforward to calculate the distance between two colors in this color space. Given a sample color with CIELAB values L_s^*, a_s^*, and b_s^* and a reference color L_r^*, a_r^*, and b_r^*, the distance between them are calculated as follows:

$$\Delta E_{ab}^* = \sqrt{(\Delta L^*)^2 + (\Delta a^*)^2 + (\Delta b^*)^2}, \tag{1}$$

where $\Delta L^* = L_s^* - L_r^*$, $\Delta a^* = a_s^* - a_r^*$, and $\Delta b^* = b_s^* - b_r^*$.

The most common way of using ΔE_{ab}^* as an image difference metric is by calculating the color difference in each pixel, and then to pool these to one value by taking the average:

$$\Delta E_{ab}^* = \frac{\sum_{x=1}^{m} \sum_{y=1}^{n} \Delta E_{ab_{(x,y)}}^*}{m \cdot n}, \tag{2}$$

▲ IS&T Member.

Received Feb. 21, 2012; accepted for publication Dec. 4, 2012; published online Jan. 21, 2013.
1062-3701/12/56(5)/050501/12/$20.00

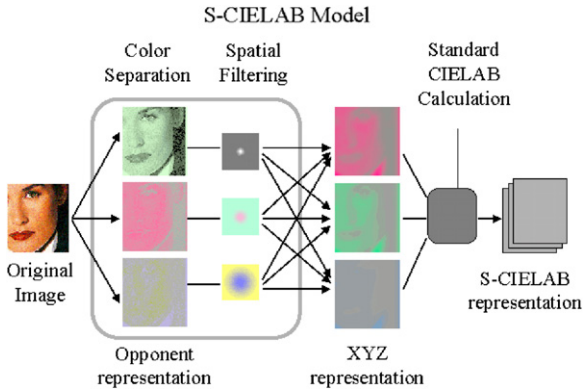

Figure 1. Flowchart of *S-CIELAB*. First, the image is converted to an opponent representation, where the spatial filters are applied to simulate the HVS. The filtered images are transformed in the CIEXYZ color space, and further to the CIELAB color space, where the CIELAB color difference formula is applied to get an *S-CIELAB* representation. Figure reproduced from Ref. 16.

where m is the width of the image and n is the height of the image. Other commonly used pooling methods are maximum, median, or minimum.[8]

When it was discovered that the CIELAB ΔE^*_{ab} did not correlate well enough with the perceptual color difference, the CIE proposed an extension, the ΔE_{94}.[9] Later, it was discovered that the ΔE_{94} suffered from similar problems as the ΔE^*_{ab}, and a new color difference formula, the CIE ΔE_{00},[10,11] was proposed. In addition to these, many other formulas exist, such as the ΔE_E,[12] CMC,[13] and BFD.[14]

Color difference formulas have commonly been used in image difference metrics. *S-CIELAB* from[15] is perhaps the best-known example: in it, a spatial pre-processing of the image, to simulate the HVS, is carried out before the difference is calculated by using the CIELAB color difference formula. The images are first separated into an opponent-color space, and each opponent-color image is convolved with a kernel determined by the visual spatial sensitivity of that color dimension. Finally, the filtered images are transformed into CIEXYZ, and further into CIELAB, where the pixelwise ΔE^*_{ab} is calculated. A flowchart of *S-CIELAB* is shown in Figure 1.

Some years later, Johnson and Fairchild[17] proposed an extension of *S-CIELAB*, in which the spatial filtering was refined by using the frequency domain rather than the spatial domain and the filters were modified to enhance the image differences where the human visual system is most sensitive to them. Johnson and Fairchild[17] also proposed to account for localized attention and local and global contrast.

Another image difference metric is the *hue angle algorithm* proposed by Hong and Luo,[18,19] which also is based on the CIELAB color difference. This metric corrects some of the drawbacks with the CIELAB color difference formula, for example that all pixels are weighted equally. The metrics shows good results for two different images.[18,19] Nevertheless, it does not include spatial filtering of the image and is therefore unsuitable to evaluate for instance halftoned images, where the viewing distance is crucial for the visual impression of quality. The *hue angle algorithm* has been shown to have problems in calculating perceived image difference.[20–22] Due to this we propose a new image difference metric based on the *hue angle algorithm* with spatial filtering simulating the human visual system, called the Spatial Hue Angle MEtric (*SHAME*).

For a complete review of image difference metrics we refer the reader to Pedersen and Hardeberg.[2,3]

THE PROPOSED METRIC

The new color image difference metric is based on the *hue angle algorithm* and the *S-CIELAB* workflow, in which a spatial filtering is applied before calculating the difference. We will first give an overview of the *hue angle algorithm*, and then describe two spatial filtering methods that are used.

The hue angle algorithm

Hong and Luo[18,19] proposed a full-reference color image difference metric based on the CIELAB color difference formula.[7] This metric is based on the known fact that systematic errors over the entire image are quite noticeable and unacceptable. The metric is based on some conjectures, summarized from Hong and Luo.[18,19] These are the following.

- Pixels or areas of high significance can be identified, and suitable weights can be assigned to these.
- Pixels in larger areas of the same color should be given a higher weight than those in smaller areas.
- Larger color differences between the pixels should get higher weights.
- Hue is an important color perception for discriminating colors within the context.

The first step is to transfer each pixel in the image from L^*, a^*, b^* to L^*, C^*_{ab}, h_{ab}. Based on the hue angles (h_{ab}) of the original a histogram is computed, and sorted in ascending order based on the number of pixels with same hue angle to an array k. Then weights can be applied to four different parts (quartiles) of the histogram, and by doing this Hong and Luo corrected the drawback that the CIELAB formula weights the whole image equally. The first quartile, containing n hue angles, is weighted with 1/4 (that is, the smallest areas with the same hue angle) and saved to a new array *hist*. The second quartile, with m hue angles, is weighted with 1/2. The third quartile, containing l hue angles, is given 1 as a weight, and the last quartile with the remaining hue angles is weighted with 9/4. These weights were empirically derived by Hong and Luo. The array *hist* is then computed as

$$hist(i) = \begin{cases} k(i) * 1/4, & i \in \{0, \ldots, n\} \\ k(i) * 1/2, & i \in \{n+1, \ldots, n+m\} \\ k(i) * 1, & i \in \{n+m+1, \ldots, n+m+l\} \\ k(i) * 9/4, & \text{otherwise.} \end{cases} \quad (3)$$

The average color difference, computed using ΔE^*_{ab}, is calculated for all pixels having the same hue angle and stored

in CD[*hue*]. Then the overall color difference for the image, CD_{image}, is calculated by multiplying the weights based on the quartiles for every pixel with the average CIELAB color difference for the hue angle:

$$CD_{image} = \sum hist[hue] * CD[hue]^2/4, \quad (4)$$

where the sum is over all hue angles.

Spatial filtering

We propose two different spatial filtering methods for the new metric, and evaluate these as a pre-processing step before applying the *hue angle algorithm*. The first spatial filtering is adopted from *S-CIELAB*.[15] The advantage of using this spatial filtering is that it has been extensively used and evaluated,[5,20–35] and it has been shown to produce good results for a wide variety of distortions.

First, the input image goes through color space transformations. The RGB image is transformed into the CIEXYZ color space, before it is further transformed into the opponent-color space (O_1, O_2, and O_3):

$$O_1 = 0.279X + 0.72Y - 0.107Z \quad (5)$$

$$O_2 = -0.449X + 0.29Y - 0.077Z \quad (6)$$

$$O_3 = 0.086X - 0.59Y + 0.501Z. \quad (7)$$

Now, the image contains a channel with the luminance information (O_1), one with the red–green information (O_2), and one with blue–yellow information (O_3). The next step consists of applying the spatial filters, where the information in each channel is filtered by a two-dimensional separable spatial kernel:

$$f = k \sum_i w_i E_i, \quad (8)$$

where

$$E_i = k_i e^{[-(x^2+y^2)/\sigma_i^2]}, \quad (9)$$

and k_i normalize E_i such that the filter sums to 1. The parameters w_i and σ_i are dependent on the color planes, and are defined in Table I. k is a scale factor, which normalizes each color plane so its two-dimensional kernel f sums to 1.

The second spatial filtering proposed is adopted from Johnson and Fairchild.[17] By specifying and implementing the spatial filters using contrast sensitivity functions (CSFs) in the frequency domain rather than in the spatial domain as the first spatial filtering, more precise control of the filters is obtained,[17] but usually at the cost of computational complexity. The luminance filter is based on research by Movshon and Kiorpes,[36] and it is a three-parameter exponential function:

$$CSF_{lum}(p) = a \cdot p^c \cdot e^{-b \cdot p}, \quad (10)$$

where $a = 75$, $b = 0.22$, $c = 0.78$, and p is cycles per degree (CPD).

Table I. The parameters used for the spatial filtering, where w_i is the weight of the plane and σ_i is the spread in degrees of visual angle as described by Zhang and Wandell.[15]

Plane	Weight w_i	Spread σ_i
Luminance	0.921	0.0283
	0.105	0.133
	−0.108	4.336
Red–green	0.531	0.0392
	0.330	0.494
Blue–yellow	0.488	0.0536
	0.371	0.386

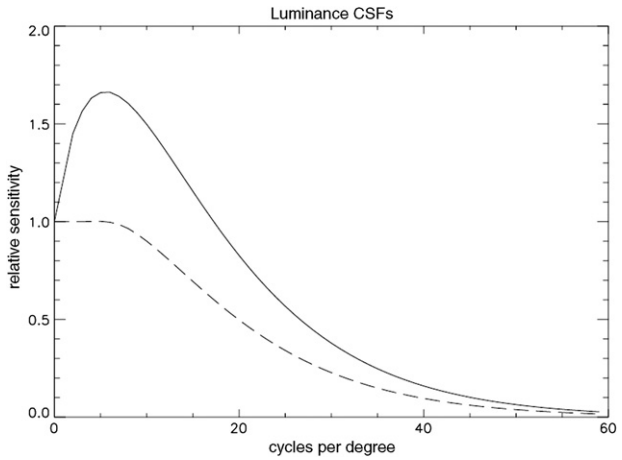

Figure 2. Luminance CSF from Eq. (10) with (solid) and without (dotted) the normalization. Figure reproduced from Johnson and Fairchild.[17]

The luminance CSF is normalized so that the DC modulation is set to 1.0. This will also enhance any image differences where the human visual system is most sensitive to them.[17] Figure 2 shows the luminance CSF from Eq. (10) with and without the normalization, showing that our HVS is more sensitive to low–medium spatial frequencies. For the chrominance CSF a sum of two Gaussian functions is used:

$$CSF_{chroma}(p) = a_1 \cdot e^{-b_1 \cdot p^{c_1}} + a_2 \cdot e^{-b_2 \cdot p^{c_2}}, \quad (11)$$

where different parameters for a_1, a_2, b_1, b_2, c_1, and c_2 have been used, as seen in Table II. Research indicates that metrics using the modified filters from Johnson and Fairchild[17] show increased correlation with human observers.[5,17,37]

Applying spatial filtering to the hue angle algorithm

The images are spatially filtered with one of the previously introduced methods. This is done by applying color transformations from the RGB space to the CIEXYZ color space, and further into the opponent-color space (O_1, O_2, and O_3) where the filters are applied. This is carried out for both the original and the reproduction, which results in filtered versions of both. The filtered original and filtered reproduction are used as input to the *hue angle algorithm*,

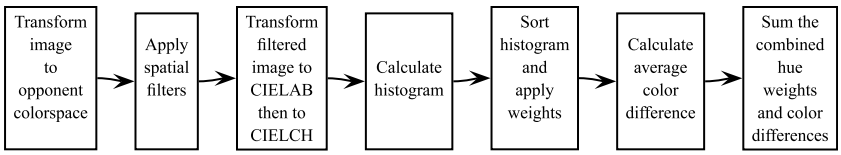

Figure 3. Workflow of the proposed metric.

Table II. The parameters used for the spatial filtering of the chrominance channels in the frequency domain.

Parameter	Red–green	Blue–yellow
a_1	109.141,30	7.032,845
b_1	−0.000,38	−0.000,004
c_1	3.424,36	4.258,205
a_2	93.597,11	40.690,950
b_2	−0.003,67	−0.103,909
c_2	2.167,71	1.648,658

as shown in Figure 3. This new metric can be considered as an extension of the *S-CIELAB* flowchart on Fig. 1.

The *hue angle algorithm*, filtered with the first filter and the second filter, is from now on referred to as *SHAME* and *SHAME-II*, respectively. The new metric will theoretically have several key features from both the *S-CIELAB* and the *hue angle measure*.

- Weight allocation: pixels in larger areas of the same color should be weighted higher.
- Simulation of the spatial properties of the human visual system.
- Undetectable distortions are ignored.
- Suitable for different kind of distortions, not only color patches.
- Quality values are pooled to one value for easy interpretation.

EXPERIMENTAL SETUP

We will evaluate the performance of the new image difference metric by comparing its results to the results of human observers. The performance will also be compared to other state-of-the-art metrics.

Image datasets

Many different image databases have been proposed for evaluation of image difference metrics. We have selected five state-of-the-art databases for the evaluation of the new metric. These databases cover a wide range of distortions and quality attributes.

TID2008

The TID2008 database[38] has been used for evaluation of the proposed metric. This database contains a total of 1700 images, with 25 reference images with 17 types of distortion (Table III) over four distortion levels. The mean opinion scores (MOSs) are the results of 654 observers attending the experiments. For the viewing distance, since this was not fixed in the TID2008 database, we have used a standard viewing distance of 50 cm on a normal 19 inch screen.

Dugay

For the dataset from Dugay, 20 original images were gamut mapped with five different algorithms.[39,40]

- HPminDE (Hue preserving minimum ΔE^*_{ab} clipping) which is a baseline gamut mapping algorithm proposed by the CIE.[41]
- SGCK (sigmoidal lightness mapping and cusp knee scaling)[41] is an advanced spatially invariant sequential gamut compression algorithm.
- A spatial gamut mapping algorithm proposed by Zolliker and Simon[42] with the intent to recover local contrast while preserving lightness, saturation, and global contrast.
- A hue- and edge-preserving spatial color gamut mapping algorithm proposed by Kolås and Farup.[43]
- A multiscale algorithm preserving hue and local relationship between closely related pixel colors proposed by Farup et al.[44]

The images were evaluated by 20 observers in a comparison experiment, from a viewing distance of 50 cm. Gamut mapping results in many simultaneous changes, making the changes in this database difficult to predict by image quality metrics.

Pedersen et al.

Pedersen et al.[21] proposed a dataset with four original images, three portraits, and one illustration. The originals were altered in lightness: each image had four versions with global lightness differences, and four versions with local lightness changes. The lightness changes were 3 and 5 ΔE^*_{ab}. Four versions were brighter than the original, and four darker. A total of 25 observers were recruited for the experiment, and they were asked in a pair-wise comparison experiment to choose the image most similar to the original. The viewing distance was set to 80 cm.

Ajagamelle

This database from Ajagamelle contains a total of ten original images covering a wide range of characteristics and scenes.[45] 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. These images were evaluated by 14 observers from a viewing distance of 70 cm.

Table III. Overview of the distortions in the TID database and how they are related to the tested subsets. The database contains 17 types of distortion over four distortion levels. The sign "+" indicates that the distortion type was used to alter the images of the subset and the sign "−" that it was not considered for this subset.

	Type of distortion	Noise	Noise2	Safe	Hard	Simple	Exotic	Exotic2	Full
1	Additive Gaussian noise	+	+	+	−	+	−	−	+
2	Noise in color components	−	+	−	−	−	−	−	+
3	Spatially correlated noise	+	+	+	+	−	−	−	+
4	Masked noise	−	+	−	+	−	−	−	+
5	High-frequency noise	+	+	+	−	−	−	−	+
6	Impulse noise	+	+	+	−	−	−	−	+
7	Quantization noise	+	+	−	+	−	−	−	+
8	Gaussian blur	+	+	+	+	+	−	−	+
9	Image denoising	+	−	−	+	−	−	−	+
10	JPEG compression	−	−	+	−	+	−	−	+
11	JPEG2000 compression	−	−	+	−	+	−	−	+
12	JPEG transmission errors	−	−	−	+	−	−	+	+
13	JPEG2000 transmission errors	−	−	−	+	−	−	+	+
14	Noneccentricity pattern noise	−	−	−	+	−	+	+	+
15	Local block-wise distortion	−	−	−	−	−	+	+	+
16	Mean shift	−	−	−	−	−	+	+	+
17	Contrast change	−	−	−	−	−	+	+	+

IVC

The IVC database[46] contains a total of 235 images, with 10 original images distorted by three types of lossy compression techniques (JPEG, JPEG2000, and Locally Adaptive Resolution) and blurring process. For this database we only use the color images, resulting in 180 images. Subjective evaluations were made by 15 observers using the Double Stimulus Impairment Scale method with 5 categories at a distance of six times the height of the screen.

Image difference metrics

The new metric, with the two different spatial filtering methods, is compared against the original *hue angle algorithm*,[18,19] pixelwise ΔE^*_{ab}, Spatial-CIELAB (*S-CIELAB*)[15] and Spatial-CIELAB from Johnson and Fairchild ($S\text{-}CIELAB_{\text{Johnson}}$)[17] to see if the segmentation done according to the hue angles and the spatial filtering improves the performance of the metric. We also include two other color image difference metrics, the Spatial-ΔE_E (*S-DEE*),[37] based on $S\text{-}CIELAB_{\text{Johnson}}$ and the ΔE_E color difference formula,[12] and the *Adaptive Bilateral Filter* (*ABF*),[47,48] based on bilateral filtering and ΔE^*_{ab}. Additionally, we compare *SHAME* against the Structural Similarity Metric (*SSIM*)[49] and Universal Image Quality index (*UIQ*),[50] both based on structural similarity. The *Peak-Signal-to-Noise-Ratio* (*PSNR*) is a commonly used metric in the industry, and therefore we include it in our evaluation. Two extensions of the *PSNR* are added to the evaluation list as well, the *PSNR*-Human Visual System metric (*P-HVS*)[51] and *P-HVS* with contrast masking (*P-HVS-M*).[52] Lastly, the *Visual-Signal-to-Noise-Ratio* (*VSNR*),[53] which is based on near-threshold and suprathreshold properties of the HVS, and the *Visual Information Fidelity* (*VIF*),[54] which quantifies the Shannon information present in the reproduction relative to the information present in the original, are included in the evaluation. Using these 12 state-of-the-art metrics will provide an extensive evaluation of the proposed metric, and it will show differences between the two proposed spatial filtering methods used in *SHAME*.

For *SHAME* and *SHAME-II*, the images have been transformed from 8-bit sRGB images to CIELAB using Matlab and the srgb2lab function.

Performance measures

In order to measure the performance of the image difference metrics, we compare the results of the metrics to the results of observers. Two standard types of correlation coefficient are computed[55] and used as performance measures.

(1) *The Pearson product moment*: this 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.[56]
(2) *Spearman rank*: this 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. It 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.[56]

The relationship between the metrics and the observers are not necessarily linear. In order to remove any non-linearities from the subjective experiments and to obtain a fair comparison of the metrics in a common analysis

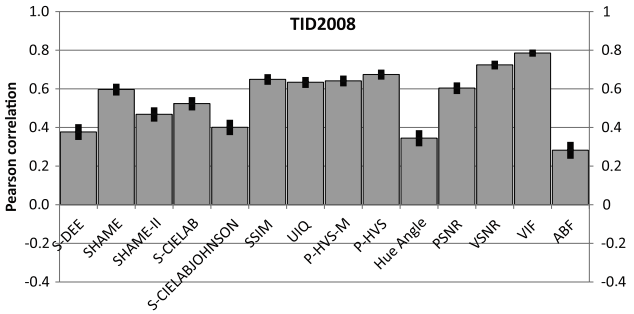

Figure 4. Pearson and Spearman correlation coefficients for on the TID2008 database. *VIF* is the highest-performing metric for both correlation types. *SHAME* outperforms the *hue angle algorithm*, *S-CIELAB*, and *SHAME-II*.

space, we investigate the relationship between the metrics and observers by using non-linear regression.[56] We use the same mapping function as Sheikh et al.: [54]

$$f(x) = \theta_1 \left(\frac{1}{2} - \frac{1}{1 + e^{\theta_2 (x - \theta_3)}} \right) + \theta_4 X + \theta_5, \qquad (12)$$

where θ_i, $i = 1, 2, 3, 4, 5$, are parameters to be fitted. In addition to the correlation values we will also report the 95% confidence intervals for the correlation values, which are calculated using Fisher's Z transformation as described by the Video Quality Expert Group.[57]

RESULTS

TID2008 database

VIF is the best-performing metric for the TID 2008 database, as seen in Table IV and Figure 4. *SHAME* has a significantly higher Pearson correlation than *SHAME-II*, and it is significantly better than *S-CIELAB* and the *hue angle algorithm*. This shows that the weighting carried out by the *hue angle algorithm* combined with the filtering from *S-CIELAB* improves the performance. It is also interesting to notice that *SHAME* is the best-performing metric among those based on color differences.

The TID2008 database contains different subsets, as seen in Table III. For the five first datasets (noise, noise2, safe, hard, and simple) *SHAME* and *SHAME-II* perform well (Table V); *SHAME* has a slightly higher Pearson correlation than *SHAME-II*. We can also notice that *SHAME* and *SHAME-II* have correlation values higher than *S-CIELAB*, *S-CIELAB*$_{\text{Johnson}}$, and the *hue angle algorithm*. This indicates that the combination of weighting and spatial filtering improves the performance of the metric. For the two last datasets (exotic and exotic2) all metrics have a decrease in performance, which is mostly caused by scale differences where different images are rated similarly by observers but very differently by the metrics.

For the Spearman correlation (Table VI), *SHAME* and *SHAME-II* have a decrease in performance compared to the Pearson correlation values (Table V). It is interesting to notice that *SHAME-II* has slightly higher values than *SHAME*, this being the opposite of what is found for the Pearson correlation values.

Table IV. Pearson and Spearman correlation coefficients for on the TID2008 database. *VIF* is the highest-performing metric for both correlation types. *SHAME* outperforms the *hue angle algorithm*, *S-CIELAB*, and *SHAME-II*.

Metric	Pearson correlation	Spearman correlation
S-DEE	0,38	0,29
SHAME	0,60	0,38
SHAME-II	0,47	0,41
S-CIELAB	0,52	0,46
S-CIELAB$_{\text{Johnson}}$	0,40	0,31
SSIM	0,65	0,64
UIQ	0,63	0,60
P-HVS-M	0,64	0,64
P-HVS	0,67	0,67
Hue angle algorithm	0,34	0,31
PSNR	0,60	0,59
VSNR	0,72	0,72
VIF	0,79	0,75
ABF	0,28	0,26

As an example where *SHAME* and *SHAME-II* are in agreement, we show the metric values for a one-reference image of the TID2008 database for JPEG compression over four levels (Figure 5). We also show the results of *S-CIELAB* in order to have a reference metric to compare against. For this image and distortion the metrics are in agreement with the observers as well, showing excellent correlation against the observers (Figure 6).

We also show an example where the metrics are not in agreement (Figure 7). This image, being the 14th in the TID2008 database, has four levels of local block-wise distortions of different intensity. Comparing the results of the metrics to the observers (Figure 8), we see that the metrics perform differently. The Pearson correlation coefficients are for *SHAME* 0.63, *SHAME-II* −0.30, and *S-CIELAB* 0.99. For *SHAME* and *SHAME-II*, the correlation is lower than for *S-CIELAB*, and this can be explained by the sorting and weighting performed in these metrics. The local distortions are small, and the hue of the local distortions does not occur in the hues that have the highest pixel counts, and therefore they are given a low weight in *SHAME* and *SHAME-II*. The local block-wise distortions of different intensity is also a part of the exotic and exotic2 datasets, which might also partly explain the low correlation found in these datasets for *SHAME* and *SHAME-II*.

Dugay

Figure 9 and Table VII shows the results from the dataset with gamut mapped images. In general all metrics have a low performance. This is probably because the task is very complex, since multiple attributes are changed simultaneous and the observers may judge them differently.[39,40] Previous research has shown that image difference metrics have

Table V. Pearson correlation of the different subsets of the TID2008 database. $N = 700, 800, 700, 800, 400, 400, 600$ for the noise, noise2, safe, hard, simple, exotic, and exotic2 datasets, respectively.

Metric	Noise	Noise2	Safe	Hard	Simple	Exotic	Exotic2
S-DEE	0,64	0,64	0,66	0,69	0,64	0,22	0,17
SHAME	0,86	0,84	0,85	0,84	0,86	0,22	0,25
SHAME-II	0,75	0,73	0,72	0,72	0,70	0,25	0,20
S-CIELAB	0,84	0,82	0,83	0,83	0,85	0,55	0,35
$S\text{-}CIELAB_{Johnson}$	0,66	0,66	0,71	0,59	0,69	0,31	0,20
SSIM	0,58	0,64	0,66	0,82	0,78	0,27	0,59
UIQ	0,54	0,61	0,67	0,76	0,81	0,17	0,59
P-HVS-M	0,93	0,89	0,93	0,82	0,95	0,53	0,43
P-HVS	0,92	0,89	0,92	0,84	0,94	0,53	0,45
Hue angle algorithm	0,55	0,52	0,53	0,68	0,65	0,52	0,21
PSNR	0,76	0,74	0,76	0,68	0,85	0,39	0,46
VSNR	0,85	0,84	0,86	0,76	0,91	0,61	0,61
VIF	0,82	0,90	0,92	0,84	0,95	0,56	0,68
ABF	0,59	0,47	0,60	0,57	0,67	0,62	0,50

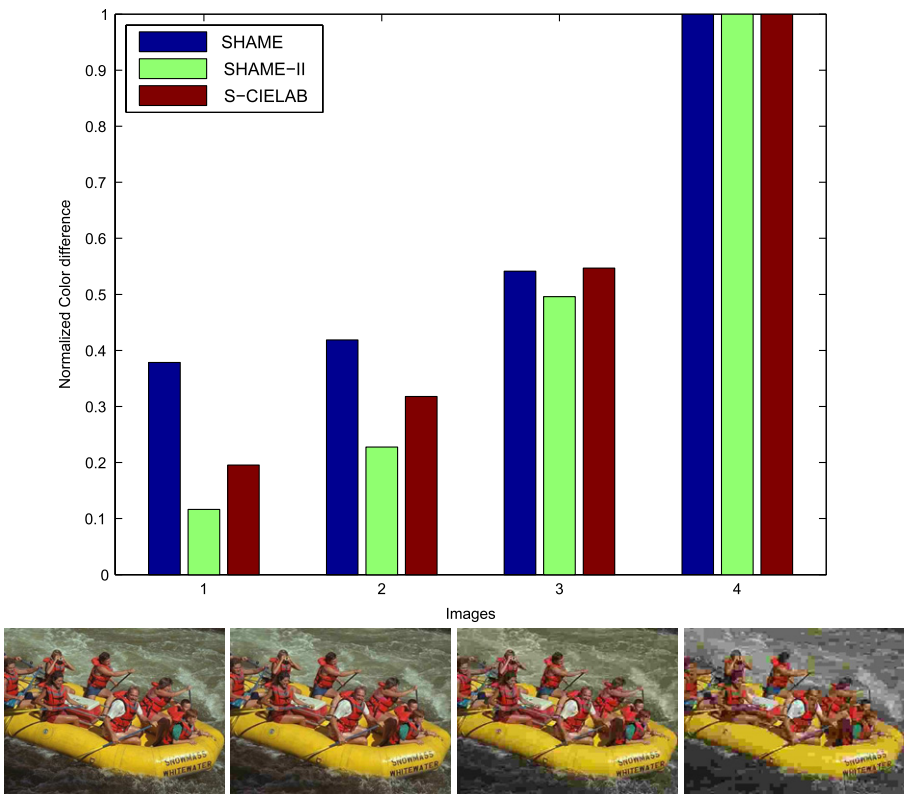

Figure 5. Example of agreement between *SHAME*, *SHAME-II*, and *S-CIELAB* for the 14th image with JPEG compression. This is for a one-reference image, the 14th in the TID2008 database, with four different levels of JPEG compression (shown below with increasing compression corresponding to image 1–4 in the barplot). The metric values have been normalized by the highest value for easy comparison between the metrics.

problems when multiple distortions occur simultaneously, as in gamut mapping.[23,26] Both *SHAME* and *SHAME-II* have confidence intervals overlapping with UIQ, which is the metric with the highest correlation.

Pedersen

For the Pedersen database, we can see that *SHAME-II* has the highest Pearson correlation (Figure 10 and Table VIII), but several other metrics overlap the 95% confidence interval

Table VI. Spearman correlation of the different subsets of the TID2008 database. $N = 700, 800, 700, 800, 400, 400, 600$ for the noise, noise2, safe, hard, simple, exotic, and exotic2 datasets, respectively.

Metric	Noise	Noise2	Safe	Hard	Simple	Exotic	Exotic2
S-DEE	0,43	0,47	0,42	0,49	0,37	0,34	−0,09
SHAME	0,55	0,52	0,54	0,43	0,49	0,28	−0,08
SHAME-II	0,66	0,62	0,60	0,66	0,58	0,25	0,28
S-CIELAB	0,79	0,76	0,77	0,79	0,80	0,27	0,28
S-CIELAB$_{\text{Johnson}}$	0,62	0,56	0,59	0,64	0,58	0,23	0,06
SSIM	0,57	0,64	0,64	0,82	0,77	0,29	0,54
UIQ	0,53	0,60	0,65	0,76	0,80	0,26	0,53
P-HVS-M	0,93	0,88	0,93	0,83	0,94	0,52	0,41
P-HVS	0,92	0,88	0,93	0,84	0,93	0,53	0,41
Hue angle algorithm	0,53	0,46	0,48	0,69	0,66	0,60	0,23
PSNR	0,75	0,73	0,75	0,69	0,85	0,41	0,40
VSNR	0,85	0,84	0,86	0,77	0,90	0,58	0,59
VIF	0,82	0,90	0,91	0,84	0,94	0,51	0,67
ABF	0,59	0,47	0,58	0,55	0,66	0,72	0,63

Table VII. *SHAME* and *SHAME-II* compared against other metrics for a set of gamut mapped images from Dugay.[39,40]

Metric	Pearson correlation	Spearman correlation
S-DEE	0,07	0,00
SHAME	0,25	0,27
SHAME-II	0,24	0,28
S-CIELAB	0,22	0,21
S-CIELAB$_{\text{Johnson}}$	0,02	0,03
SSIM	0,16	0,05
UIQ	0,43	0,19
P-HVS-M	0,18	0,21
P-HVS	0,18	0,22
Hue angle algorithm	0,07	0,12
PSNR	0,10	0,03
VSNR	0,10	0,10
VIF	0,31	0,27
ABF	0,08	0,11

of *SHAME-II*. *SHAME-II* is also significantly better than the *hue angle algorithm* since their confidence intervals do not overlap, indicating that incorporation of models of the human visual system improves the metric. The Spearman correlation for *SHAME-II* is also the highest (Table VIII), indicating a similar ranking to that of the observers. Since the changes in this database are according to color differences, a metric calculating image difference with a color difference formula is likely to have a better performance than metrics using other approaches such as SSIM and VSNR.

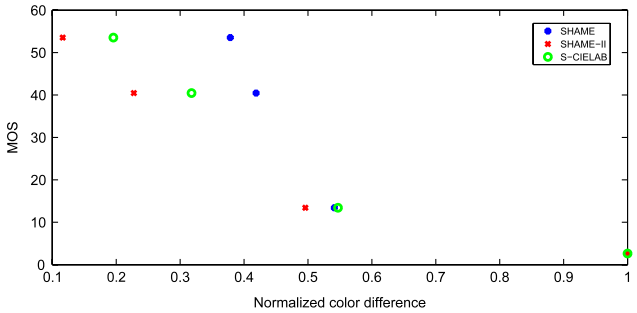

Figure 6. Values from *SHAME*, *SHAME-II* and *S-CIELAB* for the 14th image with JPEG compression plotted against MOS. As we can see from the plot, all metrics are in agreement with the observers. The Pearson correlation coefficients are for *SHAME* 0.86, *SHAME-II* 0.93, and *S-CIELAB* 0.93. All have a perfect Spearman correlation coefficient of 1.

Ajagamelle

For the Ajagamelle database,[45] *P-HVS* has the highest Pearson and Spearman correlation values (Figure 11 and Table IX). However, with a 95% confidence interval it cannot be said to be significantly different from many other metrics, including both *SHAME* and *SHAME-II* (Fig. 11). The hue angle algorithm has a lower correlation value than the two SHAME variations, giving an indication that the spatial filtering contributes to increased performance, but they are not significantly different given the 95% confidence interval. The changes in this database are related to color differences (the lightness and saturation changes), giving the metrics based on color differences an advantage, while the metrics based on structural information and contrast are more adapted to the changes in contrast.

IVC

VIF has the highest Pearson and Spearman correlation for the IVC database, as seen in Figure 12 and Table X. None

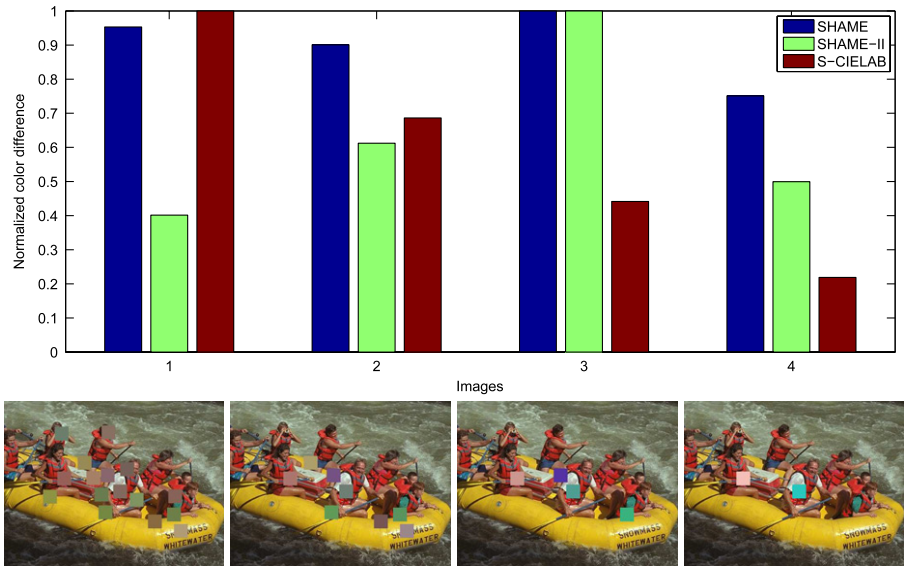

Figure 7. Example of disagreement between *SHAME*, *SHAME-II*, and *S-CIELAB* for the 14th image with local block-wise distortions of different intensity (shown below from left to right corresponding to image 1–4).

Table VIII. Pearson and Spearman correlation values for the lightness changed image from Pedersen et al.[21,22]

Metric	Pearson correlation	Spearman correlation
S-DEE	0,66	0,70
SHAME	0,81	0,71
SHAME-II	0,94	0,94
S-CIELAB	0,85	0,81
S-CIELAB$_{\text{Johnson}}$	0,81	0,76
SSIM	0,22	0,49
UIQ	0,45	0,49
P-HVS-M	0,75	0,68
P-HVS	0,63	0,67
Hue angle algorithm	0,83	0,73
PSNR	0,66	0,72
VSNR	0,02	0,02
VIF	0,39	0,35
ABF	0,86	0,81

of the other metrics overlap the confidence interval of *VIF*, but many metrics are close. *SHAME* performs well for this database, with Pearson and Spearman correlation above 0.7, slightly higher than *S-CIELAB* but not significantly different. *SHAME-II* is among the worst-performing metrics, and there is significant difference between the filtering from *S-CIELAB* and *S-CIELAB*$_{\text{Johnson}}$. However, it is interesting to see that the correlations for *SHAME* and *SHAME-II* are slightly higher than those of *S-CIELAB* and *S-CIELAB*$_{\text{Johnson}}$, indicating that the weighting by the *hue angle algorithm* contributes to an increased performance, but the differences are not statistically significant.

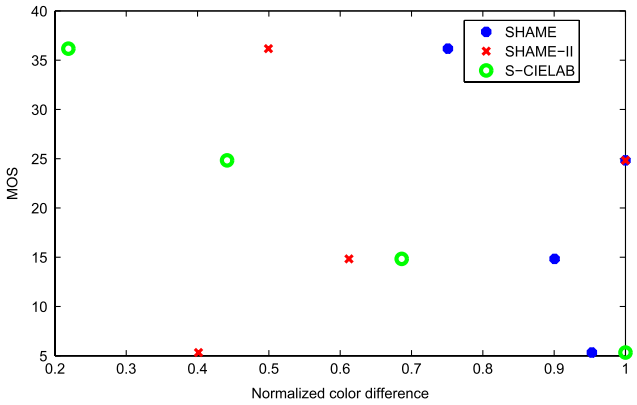

Figure 8. Values from *SHAME*, *SHAME-II*, and *S-CIELAB* for the 14th image with local block-wise distortions of different intensity plotted against MOS. As we can see from the plot, *S-CIELAB* is highly correlated with observers (correlation coefficient of *0.99*), while *SHAME* has a correlation coefficient of *0.63* and *SHAME-II* is not correlated (*−0.3*).

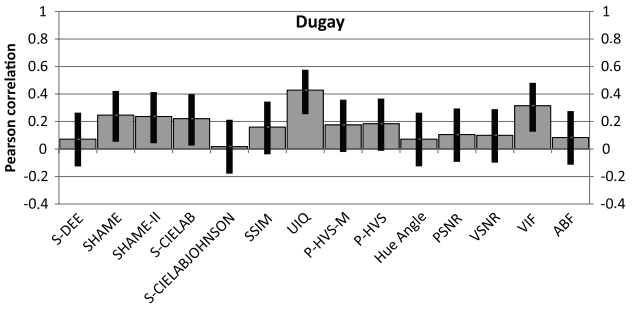

Figure 9. Pearson correlation values for the gamut mapped images from Dugay. All metrics have a low performance on the gamut mapped images, indicating that calculating the difference between an original and a gamut mapped image is very difficult for image difference metrics.

Overall observations

In many of the databases the performance of *SHAME* and *SHAME-II* is similar, such as for the Ajagamelle and

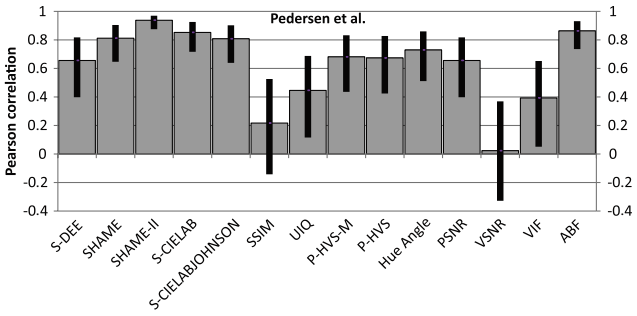

Figure 10. *SHAME* and *SHAME-II* compared against other metrics for the lightness changed images from Refs. 21,22. We notice that *SHAME-II* has the highest Pearson correlation, and it is significantly better than the *hue angle algorithm*.

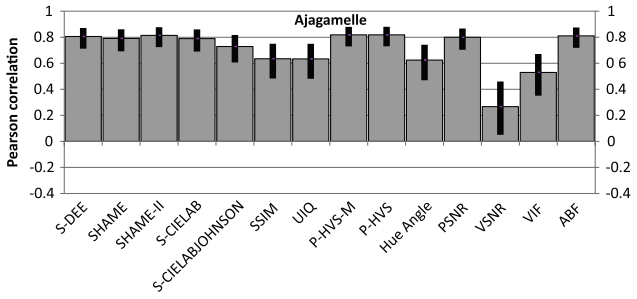

Figure 11. *SHAME* and *SHAME-II* compared against other metrics for the Ajagamelle database.[45] Both *SHAME* and *SHAME-II* perform similar to other state-of-the-art metrics, such as *S-DEE* and *S-CIELAB*.

Table IX. Pearson and Spearman correlation values for the Ajagamelle database.[45] Both *SHAME* and *SHAME-II* perform similar to other state-of-the-art metrics, such as *S-DEE* and *S-CIELAB*.

Metric	Pearson correlation	Spearman correlation
S-DEE	0,81	0,75
SHAME	0,79	0,73
SHAME-II	0,81	0,79
S-CIELAB	0,79	0,73
$S\text{-}CIELAB_{Johnson}$	0,73	0,68
SSIM	0,64	0,72
UIQ	0,63	0,73
P-HVS-M	0,82	0,79
P-HVS	0,82	0,80
Hue angle algorithm	0,62	0,72
PSNR	0,80	0,71
VSNR	0,27	0,32
VIF	0,53	0,59
ABF	0,81	0,76

Dugay databases. In IVC, *SHAME* is significantly better than *SHAME-II*, and this is also seen for the complete TID2008 database for the Pearson correlation values. However, when looking at the Spearman correlation values for the datasets in TID2008, *SHAME-II* has for all except one dataset a higher

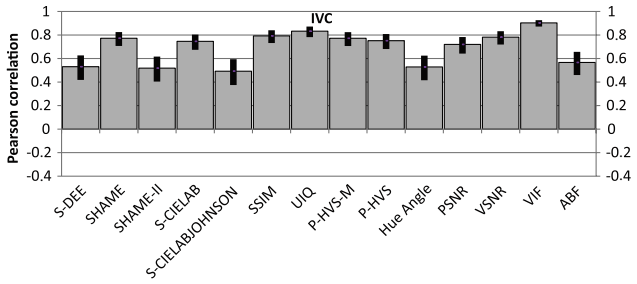

Figure 12. *SHAME* and *SHAME-II* compared against other metrics for the IVC database.[46] We notice that *SHAME* outperforms *SHAME-II*, $S\text{-}CIELAB_{Johnson}$, and *the hue angle algorithm*. *VIF* is the only metric that is statistically significantly better than *SHAME*.

Table X. Pearson and Spearman correlation values for the IVC database.[46]

Metric	Pearson correlation	Spearman correlation
S-DEE	0,53	0,49
SHAME	0,77	0,76
SHAME-II	0,52	0,51
S-CIELAB	0,75	0,74
$S\text{-}CIELAB_{Johnson}$	0,49	0,46
SSIM	0,79	0,78
UIQ	0,83	0,83
P-HVS-M	0,77	0,77
P-HVS	0,75	0,75
Hue angle algorithm	0,53	0,52
PSNR	0,72	0,69
VSNR	0,78	0,78
VIF	0,90	0,90
ABF	0,57	0,55

correlation than *SHAME*. For many of the other databases *SHAME-II* maintains a higher Spearman correlation than *SHAME*, indicating that it has a better agreement with the ranking of the observers. With the current databases and their results it is difficult to prefer one of the metrics over the other. Selection of which metric to use will most likely be dependent on the application and the distortion.

CONCLUSION AND FURTHER RESEARCH

The proposed metric, *SHAME*, uses well-known spatial filtering methods to improve a color image difference metric, which results in several advantages. Extensive testing of the proposed metrics (*SHAME* and *SHAME-II*) shows improvement over or similar performance to traditional metrics, such as *SSIM* and *S-CIELAB*, for several existing databases. Compared to other metrics based on color differences the proposed metrics (*SHAME* and *SHAME-II*) are always amongst the metrics with the best performance. We have demonstrated the importance of weighting areas of

interest and the importance of spatial filtering for color image difference metrics.

State-of-the-art image difference metrics also show weaknesses when judging the difference between an original and a modified version of it when more than one distortion occurs, and more research should be carried out to improve the metrics in this field, in terms of both difference calculation and spatial filtering. An interesting possible direction would be to adapt the weighting of the quartiles to characteristics of the image, making the weights image dependent.

ACKNOWLEDGMENTS

The authors would like to thank Gabriele Simone and Fritz Albregtsen for their advice, suggestions, and feedback regarding this project.

REFERENCES

[1] G. M. Johnson and M. D. Fairchild, "From color image difference models to image quality metrics," *Proc. ICIS'02: Int'l. Congress on Imaging Science* (Tokyo, Japan, 2002), pp. 326–327.
[2] M. Pedersen and J. Y. Hardeberg, "Survey of full-reference image quality metrics: Classification and evaluation," Found. Trends Comput. Graph. Vision **7**, 1–80 (2012).
[3] M. Pedersen and J. Y. Hardeberg, Survey of full-reference image quality metrics. GUC Reports 5, Gjøvik University College, (2009) ISSN: 1890-520X.
[4] K.-H. Thung and P. Raveendran, "A survey of image quality measures," *Int'l. Conf. for Technical Postgraduates (TECHPOS)* (Kuala Lumpur, Malaysia, 2009, IEEE), pp. 1–4.
[5] M. Pedersen and J. Y. Hardeberg, "A new spatial hue angle metric for perceptual image difference," *Computational Color Imaging* Lecture Notes in Computer Science (Saint Etienne, France, 2009 Springer Berlin/Heidelberg), Vol. 5646, pp. 81–90.
[6] M. Pedersen and J. Y. Hardeberg, "SHAME: a new spatial hue angle metric for perceptual image difference," J. Vision **9**, 343 (2009) ISSN 1534-7362.
[7] CIE, Colorimetry. Technical Report 15, 2004.
[8] M. Gong and M. Pedersen, "Spatial pooling for measuring color printing quality attributes," J. Visual Commun. Image Representation **23**, 685–696 (2012).
[9] CIE, Industrial colour-difference evaluation, publication CIE bureau central de la CIE, 1995, pp. 116–95.
[10] M. R. Luo, G. Cui, and B. Rigg, "The development of the CIE 2000 colour-difference formula: CIEDE2000," Color Res. Appl. **26**, 340–350 (2001).
[11] CIE "CIE 142-2001: Improvement to industrial colour-difference evaluation," Technical Report, 2001. ISBN 978 3 901906 08 4.
[12] C. Oleari, M. Melgosa, and 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).
[13] F. J. J. Clarke, R. McDonald, and B. Rigg, "Modification to the JPC 79 colour difference formula," J. Soc. Dyers Colourists **100**, 128–132 (1984).
[14] M. R. Luo and B. Rigg, "BFD(l:c) colour-difference formula: Part 1—development of the formula," J. Soc. Dyers Colourists **103**, 86–94 (1987).
[15] X. Zhang and B. A. Wandell, "A spatial extension of CIELAB for digital color image reproduction," *Soc. Inform. Display 96 Digest* (San Diego, CA, 1996), pp. 731–734.
[16] X. Zhang, Introduction to s-cielab, http://white.stanford.edu/~brian/scielab/introduction.html. Visited 21/02/12.
[17] G. M. Johnson and M. D. Fairchild, "Darwinism of color image difference models," *Proc. IS&T/SID 9th Color Imaging Conf.* (IS&T, Springfield, VA 2001), pp. 108–112.
[18] G. Hong and M. R. Luo, "Perceptually based colour difference for complex images," *9th Congress of the Int'l. Colour Assoc.*, edited by R. Chung and A. Rodrigues, Proc. SPIE 4421, pp. 618–621 (2002).
[19] G. Hong and M. R. Luo, "New algorithm for calculating perceived colour difference of images," Imaging Sci. J. **54**, 86–91 (2006).
[20] M. Pedersen and J. Y. Hardeberg, "Rank order and image difference metrics," *Proc. CGIV2008: 4th European Conf. on Colour in Graphics, Imaging, and Vision* (IS&T, Springfield, VA 2008) pp. 120–125.
[21] M. Pedersen, J. Y. Hardeberg, and P. Nussbaum, "Using gaze information to improve image difference metrics," edited by B. Rogowitz and T. Pappas, *Proc. SPIE 6806* p. 680611 (2008).
[22] M. Pedersen, "Importance of region-of-interest on image difference metrics," Master's thesis (Gjøvik University College, 2007).
[23] J. Y. Hardeberg, E. Bando, and M. Pedersen, "Evaluating colour image difference metrics for gamut-mapped images," Coloration Technol. **124**, 243–253 (2008).
[24] E. Bando, J. Y. Hardeberg, and D. Connah, "Can gamut mapping quality be predicted by color image difference formulae?" edited by B. Rogowitz, T. N. Pappas and S. Daly, *Proc. SPIE 5666*, pp. 180–191 (2005).
[25] J. Bai, T. Nakaguchi, N. Tsumura, and Y. Miyake, "Evaluation of image corrected by retinex method based on *S-CIELAB* and gazing information," *IEICE Trans. on Fundamentals of Electronics, Communications and Computer Sciences*, Vol. E89-A (11): 2006, pp. 2955–2961.
[26] N. Bonnier, F. Schmitt, H. Brettel, and S. Berche, "Evaluation of spatial gamut mapping algorithms," *Proc. IS&T/SID Fourteenth Color Imaging Conf.* (IS&T, Springfield, VA 2006) pp. 56–61.
[27] X. Zhang and B. A. Wandell, "Color image fidelity metrics evaluated using image distortion maps," Signal Processing—Special issue on image and video quality metrics **70**, 201–214 (1998).
[28] X. Zhang, D. A. Silverstein, J. E. Farrell, and B. A. Wandell, "Color image quality metric *S-CIELAB* and its application on halftone texture visibility," *COMPCON97 Digest of Papers* (Washington, DC, USA, 1997. IEEE), pp. 44–48.
[29] X.-F. Feng, J. Speigle, and A. Morimoto, "Halftone quality evaluation using color visual models," *Proc. PICS: Image Processing, Image Quality, Image Capture, Systems Conf.* (IS&T, Springfield, VA 2002) pp. 5–10.
[30] Q. Yu, K. J. Parker, R. Buckley, and V. Klassen, "A new metric for color halftone visibility," *Proc. PICS: Image Processing, Image Quality, Image Capture, Systems Conf.* (IS&T, Springfield, VA 1998) pp. 226–230.
[31] J.-S. Kim, M.-S. Cho, S. Westland, and M. R. Luo, "Image quality assessment for photographic images," *Int'l. Colour Assoc.* (Granada, Spain, 2005), pp. 1095–1098.
[32] J.-S. Kim, M.-S. Cho, and B.-K. Koo, "Experimental approach for human perception based image quality assessment," *Int'l. Conf. on Entertainment Computing, Lecture Notes in Computer Science* (Cambridge, UK, 2006, Springer, Berlin/Heidelberg), edited by R. Harper, M. Rauterberg and M. Combetto, Vol. 4161, pp. 59–68.
[33] E. W. Jin and S. Field, "A groundtruth database for testing objective metrics for image difference," *Proc. PICS: Image Processing, Image Quality, Image Capture, Systems Conf.* (IS&T, Springfield, VA 2003), pp. 114–119.
[34] D. W. Hertel, Exploring s-CIELAB as a scanner metric for print uniformity, *Image Quality and System Performance II*, edited by R. Rasmussen and Y. Miyake, *Proc. SPIE 5668*, pp. 51–60 (2005).
[35] S. Bouzit and L. MacDonald, "Colour difference metrics and image sharpness," *Proc. IS&T/SID Eighth Color Imaging Conf.* (IS&T, Springfield, VA, 2000), pp. 262–267.
[36] J. A. Movshon and L. Kiorpes, "Analysis of the development of spatial sensitivity in monkey and human infants," J. Opt. Soc. Am. A **5**, 2166–2172 (1988).
[37] G. Simone, C. Oleari, and I. Farup, "Performance of the euclidean color-difference formula in log-compressed OSA-UCS space applied to modified-image-difference metrics," *11th Congress Int'l. Colour Assoc.* (Sydney, Australia, 2009).
[38] N. Ponomarenko, V. Lukin, K. Egiazarian, J. Astola, M. Carli, and F. Battisti, "Color image database for evaluation of image quality metrics," *Int'l. Workshop on Multimedia Signal Processing* (Cairns, Queensland, Australia, 2008), pp. 403–408.
[39] F. Dugay, I. Farup, and J. Y. Hardeberg, "Perceptual evaluation of color gamut mapping algorithms," Color Res. Appl. **33**, 470–476 (2008).
[40] F. Dugay, "*Perceptual evaluation of colour gamut mapping algorithms*," Master thesis (Gjøvik University College and Grenoble Institute of Technology, 2007).
[41] CIE. "Guidelines for the evaluation of gamut mapping algorithms," Technical Report ISBN: 3-901-906-26-6, CIE TC8-03, 156:2004.
[42] P. Zolliker and K. Simon, "Adding local contrast to global gamut mapping algorithms," *Proc. CGIV2006: 3rd European Conf. on Colour in Graphics, Imaging, and Vision* (IS&T, Springfield, VA 2006).

[43] Ø. Kolås and I. Farup, "Efficient hue-preserving and edge-preserving spatial gamut mapping," *Proc. IS&T/SID Fifteenth Color Imaging Conf.* (IS&T, Springfield, VA 2007).
[44] I. Farup, C. Gatta, and A. Rizzi, "A multiscale framework for spatial gamut mapping," IEEE Trans. Image Process. **16**, 2423–2435 (2007).
[45] S. A. Ajagamelle, "*Analysis of the difference of gaussians model in perceptual image difference metrics*," Master thesis (Gjøvik University College and Grenoble Institute of Technology, 2009).
[46] P. Le Callet and F. Autrusseau, Subjective quality assessment IRCCYN/IVC database, 2005. http://www.irccyn.ec-nantes.fr/ivcdb/.
[47] Z. Wang and J. Y. Hardeberg, "An adaptive bilateral filter for predicting color image difference," *Proc. IS&T/SID Seventeenth Color Imaging Conf.* (IS&T, Springfield, VA 2009), pp. 27–31.
[48] Z. Wang and J. Y. Hardeberg, "Development of an adaptive bilateral filter for evaluating color image difference," J. Electronic Imaging **21**, (2012) 023021–1–023021–10.
[49] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," IEEE Trans. Image Process. **13**, 600–612 (2004).
[50] Z. Wang and A. C. Bovik, "A universal image quality index," IEEE Signal Process. Lett. **9**, 81–84 (2002).
[51] K. Egiazarian, J. Astola, N. Ponomarenko, V. Lukin, F. Battisti, and M. Carli, "Two new full-reference quality metrics based on HVS," *Proc. Second Int'l. Workshop on Video Processing and Quality Metrics* (Scottsdale, Arizona 2006), pp. 22–24.
[52] N. Ponomarenko, F. Silvestri, K. Egiazarian, M. Carli, J. Astola, and V. Lukin, "On between-coefficient contrast masking of DCT basis functions," *Third Int'l. Workshop on Video Processing and Quality Metrics for Consumer Electronics VPQM-07* (Scottsdale, Arizona 2007), pp. 1–4.
[53] D. M. Chandler and S. S. Hemami, "VSNR: A wavelet-based visual signal-to-noise ratio for natural images," IEEE Trans. Image Process. **16**, 2284–2298 (Sep 2007).
[54] H. R. Sheikh and A. C. Bovik, "Image information and visual quality," IEEE Trans. Image Process. **15**, 430–444 (2006).
[55] M. G. Kendall, A. Stuart, and J. K. Ord, *Kendall's Advanced Theory of Statistics: Classical Inference and Relationship*, 5th ed. edited by A. Hodder (Arnold Publication, 1991), Vol. 2.
[56] International Telecommunication Union, "Recommendation ITU-R BT.1676: Methodological framework for specifying accuracy and cross-calibration of video quality metrics," Technical Report, 2004.
[57] 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," Technical Report, 2009.