A New Spatial Hue Angle Metric for Perceptual Image Difference

Marius Pedersen[1,2] and Jon Yngve Hardeberg[1]

[1] Gjøvik University College, Gjøvik, Norway
[2] Océ Print Logic Technologies S.A., Créteil, France

Abstract. Color image difference metrics have been proposed to find differences between an original image and a modified version of it. 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 could therefore miscalculate the difference between an original image and a modified version of it. Because of this we propose a new color image difference metrics based on the *hue angle algorithm* that takes into account the spatial properties of the human visual system. The proposed metric, which we have named *SHAME* (Spatial Hue Angle MEtric), have been subjected to extensive testing. The results show improvement in performance compared to the original metric proposed by Hong and Luo.

1 Introduction

During the last two decades many different color image difference metrics have been proposed, some for overall image quality and some for specific distortions. New and improved metrics are created every year, but so far no one has been able to create an universal color image difference metric.

The CIE published the CIELAB ($L^*a^*b^*$) color space specification [1], with the idea of a perceptually uniform color space. In a color space like this it is straightforward to calculate the distance between two colors, by using the Euclidean distance. This metric is known as ΔE_{ab}^*, and has also been used to calculate the difference between color images by calculating the color difference of all pixels.

A spatial extension to the CIELAB color difference formula (*S-CIELAB*) was proposed by Zhang and Wandell [2], and it introduced a spatial pre-processing to the CIELAB color difference formula [1] by using a spatial filter to simulate the human visual system. The image is 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 image is transformed into CIE-XYZ, and further into CIELAB, where a pixelwise ΔE_{ab}^* is calculated.

The *hue angle algorithm* proposed by Hong and Luo [3], 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. Even though the metric shows good results for two different images [3], it does not include spatial filtering of the image and is therefore unsuitable for halftoned images where the viewing

distance is crucial for the visual impression of artifacts. It has been shown to have problems in calculating perceived image difference [4,5,6]. Due to this we propose a new image difference metric with spatial filtering simulating the human visual system called *SHAME* (**s**patial **h**ue **a**ngle **me**tric).

2   The Proposed Metric

A new color image difference metric is proposed based on the *hue angle algorithm* and two different spatial filtering methods are tested. We give an overview of the *hue angle algorithm*, and then the two spatial filtering methods.

2.1   The Hue Angle Algorithm

Hong and Luo [3] proposed a full-reference color image difference metric built on the CIELAB color difference formula [1]. 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 [3] these are:

- 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 difference 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 angle (h_{ab}) a histogram from the 360 hue angles 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.

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

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

$$CD_{image} = \sum_0^{359} hist[hue] * CD[hue]^2/4.$$

2.2   Spatial Filtering

We propose two different spatial filtering methods for the *hue angle algorithm*. The first spatial filtering is adopted from *S-CIELAB* [2]. The image goes through color space transformations, first the RGB image is transformed into CIEXYZ and further into the opponent color space (O_1,O_2,O_3) [2].

$$O_1 = 0.279X + 0.72Y - 0.107Z$$

$$O_2 = -0.449X + 0.29Y - 0.077Z$$

$$O_3 = 0.086X - 0.59Y + 0.501Z$$

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). Then a spatial filter is applied, where data in each channel is filtered by a 2-dimensional separable spatial kernel:

$$f = k \sum_i w_i E_i$$

where

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

and k_i normalize E_i such that the filter sums to 1. The parameters w_i and σ_i are different for the color planes as seen in Table 1. k is a scale factor, which normalize each color plane so its two-dimensional kernel f sums to one.

Table 1. 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 [2]

Plane	Weights w_i	Spreads 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

The second spatial filtering proposed is adopted from Johnson and Fairchild [7]. By specifying and implementing the spatial filters using contrast sensitivity functions (CSF) in the frequency domain, rather than in the spatial domain as the first spatial filtering, more precise control of the filters is obtained [7] but usually at the cost of computational complexity. The luminance filter is a three parameter exponential function, based on research by Movshon and Kiorpes [8].

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

where $a = 75$, $b = 0.22$, $c = 0.78$ and p is represented as cycles per degree (cpd). The luminance CSF is normalized so that the DC modulation is set to 1.0, resulting in a low pass filter instead of a bandpass filter. This will also enhance any image differences where the human visual system is most sensitive to them [7]. For the chrominance CSF, a sum of two Gaussian functions are used.

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

where different parameters for a_1, a_2, b_1, b_2, c_1 and c_2 have been used as seen in Table 2.

Table 2. The parameters used for the spatial filtering in the frequency domain of the chrominance channels

Parameter	Red-Green	Blue-Yellow
a_1	109.14130	7.032845
b_1	-0.00038	-0.000004
c_1	3.42436	4.258205
a_2	93.59711	40.690950
b_2	-0.00367	-0.103909
c_2	2.16771	1.648658

2.3   Applying Spatial Filtering to the Hue Angle Algorithm

The images are spatially filtered with the previously introduced spatial filtering methods. This results in a filtered original and a filtered modified version of the original, which are used as input to the *hue angle algorithm*, as shown in Figure 1.

The *hue angle algorithm*, filtered respectively with the first and second filter, is from now on referred to as *SHAME-I* and *SHAME-II*. 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
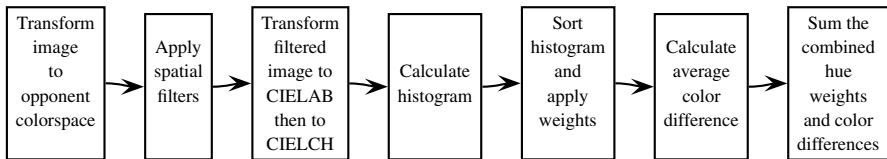- Generates one value for easy interpretation

Fig. 1. Workflow of the proposed metrics

3 Experimental Results and Discussion

Many different image databases have been proposed for evaluation of image difference metrics. For the evaluation we have used one of these databases [9] together with a dataset of gamut mapped images [10,11,4] and a dataset with lightness changed images [5,6]. Three types of correlation are computed for the results, the Pearson product-moment correlation coefficient, the Spearman's rank correlation coefficient and the Kendall tau rank correlation coefficient [12]. The first assumes that the variables are ordinal, and finds the linear relationship between variables. The second, Spearman, is a non-parametric measure of correlation that uses the ranks as basis instead of the actual values. It describes the relationship between variables without making any assumptions about the frequency distribution of the variables. The third, Kendall, is a non-parametric test used to measure the degree of correspondence between two rankings, and assessing the significance of this.

The new metric, with the two different spatial filtering methods, is compared against the original *hue angle algorithm* [3], pixelwise ΔE_{ab}^*, *S-CIELAB* [2] and *S-CIELAB$_{Johnson}$* [7] to see if the segmentation done according to the hue angles improves the performance of the metric. We also compare *SHAME* to *SSIM* [13] and *UIQ* [14], both being state of the art metrics. The evaluation performed will show potential differences between the two proposed spatial filtering methods used in *SHAME*, but also how they perform against other state of the art metrics.

3.1 Evaluation Using the TID2008 Database

The TID2008 database [9] 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 distortions over 4 distortion levels. The mean opinion scores (MOS) 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 32 samples per degree, equal to approximately 60 cm on a normal 17 inch screen.

The *hue angle algorithm* has a low overall correlation for the TID2008 database as seen on Figure 2. When looking at specific distortions the metric does not perform well, the highest Pearson correlation is 0.375 on the Hard dataset containing noise, compression, blurring and transmission errors. This indicates that the *hue angle algorithm* should be improved for the distortions found in the TID2008 database.

SHAME-I shows a better correlation for the full database, with a Pearson correlation of 0.544 (Figure 3). When looking at the specific distortions, Noise, Noise2, Safe, Hard and Simple *SHAME-I* has high correlation coefficients, indicating that it is able to predict perceived image difference. For the Exotic dataset, containing pattern noise, local block-wise distortions of different intensity, mean shift and contrast change, we get low correlation coefficient indicating problems with the metric for these distortions.

When looking at the different distortions *SHAME-I* performs very well for the JPEG and JPEG2000 transmission error, these distortions are a part of exotic2 but not of the exotic dataset. For the groups pattern noise and local block-wise distortions we get a good correlation with the MOS, but not for mean shift and contrast change. In the distortions mean shift and contrast change we have a large difference between the
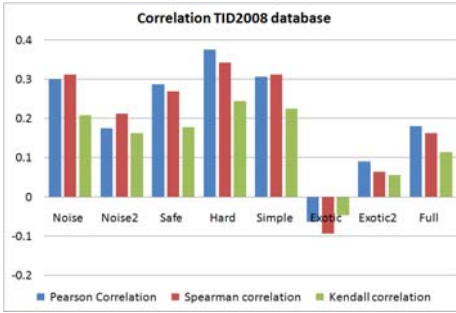
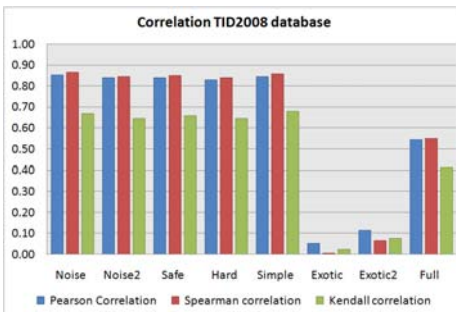| Dataset	Pearson correlation	Spearman correlation	Kendall correlation
Noise	0.299	0.311	0.207
Noise2	0.174	0.212	0.161
Safe	0.286	0.269	0.177
Hard	0.375	0.342	0.243
Simple	0.306	0.312	0.224
Exotic	-0.063	-0.093	-0.046
Exotic2	0.089	0.064	0.056
Full	0.179	0.161	0.113

Fig. 2. Pearson, Spearman and Kendall correlation coefficients for *hue angle algorithm* based on the TID2008 database. The *hue angle algorithm* has a low or medium performance for the different datasets, and low correlation for the full database.

scenes, and due to this a low correlation is found. This indicates that more work is needed for these types of distortions in order to develop better image difference metrics.

The same analysis is valid for *SHAME-II*, but it has a higher correlation for all datasets and for the full database as seen in Figure 4. It should be noted that the improvement in most cases is minimal, even so the general performance indicates that a precise spatial filtering is important for image difference metrics.

The *hue angle algorithm* was proposed to correct some of the drawbacks of the ΔE_{ab}^* color difference formula. When looking at the overall results from the TID2008 database the results for these two metrics are very similar (Figure 5). For this database the extension done in the *hue angle algorithm* does not improve the ΔE_{ab}^*. *SHAME-I* and *SHAME-II* has significantly better correlation than the *hue angle algorithm* and ΔE_{ab}^*.

The *S-CIELAB* has been shown to perform better than the ΔE_{ab}^* [2], and since the same filtering is used for *SHAME-I* the *S-CIELAB* should also be used for comparison. From the results in Figure 5 we can see that both *SHAME-I* and *SHAME-II* perform better than *S-CIELAB* and *S-CIELAB*$_{Johnson}$. This shows that the segmentation done

Dataset	Pearson correlation	Spearman correlation	Kendall correlation
Noise	0.852	0.865	0.669
Noise2	0.840	0.845	0.646
Safe	0.840	0.849	0.658
Hard	0.828	0.839	0.645
Simple	0.844	0.857	0.680
Exotic	0.052	0.006	0.023
Exotic2	0.114	0.065	0.076
Full	0.544	0.550	0.414

Fig. 3. Pearson, Spearman and Kendall correlation coefficients for *SHAME-I* based on the TID2008 database. *SHAME-I* has high correlation coefficients for the datasets, except for Exotic and Exotic2. For the full database it has an average performance.

Correlation TID2008 database
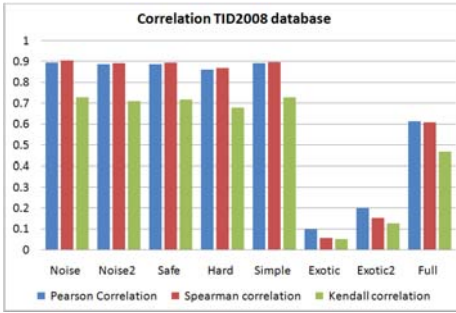
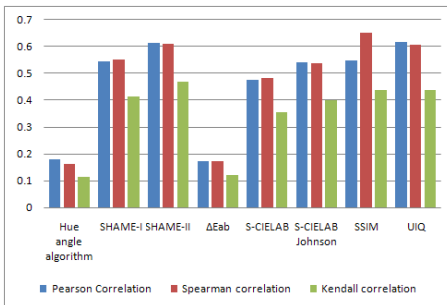| Dataset	Pearson correlation	Spearman correlation	Kendall correlation
Noise	0.893	0.905	0.726
Noise2	0.885	0.891	0.709
Safe	0.887	0.894	0.717
Hard	0.859	0.867	0.678
Simple	0.891	0.895	0.726
Exotic	0.098	0.057	0.053
Exotic2	0.199	0.152	0.126
Full	0.613	0.609	0.468

Fig. 4. Pearson, Spearman and Kendall correlation coefficients for *SHAME-II* based on the TID2008 database. *SHAME-II* gets high correlation coefficients for the datasets, except for exotic and exotic2. For the full database *SHAME-II* has an average performance.

according to the hue angle improves the metric when the images are spatially filtered. This also supports the fact that the whole image is not important when judging image difference, but that some areas are more important than others [5,6].

3.2 Evaluation Using Gamut Mapped Images

The TID2008 database contains only one distortion for each image, in order to test the metrics extensively we have used a dataset with gamut mapped images from Dugay [10,11]. 20 different images were gamut mapped with 5 different algorithms. The 20 different images were evaluated by 20 observers in a pair comparison experiment. This is a more complex task for the observers since many artifacts must be considered, and also a demanding task for the image difference metrics.

In Figure 6 shows the results from the dataset with gamut mapped images. In general all metrics have a low performance. This was probably because the task is very

Metric	Pearson correlation	Spearman correlation	Kendall correlation
Hue angle	0.179	0.161	0.113
SHAME-I	0.544	0.550	0.414
SHAME-II	0.613	0.609	0.468
ΔE_{ab}^{*}	0.174	0.173	0.121
S-CIELAB	0.476	0.482	0.354
S-CIELAB$_{Johnson}$	0.542	0.538	0.400
SSIM	0.547	0.653	0.437
UIQ	0.616	0.606	0.438

Fig. 5. Comparison of all tested image quality metrics. We can see that *SHAME-I* and *SHAME-II* clearly perform better than the *hue angle algorithm*, and that they perform similar to *SSIM* and *UIQ*. It is also interesting to see how the new metric with the two spatial filtering methods perform compared to the *S-CIELAB* and the improved *S-CIELAB$_{Johnson}$*, from the Figure we can see that *SHAME-I* and *SHAME-II* have better correlation than these.
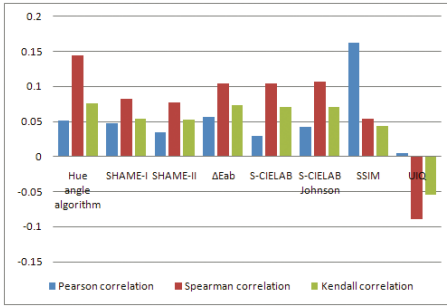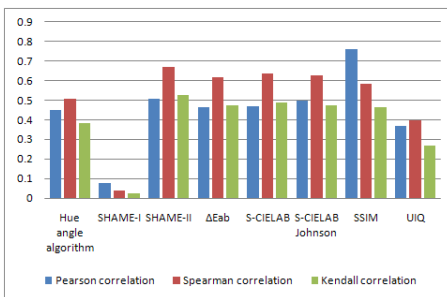
Metric	Pearson correlation	Spearman correlation	Kendall correlation
Hue angle	0.052	0.114	0.076
SHAME-I	0.047	0.082	0.054
SHAME-II	0.035	0.077	0.053
S-CIELAB	0.056	0.105	0.073
S-CIELAB$_{Johnson}$	0.029	0.104	0.071
ΔE_{ab}^{*}	0.042	0.107	0.071
SSIM	0.163	0.054	0.044
UIQ	0.005	-0.089	-0.055

Fig. 6. *SHAME-I* and *SHAME-II* compared against other metrics for a set of gamut mapped images. All metrics have a low performance on the gamut mapped images, indicating that the calculating the difference between an original and a gamut mapped image is very difficult for image difference metrics.

complex, in gamut mapping multiple artifacts can occur and the observers may judge them differently [10,11]. Previous research has shown that image difference metrics have problems when multiple distortions occur simultaneously, as in gamut mapping [15,16]. This is not the case for TID2008 since only one artifact at the time occur in the images.

3.3   Evaluation Using Luminance Changed Images

The last dataset used for the evaluation has previously been used by Pedersen [6] and Pedersen et al.[5], where four images where modified in lightness, both globally and locally, resulting in 32 reproductions. This dataset differs from the previous due to the controlled changes only in lightness, and this should be easier for the metrics to judge than the gamut mapped images.

SHAME-II has a higher correlation than *SHAME-I* and the *hue angle algorithm*, indicating that spatial filtering done in *SHAME-II* improves the *hue angle algorithm*.

Metric	Pearson correlation	Spearman correlation	Kendall correlation
Hue angle	0.452	0.507	0.383
SHAME-I	0.078	0.036	0.024
SHAME-II	0.509	0.670	0.528
ΔE_{ab}^{*}	0.464	0.618	0.472
S-CIELAB	0.467	0.637	0.488
S-CIELAB$_{Johnson}$	0.500	0.629	0.472
SSIM	0.762	0.586	0.464
UIQ	0.370	0.396	0.270

Fig. 7. *SHAME-I* and *SHAME-II* compared against other metrics for the lightness changed image from [5,6]. We notice that *SHAME-II* outperforms *SHAME-I*, but only a minor improvement over the *hue angle algorithm*. The *SSIM* is better than *SHAME-II* for the Pearson correlation, but *SHAME-II* is better for Spearman and Kendall, indicating that the ranking by *SHAME-II* is more correct than the ranking by *SSIM*.

SHAME-I does not have the same high correlation, and is clearly worse than the rest. When analyzing the results we can see that the *SHAME-I* metric miscalculated images that had a low mean luminance compared to images with high mean luminance. We can also notice that *SHAME-II* has a higher Spearman and Kendall correlation than *SSIM*, but a lower Pearson. This indicates that the ranking done by *SHAME-II* is more correct than the ranking by *SSIM*, but that *SSIM* has a more correct frequency distribution. The results indicate that the more precise spatial filtering and the bandpass nature of the filter in *SHAME-II* is important for the performance of the metric, therefore the filtering in *SHAME-II* should be preferred over *SHAME-I*.

4   Conclusion and Further Research

The proposed metric, *SHAME*, use well-known spatial filtering methods to improve a color image difference metric, which results in several advantages. Extensive testing of the proposed metrics show an improvement over the traditional metrics, as pixelwise ΔE_{ab}^* and *S-CIELAB*. We have demonstrated the importance of weighting areas of interest and the importance of spatial filtering for color image difference metrics. The results indicate that precise control of the spatial filters will improve the performance of the metric, and therefore *SHAME-II* gives an advantage over *SHAME-I*.

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, more research should be carried out to improve the metrics in this field, both in terms of difference calculation and spatial filtering.

Acknowledgments

References

1. CIE: Colorimetry. Technical Report 15 (2004)
2. Zhang, X., Wandell, B.: A spatial extension of CIELAB for digital color image reproduction. In: Soc. Inform. Display 96, San Diego, 731–734 (1996), http://white.stanford.edu/~brian/scielab/scielab.html
3. Hong, G., Luo, M.: Perceptually based colour difference for complex images. In: Chung, R., Rodrigues, A. (eds.) 9th Congress of the International Colour Association. Proceedings of SPIE, vol. 4421, pp. 618–621 (2002)
4. Pedersen, M., Hardeberg, J.Y.: Rank order and image difference metrics. In: CGIV 2008 Fourth European Conference on Color in Graphics, Imaging and Vision, Terrassa, Spain, IS&T, pp. 120–125 (June 2008)

 5. Pedersen, M., Hardeberg, J.Y., Nussbaum, P.: Using gaze information to improve image difference metrics. In: Rogowitz, B., Pappas, T. (eds.) Human Vision and Electronic Imaging VIII (HVEI 2008), San Jose, USA. SPIE proceedings, vol. 6806. SPIE (January 2008)
 6. Pedersen, M.: Importance of region-of-interest on image difference metrics. Master's thesis, Gjøvik University College (2007)
 7. Johnson, G.M., Fairchild, M.D.: Darwinism of color image difference models. In: The 9th Color Imaging Conference: Color Science and Engineering: Systems, Technologies, Applications, pp. 108–112 (2001)
 8. Movshon, J.A., Kiorpes, L.: Analysis of the development of spatial sensitivity in monkey and human infants. J. Opt. Soc. Am. A 5, 2166–2172 (1988)
 9. Ponomarenko, N., Lukin, V., Egiazarian, K., Astola, J., Carli, M., Battisti, F.: Color image database for evaluation of image quality metrics. In: International Workshop on Multimedia Signal Processing, Cairns, Queensland, Australia (October 2008), http://www.ponomarenko.info/tid2008.htm
10. Dugay, F.: Perceptual evaluation of colour gamut mapping algorithms. Master thesis, Gjøvik University College and Grenoble Institute of Technology (2007)
11. Dugay, F., Farup, I., Hardeberg, J.Y.: Perceptual evaluation of colour gamut mapping algorithms. Color Research & Application 33(6), 470–476 (2008)
12. Kendall, M.G., Stuart, A., Ord, J.K.: Kendall's Advanced Theory of Statistics: Classical inference and relationship, 5th edn., vol. 2. A Hodder Arnold Publication (1991)
13. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: from error visibility to structural similarity. IEEE Transactions on Image Processing 13(4), 600–612 (2004)
14. Wang, Z., Bovik, A.: A universal image quality index. IEEE Signal Processing Letters 9, 81–84 (2002)
15. Hardeberg, J.Y., Bando, E., Pedersen, M.: Evaluating colour image difference metrics for gamut-mapped images. Coloration Technology 124(4), 243–253 (2008)
16. Bonnier, N., Schmitt, F., Brettel, H., Berche, S.: Evaluation of spatial gamut mapping algorithms. In: 14th Color Imaging Conference. IS&T/SID, vol. 14, pp. 56–61 (2006)