Foundations and Trends® in
Computer Graphics and Vision
Vol. 7, No. 1 (2011) 1–80

DOI: 10.1561/0600000037

now
the essence of knowledge

Full-Reference Image Quality Metrics: Classification and Evaluation

By Marius Pedersen and Jon Yngve Hardeberg

Contents

1 Introduction **2**

2 Classification of Image Quality Metrics **4**

2.1 Existing Classification of Image Quality Metrics 4
2.2 Proposal for Classification of Image Quality Metrics 6

3 Survey of Image Quality Metrics **8**

3.1 Mathematically Based Metrics 9
3.2 Low-level Metrics 14
3.3 High-level Metrics 20
3.4 Other Approaches 25

4 Evaluation of Image Quality Metrics **29**

4.1 TID2008 Database, Ponomarenko et al. 32
4.2 Luminance Changed Images, Pedersen et al. 38
4.3 JPEG and JPEG2000 Compressed Images, Caracciolo 40
4.4 IVC Database, Le Callet et al. 43
4.5 Images Altered in Contrast, Lightness, and Saturation, Ajagamelle et al. 45

4.6 Gamut Mapped Images, Dugay et al. 47
4.7 Overall Performance of the Metrics 50

5 Conclusion 51

A Image Quality Metric Overview 52

Acknowledgements 65

References 66

Foundations and Trends® in
Computer Graphics and Vision
Vol. 7, No. 1 (2011) 1–80

DOI: 10.1561/0600000037

now
the essence of knowledge

Full-Reference Image Quality Metrics: Classification and Evaluation

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

[1] *Gjøvik University College, Norwegian Color Research Laboratory, P.O. Box 191, N-2802 Gjøvik, Norway, marius.pedersen@hig.no*
[2] *Gjøvik University College, Norwegian Color Research Laboratory, P.O. Box 191, N-2802 Gjøvik, Norway, jon.hardeberg@hig.no*

Abstract

The wide variety of distortions that images are subject to during acquisition, processing, storage, and reproduction can degrade their perceived quality. Since subjective evaluation is time-consuming, expensive, and resource-intensive, objective methods of evaluation have been proposed. One type of these methods, image quality (IQ) metrics, have become very popular and new metrics are proposed continuously. This paper aims to give a survey of one class of metrics, full-reference IQ metrics. First, these IQ metrics were classified into different groups. Second, further IQ metrics from each group were selected and evaluated against six state-of-the-art IQ databases.

1

Introduction

Advances are rapid in the imaging industry, and new and more advanced products are continuously being introduced into the market. In order to verify that the new technologies produce higher-quality images than the current technology, some kind of quality assessment is required.

There are two main methods of assessing image quality (IQ): subjective or objective. The first is carried out by human observers, while the second does not involve observers. To make an objective assessment, one can use measuring devices to obtain numerical values; another method is to use IQ metrics. These IQ metrics are usually developed to take into account the human visual system (HVS), and thus have the goal of correlating with subjective assessment.

Many IQ metrics have been proposed in the literature; a brief summary of more than 100 metrics was given by Pedersen and Hardeberg [125]. These metrics stem from different ideas, and they have been made for different purposes, such as to quantify distortions, produce benchmarks, monitor quality, optimize a process, or indicate problem areas. Because different metrics have different goals, it is important to keep in mind their areas of use when evaluating their performance.

Existing surveys, such as the one by Wang and Bovik [166], mostly focus on grayscale IQ metrics whereas the survey by Avcibas et al. [6] covers only simple statistical metrics. In this survey, we continue the work started in Pedersen and Hardeberg [125] and we carry out a comprehensive survey and evaluation of color and grayscale IQ metrics.

Our goal is to classify IQ metrics into separate groups and to evaluate their correspondence with the percept. Such a classification can be used to select the most appropriate IQ metric for a given problem or distortion. It will also provide a better understanding of the state-of-the-art of IQ metrics, which can be used to improve or develop new metrics that correlate better with the percept. Because the original is available in many situations, we limit our survey to full-reference IQ metrics, where both the complete original and reproduction are used for the calculation of quality. Additionally, more work has been carried out on full-reference IQ metrics than on reduced-reference and no-reference metrics. The two latter types of IQ metrics are also considered to be more difficult to assess than full-reference metrics [170].

In the literature many different terms have been used, such as IQ, image difference, image fidelity, and image similarity. If possible we will use the general term IQ. In addition, we use the term metric as a general term even though not all metrics fulfill the requirements to be a metric in mathematical terms. Other terms, such as index, measure, and criterion, have also been used in the literature.

The survey is organized as follows: first, we classify IQ metrics into groups; then we go through IQ metrics within each group. This is followed by evaluations of selected IQ metrics from each group. Finally, we report our conclusions.

2

Classification of Image Quality Metrics

2.1 Existing Classification of Image Quality Metrics

Since IQ metrics are based on different approaches, they can be divided into different groups. These groups usually reflect different aspects of the metrics, such as their intended use or construction. Several different researchers have classified metrics into groups, even though it can be difficult to find sharp boundaries between the numerous IQ metrics in the literature.

Avcibas et al. [6] divided IQ metrics into six groups based on the information they use:

- Pixel difference-based measures such as mean square distortion.
- Correlation-based measures: correlation of pixels or vector angular directions.
- Edge-based measures: displacement of edge positions or their consistency across resolution levels.
- Spectral distance-based measures: the Fourier magnitude and/or phase spectral discrepancy on a block basis.

- Context-based measures, which are based on various functionals of the multidimensional context probability.
- HVS-based measures, which are either based on the HVS-weighted spectral distortion measures or (dis)similarity criteria used in image-based browsing functions.

Callet and Barba [89] divided IQ metrics into two distinct groups:

- Those that use a HVS model for low-level perception, such as subband decomposition and masking effects.
- Those that use little information about the HVS for error representation and push the effort on prior knowledge about introduced distortion.

The authors comment that metrics belonging to the last group fail to be robust since they are specialized.

Wang and Bovik [166] classified IQ metrics based on three criteria:

- Full-reference, no-reference, and reduced reference metrics.
- General-purpose and application-specific metrics.
- Bottom-up and top-down metrics.

These categories are based on information about the original image and the distortion process, and knowledge of the HVS. The authors state that it is not possible to draw sharp boundaries between the different groups, but that a crisp classification can be helpful.

Chandler and Hemami [28] divided IQ metrics into three different groups:

- Mathematically convenient metrics, which only operate on the intensity of the distortion.
- Metrics based on near-threshold psychophysics. These metrics take into account the visual detectability of the distortions.
- Metrics based on overarching principles such as structural or information extraction.

Thung and Raveendran [160] divided full-reference IQ metrics into three groups, similarly to Chandler and Hemami [28]:

- Mathematical metrics.
- HVS-based metrics.
- Others.

Recently, Seshadrinathan and Bovik [143] divided techniques for image and video quality into three main groups:

- HVS-based approaches.
- Structural approaches.
- Information theoretic approaches.

Seshadrinathan and Bovik also mention that IQ assessment can be broadly classified as bottom-up and top-down approaches, where bottom-up approaches model the HVS and top-down approaches characterize specific features of the HVS.

2.2 Proposal for Classification of Image Quality Metrics

As seen above, there are many different ways to group IQ metrics. In order to present the various approaches, we have divided the IQ metrics into four groups:

- Mathematically based metrics, which operate only on the intensity of the distortions. These metrics are usually simple. Examples include the mean squared error (MSE) and peak signal-to-noise ratio (PSNR).
- Low-level metrics, which take into account the visibility of the distortions using, for example, contrast sensitivity functions (CSFs), such as the spatial-CIELAB (S-CIELAB) [189].
- High-level metrics, which measure quality based on the idea that our HVS is adapted to extract information or structures from the image. The structural similarity (SSIM) [167], which is based on structural content, or the visual image fidelity (VIF) [145], which is based on scene statistics, are examples of metrics in this group.

- Other metrics, which are based on other strategies or combine two or more of the above groups. One example is the visual signal-to-noise ratio (VSNR) [28], which takes into account both low- and mid-level visual properties, and the final stage incorporates a mathematically based metric.

The three first groups are similar to the classification by Chandler and Hemami [28], and the first, second, and fourth are similar to the classification of Thung and Raveendran [160].

3

Survey of Image Quality Metrics

For each of the four groups we will present a number of selected metrics. Since an impressive number of IQ metrics have been proposed [121, 125], metrics will be selected based on compliance with standards, existing literature, frequency of use, and for new metrics, degree of promise.

In addition to introducing the metrics, we will identify differences between them. For this purpose, a test target has been adopted. The target (Figure 3.1) was developed by Halonen et al. [62] to evaluate print quality. In this test target ten specific color patches from the GretagMacbeth ColorChecker were incorporated as objects in the scene; blue, green, red, yellow, magenta, cyan, orange, and three neutral grays. This target should show differences between the metrics, since it contains different frequencies, colors, and objects. Since we are applying full-reference IQ metrics, a reproduction is needed. This has been obtained by taking the original image, applying an ICC profile (EuroScale Uncoated V2) with the relative colorimetric rendering intent, and then converting the image back into RGB, using Adobe Photoshop CS3 for all procedures. This should result in many different quality changes, which will provoke differences in the metrics.

(a) Original. (b) Modified.

Fig. 3.1 Test target from Halonen et al. [62] used to differentiate IQ metrics. The original image has been processed by applying an ICC profile with the relative colorimetric rendering intent to obtain a reproduction.

3.1 Mathematically Based Metrics

The first group of metrics, mathematically based ones, have been very popular because of their easy implementation, and they are convenient for optimization. These metrics usually work only on the intensity of the distortion E:

$$E(x,y) = I_{\mathrm{O}}(x,y) - I_{\mathrm{R}}(x,y), \tag{3.1}$$

where I_{O} is the original image, I_{R} is the reproduction, x and y indicate the pixel position.

Most of the metrics in this group are strict metrics — $(\rho(x,y))$ is essentially an abstract distance — and have the following properties: $\rho(x,y) = 0$ if $x = y$, symmetry, triangle inequality, and non-negativity.

3.1.1 Mean Squared Error

Mean squared error (MSE) is one of the mathematically based metrics. It calculates the cumulative squared error between the original image and the distorted image. MSE is given as

$$\mathrm{MSE} = \frac{1}{MN} \sum_{y=0}^{M-1} \sum_{x=0}^{N-1} [E(x,y)]^2, \tag{3.2}$$

where x and y indicate the pixel position, and M and N are the image width and height. Several other metrics of this type have been proposed as well, such as the PSNR (Section 3.1.2) and root mean square

(RMS). The MSE and other metrics have been used because of their ease of calculation and analytical tractability [159]. Although these simple mathematical models usually do not correlate closely with perceived IQ [28, 166], they have been influential to other more advanced metrics, such as the PSNR-HVS-M [132], which takes into account contrast masking and CSFs to model the HVS.

3.1.2 Peak Signal-to-Noise Ratio

The peak signal-to-noise ratio (PSNR) is a measure of the peak error between the compressed image and the original image. PSNR is given as

$$\mathrm{PSNR} = 20 \cdot \log 10 \left(\frac{\mathrm{MAX}_I}{\sqrt{\mathrm{MSE}}} \right), \tag{3.3}$$

where MAX_I is the maximum possible value of the image. The higher the PSNR, the better the quality of the reproduction. PSNR has usually been used to measure the quality of a compressed or distorted images.

3.1.3 ΔE^*_{ab}

Metrics that measure color also belong to the group of mathematically based metrics. One of the most commonly used color difference formulas stems from the CIELAB color space specification published by CIE [36], based on the idea of a perceptually uniform color space. In a color space like this it is easy to calculate the distance between two colors by using the Euclidean distance. This value, known in engineering as the RMS value, is the square root of the MSE. As the ΔE^*_{ab} is a color difference formula it deals with three dimensions. As input it takes a sample color with CIELAB values L^*_s, a^*_s, b^*_s and a reference color L^*_r, a^*_r, b^*_r. The distance between the two colors is given by

$$\Delta E^*_{ab} = \sqrt{(\Delta L^*)^2 + (\Delta a^*)^2 + (\Delta b^*)^2}, \tag{3.4}$$

where $\Delta L^* = L^*_s - L^*_r$, $\Delta a^* = a^*_s - a^*_r$, and $\Delta b^* = b^*_s - b^*_r$.

The CIELAB formula has served as a satisfactory tool for measuring perceptual difference between uniform color patches. However, one

problem is that although the HVS is not as sensitive to color differences in fine details as in large patches, the CIELAB color formula will predict the same visual difference in these two cases, since it does not have a spatial variable [190]. Because of the popularity of the ΔE^*_{ab} it has also been commonly used to compare two natural images by calculating the color difference of each pixel of the image. The mean of these differences is usually the overall indicator of the difference between the original and the reproduction:

$$\overline{\Delta E^*_{ab}} = \frac{1}{MN} \sum_{x=0}^{m-1} \sum_{y=0}^{n-1} \Delta E^*_{ab(x,y)}. \tag{3.5}$$

Other measures of the ΔE^*_{ab} can be the minimum or the maximum value in the computed difference. Because of its widespread use [10, 16, 33, 63, 77, 78, 120, 123, 127, 139, 156, 169, 188, 190], its position in the industry, and because it is often used a reference metric, we consider that the ΔE^*_{ab} should be included in an overview of IQ metrics. Extensions of the ΔE^*_{ab} have been proposed when it became apparent that it had problems, especially in the blue regions. First the CIE proposed the ΔE^*_{94} [34] (Section 3.1.4) and later the ΔE^*_{00} [95] (Section 3.1.5). These are increasingly complex compared to the ΔE^*_{ab}, and therefore many still use the ΔE^*_{ab}.

3.1.4 ΔE^*_{94}

The CIE ΔE_{94} [34] was developed after it became clear that the CIELAB ΔE^*_{ab} did not correlate well with the perceptual color difference. The formula, based on CIE lightness ΔL^*, chroma ΔC^*, and hue ΔH^* differences, is as follows:

$$\Delta E^*_{94} = \sqrt{\left(\frac{\Delta L^*}{k_L S_L}\right)^2 + \left(\frac{\Delta C^*}{k_C S_C}\right)^2 + \left(\frac{\Delta H^*}{k_H S_H}\right)^2}, \tag{3.6}$$

where k_L, k_C, k_H are scaling parameters, S_L, S_C, and S_H are lightness, chroma, and hue scaling functions [144]. ΔL^*, ΔC^*, and ΔH^* refer to lightness, chroma, and hue differences.

3.1.5 ΔE^*_{00}

The CIE ΔE_{00} [95, 35] was published because of the same problems as CIE ΔE_{94} [144]. CIE ΔE_{00} is calculated as

$$\Delta E^*_{00} = \sqrt{\left(\frac{\Delta L^*}{k_L S_L}\right)^2 + \left(\frac{\Delta C^*}{k_C S_C}\right)^2 + \left(\frac{\Delta H^*}{k_H S_H}\right)^2 + R_T \phi(\Delta C^* \Delta H^*)}, \tag{3.7}$$

where k_L, k_C, k_H, S_L, S_C, and S_H are the same as in ΔE_{94} and R_T is an additional scaling function depending on chroma and hue [144]. The ΔL^*, ΔC^*, and ΔH^* parameters are differences in lightness, chroma, and hue.

3.1.6 ΔE_E

Another very interesting color difference formula was proposed by Oleari et al. [116] for small and medium color differences in the log-compressed OSA-UCS space. Because of its different origin and promising results [154, 155] we include it in this overview of IQ metrics. The formula is defined as follows:

$$\Delta E_E = \sqrt{(\Delta L_E)^2 + (\Delta G_E)^2 + (\Delta J_E)^2}, \tag{3.8}$$

where

$$L_E = \frac{1}{b_L} \ln\left[1 + \frac{b_L}{a_L}(10 L_{\mathrm{OSA}})\right], \tag{3.9}$$

with $a_L = 2.890$, $b_L = 0.015$, and

$$G_E = -C_E \cos(h), \tag{3.10}$$

$$J_E = C_E \sin(h), \tag{3.11}$$

with

$$h = \arctan\left(-\frac{J}{G}\right), \tag{3.12}$$

$$C_E = \frac{1}{b_C} \ln\left[1 + \frac{b_C}{a_C}(10 C_{\mathrm{OSA}})\right], \tag{3.13}$$

with $a_C = 1.256$, $b_C = 0.050$, and

$$C_{\text{OSA}} = \sqrt{G^2 + J^2}. \tag{3.14}$$

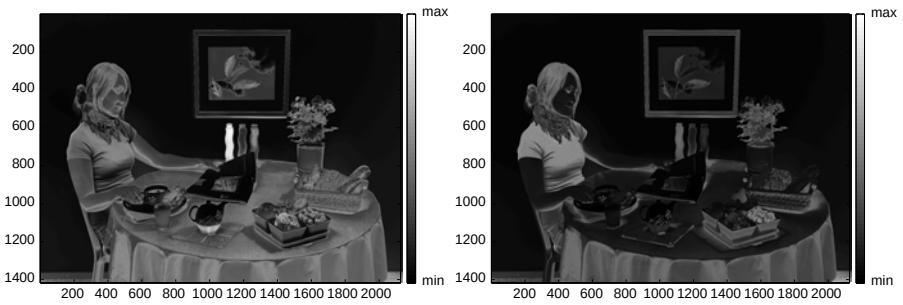
The ΔE_E formula is statistically equivalent to ΔE^*_{00} in the prediction of many available empirical datasets. However, ΔE_E is the simplest formula that provides relationships with visual processing. These analyses hold true for the CIE 1964 Supplementary Standard Observer and D65 illuminant. This formula can also be used to measure the color difference of natural images by taking the average color differences over the entire image, as with the ΔE^*_{ab} metric.

3.1.7 Difference Between the Metrics

The metrics introduced above are based on different ideas and have different origins. Because of this they will work in different ways. In order to understand the fundamental differences between the metrics, we will investigate their similarities and differences.

Since MSE and PSNR have been shown not to correlate well with perceived IQ we do not consider them in this analysis. Instead we will compare the ΔE^*_{ab} and ΔE_E metrics. Both of these metrics are based on the Euclidean distance between two color values; the difference is in the color space. Both color spaces are opponent color spaces, where both have a lightness, a red-green axis, and a blue-yellow axis. The OSA-UCS color space has a better blue linearity than the CIELAB color space [104]. The OSA-UCS system also has the unique advantage of equal perceptual spacing among the color samples [99].

Both the ΔE^*_{ab} and ΔE_E have been computed for the test target, and the result can be seen in Figure 3.2. Both results have been normalized with the maximum value, making it easier to compare the results. The first apparent difference is the difference in the blue vase. This is due to the issues in the blue in the CIELAB color space, which has been reported by others [21]. It is also interesting to notice the differences in the dark regions: these could be because the OSA-UCS color space, which ΔE_E is based on, compensate for the reference background lightness [119]. There is also a difference in the skin tones: ΔE_E finds almost no difference between the original and the reproduction, while ΔE^*_{ab} indicates a larger difference.



(a) ΔE^*_{ab} color difference map. (b) ΔE_E color difference map.

Fig. 3.2 Comparison of color difference formulas. The difference is calculated between an original image (Figure 3.1(a)) and a reproduction (Figure 3.1(b)). White indicates the highest difference between the original and the modified version of the test target, while black indicates a low difference.

It is important to state that we are not evaluating the correctness of the differences or whether they are correlated with perceived difference or quality at this point; we are merely showing the difference between the metrics. Evaluation of the IQ metrics can be found in Section 4.

3.2 Low-level Metrics

Metrics classified as low-level metrics simulate the low-level features of the HVS, such as CSFs or masking. However, most of these metrics use a mathematically based metric, for example, one of the metrics introduced above, in the final stage to calculate quality.

3.2.1 S-CIELAB

When it became apparent that the ΔE^*_{ab} was not correlated with perceived image difference for natural images, Zhang and Wandell [189] proposed a spatial extension of the formula (Figure 3.3). This metric 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 goes through color space transformations. The RGB image is transformed first into CIEXYZ, and then further into the

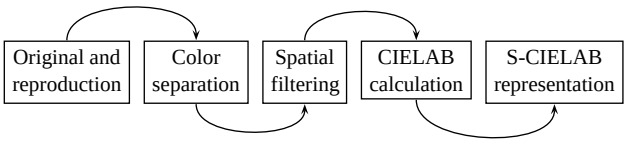


Fig. 3.3 S-CIELAB workflow. The spatial filtering is done in the opponent color space, and the color difference is calculated using ΔE^*_{ab}.

opponent color space (O_1, O_2, and O_3):

$$\begin{aligned} O_1 &= 0.279X + 0.72Y - 0.107Z \\ O_2 &= -0.449X + 0.29Y - 0.077Z \\ O_3 &= 0.086X - 0.59Y + 0.501Z \end{aligned}$$

Now the image contains one channel with the luminance information (O_1), one with the red–green information (O_2), and one with the blue–yellow information (O_3). Then a spatial filter is applied, where the data in each channel is filtered by a two-dimensional separable spatial kernel:

$$f = k \sum_i w_i E_i, \tag{3.15}$$

where

$$E_i = k_i e^{[-(x^2+y^2)/\sigma_i^2]}, \tag{3.16}$$

and k_i normalize E_i such that the filter sums to one. The parameters w_i and σ_i are different for the color planes as seen in Table 3.1. The scale factor k normalizes each color plane so that its two-dimensional kernel f sums to 1.

Finally the filtered image is transformed into CIEXYZ, and this representation is further transformed into the CIELAB color space where the ΔE^*_{ab} is applied to calculate color differences. The final result, an S-CIELAB[1] representation, consists of a color difference for each pixel. These values are usually averaged to give one value for the overall image difference.

[1] 16/02/2011: The implementation of S-CIELAB described here is available at `http://white.standford.edu/~brian/scielab/scielab.html`

Table 3.1. The parameters used for the spatial filtering in S-CIELAB, where w_i is the weight of the plane and σ_i is the spread in degrees of visual angle.

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 S-CIELAB metric was originally designed for halftoned images, but it has also been used to calculate overall IQ for a wide variety of distortions. It is often used as a reference metric because of its simple implementation and because it formed a new direction of IQ metrics. It has also been extensively used and evaluated by other researchers [1, 2, 9, 10, 16, 18, 53, 63, 65, 73, 77, 78, 120, 123, 126, 127, 155, 184, 188, 190]. Additionally, it has led to many other IQ metrics, such as S_{DOG}-CIELAB [1, 2] (Section 3.4.4), where the spatial filtering was performed by calculating the difference-of-Gaussians (DOG), and S-CIELAB$_J$ [75] (Section 3.2.1.1), where the CSFs were modified by adjusting the peak sensitivity to correlate better with complex stimuli.

3.2.1.1 S-CIELAB$_J$

Johnson and Fairchild [74] describe an extension to the S-CIELAB metric, focusing on the CSFs. They propose to filter the image in the frequency domain, rather than in the spatial domain as in S-CIELAB. They also propose changes to the CSFs. The luminance filter is a three-parameter exponential function, based on research by Movshon and Kiorpes [108]. The luminance CSF is calculated as

$$\mathrm{CSF}_{\mathrm{lum}}(p) = a \cdot p^c \cdot e^{-b \cdot p}, \tag{3.17}$$

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. This minimizes the luminance shift and will also

Table 3.2. The parameters used in S-CIELAB$_J$ 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

enhance any image differences where the HVS is most sensitive to them [74].

A sum of two Gaussian functions is used to form the chrominance CSFs:

$$\mathrm{CSF}_{\mathrm{chroma}}(p) = a_1 \cdot e^{-b_1 \cdot p^{c_1}} + a_2 \cdot e^{-b_2 \cdot p^{c_2}}, \tag{3.18}$$

where different parameters for a_1, a_2, b_1, b_2, c_1, and c_2 have been used as seen in Table 3.2. The filters are applied in the $O_1O_2O_3$ color space as for S-CIELAB.

The authors also propose to account for local and global contrast by modifying a local color correction technique that generates gamma curves for each of the opponent channels. The gamma curves are based on low-frequency information, but also on global contrast. They also propose to account for localized attention, which is based on sensitivity to the position of edges. A simple convolution, with a Sobel kernel, is proposed to filter the image.

3.2.2 S-DEE

The ΔE_E metric also has a spatial extension: Spatial-DEE (S-DEE), proposed by Simone et al. [155]. This metric follows the S-CIELAB framework, but the color difference is calculated with ΔE_E instead of ΔE^*_{ab}. The spatial filtering with CSFs is different from the original S-CIELAB, since the S-DEE incorporates CSFs from Johnson and Fairchild [75]. The final step consists of applying the ΔE_E rather than the ΔE^*_{ab} to get the color differences. The overall image difference is achieved by averaging the results over the image.

The ΔE_E metric has promising results, so has the S-DEE [155], and because it challenges the standard calculation of the CIELAB color difference it is included in this survey of IQ metrics. A variant of the S-DEE, S_{DOG}-DEE, was proposed by Ajagamelle et al. [1, 2] (Section 3.4.4). In this version, the spatial filter is carried out by calculating the DOG.

3.2.3 Adaptive Image Difference

Wang and Hardeberg [168] proposed an adaptive bilateral filter (ABF) for image difference metrics. The filter blurs the image but preserves edges, which is not the case when using CSFs. The bilateral filter is defined as

$$h(x) = k^{-1}(x) \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(\epsilon) c(\epsilon, x) s(f(\epsilon), f(x)) d\epsilon, \tag{3.19}$$

where

$$k(x) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} c(\epsilon, x) s(f(\epsilon), f(x)) d\epsilon, \tag{3.20}$$

and where the function $c(\epsilon, x)$ measures the geometric closeness between the neighborhood center x and a nearby point ϵ:

$$c(\epsilon, x) = \exp\left(-\frac{(\epsilon - x)^2}{2\sigma_d^2}\right). \tag{3.21}$$

The function $s(\epsilon, x)$ measures the photometric similarity between the neighborhood center x and a nearby point ϵ:

$$s(\epsilon, x) = \exp\left(-\frac{(f(\epsilon) - f(x))^2}{2\sigma_r^2}\right). \tag{3.22}$$

The geometric spread, σ_d, is determined by the viewing conditions in pixels per degree:

$$\sigma_d = \frac{n/2}{180/\pi \cdot \tan^{-1}(l/(2m))}, \tag{3.23}$$

where n is the width of the image, l is the physical length in meters, and m is the viewing distance in meter. The range spread, σ_r, is determined with image entropy:

$$\sigma_r = K/E, \tag{3.24}$$

where K is a constant to rescale the entropy into an optimized value and

$$E = -\sum_i p_i \log(p_i), \tag{3.25}$$

where p_i is the histogram of the pixel intensity values of an image.

The filtering is performed in the CIELAB color space, and the color difference formula used was ΔE^*_{ab}. Because of this, the metric can be said to follow the same overlaying framework as S-CIELAB.

Because the new type of filtering preserves edges, which has shown to produce good results for, among others, gamut-mapping [15], this metric is interesting to compare with others. It shows promising results compared to observers as well [168].

3.2.4 PSNR-HVS-M

Ponomarenko et al. [132] proposed a block-based IQ metric based on PSNR and local contrast. The image is divided into 8×8-pixel non-overlapping blocks. The metric uses the same formula as the PSNR, but computes the MSE in the discrete cosine transform (DCT) domain. This is done by taking the difference between the DCT coefficients of each 8×8 block in the original image and the DCT coefficients of the corresponding 8×8 block in the distorted image. Then the difference between the original and reproduction is weighted by the coefficients of the CSFs and a contrast-masking metric.

3.2.5 Comparison of Selected Metrics Within the Group

The difference between the metrics in this group is mainly in the filtering stage and in the quality calculation. Since the quality calculation is mostly carried out with a color difference formula (ΔE^*_{ab} or ΔE_E), we will focus on the filtering. The S-CIELAB, S-CIELAB$_J$, and S-DEE use CSFs, where they differ in the normalization of the filters. On the other side, ABF uses a bilateral filter preserving edges.

Figure 3.4 shows the difference between images filtered with the three different metrics. We can see that S-CIELAB (the second image from the left) is slightly darker than the original (at left), which is

Fig. 3.4 Comparison between the different spatial filtering methods. From *left* to *right* we can see the original image, the image filtered with S-CIELAB, the image filtered with S-DEE and S-CIELAB$_J$, and the image filtered with ABF. All images have been computed for the same viewing conditions. In S-CIELAB we perceive a lightness shift due to the fact that the luminance filter is not normalized. In S-DEE and S-CIELAB$_J$ the luminance filter is normalized, which corrects this issue, and the ABF has a softer look, since it smooths on each side of an edge using an edge-preserving bilateral filter.

caused by the CSF and has been reported earlier [75]. The image filtered with the S-DEE and S-CIELAB$_J$ methods does not have this lightness shift since the DC component of the CSF is normalized to 1. We can also see blurring and loss of detail in the images filtered with S-CIELAB, S-DEE, and S-CIELAB$_J$. The S-DEE and S-CIELAB$_J$ have CSFs that enhance the frequencies where our HVS is most sensitive, and therefore they will in some cases preserve more detail than S-CIELAB. The image processed with ABF, on the right, has a softer look compared to the other metrics; most noticeable is the preservation of the hard edges, which is a feature of the bilateral filter.

3.3 High-level Metrics

High-level metrics quantify quality based on the idea that our HVS is adapted to extract information or structures from the image.

3.3.1 Universal Image Quality Index

The universal image quality index (UIQ) was proposed by [165]. This is a mathematically defined IQ metric for grayscale images, with no

HVS model incorporated. Because of this the metric is independent of viewing conditions and individual observers. It is also easy to calculate and has low complexity. The metric models any distortion as a combination of loss of correlation, luminance distortion, and contrast distortion. It is defined as

$$Q = \frac{4\sigma_{xy}\bar{x}\bar{y}}{(\sigma_x^2 + \sigma_y^2)[(\bar{x})^2 + (\bar{y})^2]}, \tag{3.26}$$

where x is the original, y the reproduction, $\bar{x} = \frac{1}{N}\sum_{u=1}^{N} x_i$, $\bar{y} = \frac{1}{N}\sum_{u=1}^{N} y_i$, $\sigma_x^2 = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{x})^2$.

The final measure, Q, is in the range $[-1, 1]$, where 1 is equivalent to two identical images. The results shown by the authors indicate that the UIQ outperforms the MSE significantly for different types of distortion, such as salt-and-pepper noise, mean shift, JPEG compression, additive Gaussian noise, and blurring.

3.3.1.1 Q_{color}

Toet and Lucassen [161] introduced a color image fidelity metric based on the UIQ [165]. The UIQ is calculated on each channel in the l, α, and β channels, where these channels are found by a transformation from the LMS space [135]. The final color metric is defined as

$$Q_{\text{color}} = \sqrt{w_l(Q_1)^2 + w_\alpha(Q_\alpha)^2 + w_\beta(Q_\beta)^2}, \tag{3.27}$$

where Q is the UIQ calculation. The weights, w, are set according to the distortion in each channel.

3.3.2 SSIM

The structural similarity (SSIM) index proposed by Wang et al. [167] attempts to quantify the visible difference between a distorted image and a reference image. This metric is based on the universal image quality (UIQ) index [165]. 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 IQ is the mean of all these local windows. SSIM is specified as

$$\mathrm{SSIM}(x,y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)}, \tag{3.28}$$

where μ is the mean intensity for signals x and y, and σ is the standard deviation of the signals x and y. C_1 is a constant defined as

$$C_1 = (K_1 L)^2, \tag{3.29}$$

where L is the dynamic range of the image, and $K_1 \ll 1$. The constant C_2 is similar to C_1, and is defined as

$$C_2 = (K_2 L)^2, \tag{3.30}$$

where $K_2 \ll 1$. These constants are used to stabilize the division of the denominator. SSIM is then computed for the entire image as

$$\mathrm{MSSIM}(X,Y) = \frac{1}{W}\sum_{j=1}^{W}\mathrm{SSIM}(x_j, y_j), \tag{3.31}$$

where X is the reference image, Y the distorted images, x_j and y_j are image content in local window j, and W indicates the total number of local windows. Figure 3.5 shows the SSIM flowchart, where the input is the original signal (image) and the distorted signal (image).

SSIM has received a lot of attention since its introduction. It has gone through extensive evaluation and influenced a number of other metrics, such as the color version $\mathrm{SSIM}_{\mathrm{IPT}}$ by Bonnier et al. [16] (Section 3.3.3). Because of this importance and our interesting in the idea of using structural information to measure IQ, we include SSIM in this survey.[2]

3.3.3 $\mathbf{SSIM_{IPT}}$

Bonnier et al. [16] developed and evaluated a color extension of SSIM, where SSIM was calculated for each channel in the IPT [46] color space.

[2] 16/02/2011: The implementation of SSIM can be found at `https://ece.uwaterloo.ca/~z70wang/research/ssim/`

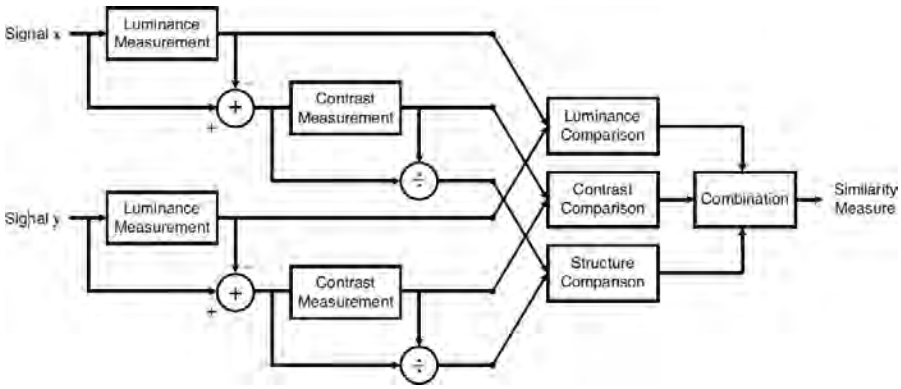


Fig. 3.5 SSIM flowchart. Signals x and y go through a luminance and contrast measurement before comparison of luminance, contrast, and structure. These results are combined in the final similarity measure. Figure reproduced from Wang et al. [167].

Then the quality values from each of the channels were combined using the geometrical mean. $\mathrm{SSIM}_{\mathrm{IPT}}$ is defined as

$$\mathrm{SSIM}_{\mathrm{IPT}}(x,y) = \sqrt[3]{\mathrm{SSIM}_I(x,y) \cdot \mathrm{SSIM}_P(x,y) \cdot \mathrm{SSIM}_T(x,y)}, \quad (3.32)$$

where SSIM_I, SSIM_P, and SSIM_T are the SSIM values for each of the channels in the IPT color space. This implementation is similar to the one used by Toet and Lucassen [161] for UIQ (Section 3.3.1), calculating the quality values for each color channel and then combining them.

3.3.4 Visual Information Fidelity

Sheikh and Bovik [145] proposed the visual information fidelity (VIF) criterion, which is an extension of the information fidelity criterion (IFC) by the same authors [146]. It quantifies the Shannon information present in the reproduction relative to the information present in the original. VIF uses a natural scene model based on a Gaussian scale mixture model in the wavelet domain, and as a HVS model they use an additive white Gaussian noise model.

The reference image is modeled by a Gaussian scale mixture in the wavelet domain. We model c, a collection of M neighboring wavelet coefficients from a local patch in a subband, as $c = \sqrt{z}u$, where u is a zero-mean Gaussian vector and $\sqrt{z}$ is an independent scalar random

variable. The VIF assumes that the distortion of the image can be described locally as a combination of uniform wavelet domain energy attenuation with added independent additive noise. The visual distortion is modeled as a stationary, zero-mean, additive white Gaussian noise process in the wavelet domain: $e = c + n$ and $f = d + n$, where e and f are the random coefficient vectors for the same wavelet subband in the perceived original and perceived distorted image. Random vectors c and d are from the same location in the same subband for the original and distorted image, and n denotes the independent white Gaussian noise with the covariance matrix $C_n = \sigma_n^2 I$. The VIF is calculated as the ratio of the summed mutual information in the subbands, which can be written as follows:

$$\text{VIF} = \frac{I(C;F|z)}{I(C;E|z)} = \frac{\sum_{i=1}^{N} I(c_i;f_i|z_i)}{\sum_{i=1}^{N} I(c_i;e_i|z_i)}, \tag{3.33}$$

where i is the index of local coefficients patches, including all subbands.

The VIF[3] criterion has been shown to perform well compared to other state-of-the-art metrics [85, 145, 147], and combined with the use of statistical information it is interesting to compare it against more traditional metrics.

3.3.5 Comparison of Selected Metrics Within the Group

It is difficult to compare the metrics within this group, since some metrics produce a map and others a single value. We will focus on SSIM, since it is the basis for many of the metrics in this group. To compare the SSIM against the other metrics, we show the resulting map from the target (Figure 3.6). Since SSIM is a grayscale metric, the images have been converted from color to grayscale.[4] SSIM shows the largest difference in the dark regions of the image (Figure 3.6). It is also interesting to notice the similarity between the ΔE_E (Figure 3.2(b)) and SSIM (Figure 3.6). To verify the similarities between these two

[3] 16/02/2011: The implementation of VIF is available at `http://live.ece.utexas.edu/research/quality/`.

[4] For all grayscale metrics we have converted the image using the rgb2gray function in Matlab, following the recommendation of Wang et al. `16/02/2011:https://ece.uwaterloo.ca/~z70wang/research/ssim/`.

Fig. 3.6 SSIM difference map between an original image (Figure 3.1(a)) and a reproduction (Figure 3.1(b)).

difference maps a two-dimensional correlation coefficient has been used:

$$r = \frac{\sum_m \sum_n (A_{mn} - \overline{A})(B_{mn} - \overline{B})}{\sqrt{(\sum_m \sum_n (A_{mn} - \overline{A})^2)(\sum_m \sum_n (B_{mn} - \overline{B})^2)}}, \tag{3.34}$$

where $\overline{A}$ is the mean of one map, and $\overline{B}$ is the mean for the other map. This method has previously been used to compare results from eye tracking maps [8, 120] and to compare an IQ metric to the regions marked by observers [23]. The correlation coefficient between these maps is -0.75 (negative because the DEE has a scale from 0 to infinity, where 0 is two idenitical images, while SSIM has a scale from 0 to 1, where 1 is two identical images), indicating a high degree of similarity between the two maps.

3.4 Other Approaches

Metrics considered in this group are based on other approaches or metrics combining two or more of the above groups.

3.4.1 Visual Signal-to-Noise Ratio

Chandler and Hemami [28] proposed the visual signal-to-noise ratio (VSNR), which is based on near-threshold and suprathreshold properties of the HVS, incorporating both low-level features and mid-level

features. The metric is calculated in 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 for each octave band based on the CSNR. The contrast is then compared to the detection threshold, and if it is higher than the distortion, it is considered to be suprathreshold (visible).

In this case a second stage is carried out, where a model of global precedence is proposed to account for the mid-level properties of the HVS. The global precedence takes into account that contrast of distortions should be proportioned across spatial frequency. The final metric is computed as the combination of the perceived contrast of the distortion and the disruption of global precedence.

We include VSNR[5] in this survey for its interesting properties. Since it is based on contrast thresholds, it will only take into account the visible difference, unlike the CSF-based metrics where the entire image is modulated. In addition, the use of both low-level features and mid-level features is compelling.

3.4.2 Hue Angle Algorithm

Hong and Luo [66, 67] proposed the hue angle algorithm, which is based on the idea that our eyes tend to be more tolerant of color errors over a smaller area whereas systematic errors over the entire image are noticeable and unacceptable. In order to apply this idea, the original and reproduction are transformed from L^*, a^*, b^* to L^*, C^*_{ab}, h_{ab}. A histogram for the 360 hue angles (h_{ab}) 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. 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 given a weight of 1/4 (i.e., the smallest areas with the same hue angle) and saved to a new array $hist$. The second quartile, with m hue angles, is given a weight of 1/2. The third

[5] 16/02/2011: The code for VSNR can be downloaded at `http://foulard.ece.cornell.edu/dmc27/vsnr/vsnr.html`

quartile, containing l hue angles, is given 1 as a weight, and the last quartile with the remaining hue angles is given a weight of 9/4.

$$\text{hist}(i) = \begin{cases} k(i) * 1/4, & i \in \{0, \ldots, n\} \\ k(i) * 1/2, & i \in \{n+1, \ldots, n+m\} \\ k(i) * 1, & i \in \{n+m+1, \ldots, n+m+l\} \\ k(i) * 9/4, & \text{otherwise} \end{cases} \tag{3.35}$$

The average color difference, computed using ΔE^*_{ab}, is calculated for all pixels having the same hue angle and stored in $CD[\text{hue}]$. Then the overall color difference for the image, CD_{image}, is calculated by multiplying the quartile weighting to every pixel with the average CIELAB color difference for the hue angle:

$$CD_{\text{image}} = \sum_{\text{hue}=1}^{360} \text{hist}[\text{hue}] * CD[\text{hue}]^2/4. \tag{3.36}$$

3.4.3 Spatial Hue Angle Metric

The spatial hue angle metric (SHAME) was proposed by Pedersen and Hardeberg [124, 126]. It follows the same framework as S-CIELAB, but incorporates a weighting based on hue (Figure 3.7). First the original and reproduction are spatially filtered with CSFs. Two different CSFs were tested, as was done for S-CIELAB (Section 3.2.1) and S-DEE (Section 3.2.2). The first, referred to as SHAME, performed worse than the latter, referred to as SHAME-II. After the spatial filtering, the hue angle algorithm [66, 67] (Section 3.4.2) is applied to account for the fact that systematic errors over the entire image are quite noticeable and unacceptable.

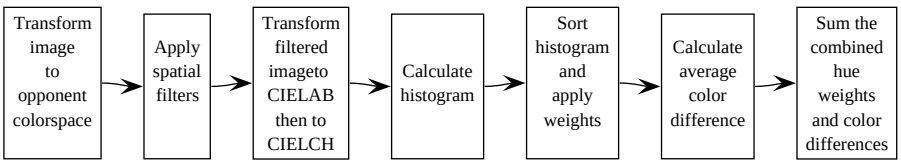


Fig. 3.7 SHAME workflow.

The approach of assigning different weights in the calculation of IQ has proven to be beneficial for IQ metrics in some cases [127], therefore SHAME is included in this survey.

3.4.4 S_{DOG}-CIELAB and S_{DOG}-DEE

Ajagamelle et al. [1] proposed two new metrics based on the S-CIELAB framework: S_{DOG}-CIELAB and S_{DOG}-DEE. They are modified to include a pyramidal subsampling, the DOG receptive field model [157]. The original and reproduction are transformed to the CIELAB color space for S_{DOG}-CIELAB and CIEXYZ for S_{DOG}-DEE. Then the images are subsampled with an antialiasing filter to various levels. A pixelwise neighborhood contrast calculation is done in each level using the DOG on the lightness and on the chromatic channels separately, thus providing local contrast maps for each level and each channel. Local contrast errors are computed using ΔE^*_{ab} for S_{DOG}-CIELAB and ΔE_E for S_{DOG}-DEE. At last, a weighted recombination of the local contrast maps is computed to obtain a single quality value.

3.4.5 Comparison of Selected Metrics Within the Group

Many of the metrics in this group produce one quality value, not a map. Therefore, it is difficult to show the differences between them. Nevertheless, they have one fundamental difference: VSNR is made for grayscale images and the others for color images. Also, the difference between the spatial filtering should be noted, where VSNR is built on contrast visibility, SHAME on CSFs, and S_{DOG}-CIELAB and S_{DOG}-DEE on the DOG.

4

Evaluation of Image Quality Metrics

The IQ metrics presented above take different approaches in their common goal to predict perceived IQ. In order to discover their performance in terms of correlation with the percept, we carried out extensive evaluations of the metrics (Table 4.1) by selecting a multitude of test images. There are a number of databases available online for evaluation of metrics (Table 4.2). There is a trend that the number of images and number of observers in the databases is increasing, and recently experiments have also been carried out on the web [49, 133]. Web-based experiments and other large experiments have the disadvantage that controlling the viewing conditions is difficult, and therefore they might be the best for evaluating metrics that take the viewing conditions into account.

Among the few public databases providing images for evaluation of IQ metrics, we have used the Tampere Image Database 2008 (TID2008) [131] and the IVC image database [88]. These two databases cover same the distortions as the other databases in Table 4.2. Additionally, we selected four datasets containing luminance-changed images [120, 127], JPEG and JPEG2000 compressed images [24], images with global variations of contrast, lightness, and saturation [2], and gamut-mapped

Table 4.1. Metrics evaluated in terms of correlation with the percept. Year gives the year of the publication where the metric was proposed, Name is the name of the metric, Author is the author's name, Type indicates the type of metric by the authors, HVS indicates if the metric has any simulation of the HVS, and Image indicates whether the metric is a color or grayscale metric. For the grayscale metrics the images were converted to grayscale images using the rgb2gray function in Matlab 7.5.0.

Year	Name	Author	Type	HVS	Image
—	MSE	—	Image difference	No	Gray
—	PSNR	—	Image difference	No	Gray
1976	ΔE^*_{ab}	CIE	Color difference	No	Color
1995	ΔE_{94}	CIE	Color difference	No	Color
1996	S-CIELAB	Zhang and Wandell [189]	Image difference	Yes	Color
2001	ΔE_{00}	CIE	Color difference	No	Color
2001	SCIELAB$_J$	Johnson and Fairchild [74]	Image quality	Yes	Color
2002	UIQ	Wang and Bovik [165]	Image quality	No	Gray
2002	Hue angle	Hong and Luo [66]	Image difference	No	Color
2003	Q_{color}	Toet and Lucassen [161]	Image fidelity	No	Color
2004	SSIM	Wang et al. [167]	Image quality	No	Gray
2006	VIF	Sheikh and Bovik [145]	Image fidelity	Yes	Gray
2006	PSNR-HVS-M	Egiazarian et al. [47]	Image quality	Yes	Gray
2006	SSIM$_{\text{IPT}}$	Bonnier et al. [16]	Image difference	No	Color
2007	VSNR	Chandler and Hemami [28]	Image fidelity	Yes	Gray
2009	ΔE_E	Oleari et al. [116]	Color difference	No	Color
2009	S-DEE	Simone et al. [155]	Image difference	Yes	Color
2009	SHAME	Pedersen and Hardeberg [126]	Image quality	Yes	Color
2009	SHAME-II	Pedersen and Hardeberg [126]	Image quality	Yes	Color
2009	ABF	Wang and Hardeberg [168]	Image difference	Yes	Color
2010	S$_{\text{DOG}}$-CIELAB	Ajagamelle et al. [1]	Image quality	Yes	Color
2010	S$_{\text{DOG}}$-DEE	Ajagamelle et al. [1]	Image quality	Yes	Color

images [44, 45]. The datasets from Pedersen [120, 127] and Ajagamelle et al. [2] are related to color difference, which is an important aspect for metrics to be able to evaluate. The dataset from Caracciolo [24] contains small compression artifacts, unlike the TID2008 database that has larger differences. This dataset will be able to tell if the metrics are able to evaluate just noticeable differences in terms of compression. The dataset from Dugay [44, 45] is based on gamut-mapping, where a number of different attributes are changed simultaneously, making it a difficult task for the metrics. To assure an extensive evaluation of the IQ metrics, these databases include a wide range of distortion, from color changes to structural changes, over a large set of images with many different characteristics.

Table 4.2. Image quality databases available online. Observers shows the total number of observers used to compute the scores.

Database	Type of distortion	Original scenes	Number of distortions	Total number of images	Observers
TID2008[1] [131]	JPEG, JPEG2000, blur, noise, etc.	25	17	1700	838
LIVE[2] [148]	JPEG, JPEG2000, blur, noise, and bit error	29	5	982	161
A57[3][27]	Quantization, blur, noise, JPEG, JPEG2000	3	6	57	7
IVC[4] [88]	JPEG, JPEG2000, LAR coding, and blurring	10	4	235	15
Toyama[5] [68]	JPEG and JPEG200	14	2	168	16
CSIQ[6] [86, 87]	JPEG, JPEG2000, noise, contrast, and blur.	30	6	866	35
WIQ[7] [48]	Wireless artifacts	7	Varying	80	30

[1]18/02/2011: `http://www.ponomarenko.info/tid2008.htm`
[2]18/02/2011: `http://live.ece.utexas.edu/research/quality`
[3]18/02/2011: `http://foulard.ece.cornell.edu/dmc27/vsnr/vsnr.html`
[4]18/02/2011: `http://www2.irccyn.ec-nantes.fr/ivcdb/`
[5]18/02/2011: `http://mict.eng.u-toyama.ac.jp/mictdb.html`
[6]18/02/2011: `http://vision.okstate.edu/index.php?loc=csiq`
[7]18/02/2011: `http://www.bth.se/tek/rcg.nsf/pages/wiq-db`

The performance of each metric is calculated as the correlation between the perceptual quality measured in psychophysical experiments and the quality values calculated by the metric. We opted for two standard types of correlation:

- The product-moment, or Pearson's, correlation coefficient, which assumes a normal distribution in the uncertainty of the data values and that the variables are ordinal.
- The Spearman's rank-correlation coefficient, which is a non-parametric measure of association based on the ranks of the data values, that describes the relationship between the variables without making any assumptions about the frequency distribution.

Confidence intervals will also be calculated for Pearson's correlation. One consideration is that the sampling distribution of Pearson's r is not

normally distributed. Therefore, Pearson's r is converted to Fisher's z' where confidence limits are derived. The values of Fisher's z' in the confidence interval are then converted back to Pearson's r.

The first step is to use Fisher's Z-transform:

$$z = \frac{1}{2}\ln\left(\frac{1+r}{1-r}\right), \tag{4.1}$$

where r is the correlation coefficient. The confidence intervals for r are calculated on the transformed r values (z'). The general formulation of confidence intervals for z' is

$$z' \pm z\sigma_{z'}, \tag{4.2}$$

where the criterion z is the desired confidence level (1.96 gives a 95% confidence interval), and $\sigma_{z'}$ is defined as

$$\sigma_{z'} = \frac{1}{\sqrt{N-3}}, \tag{4.3}$$

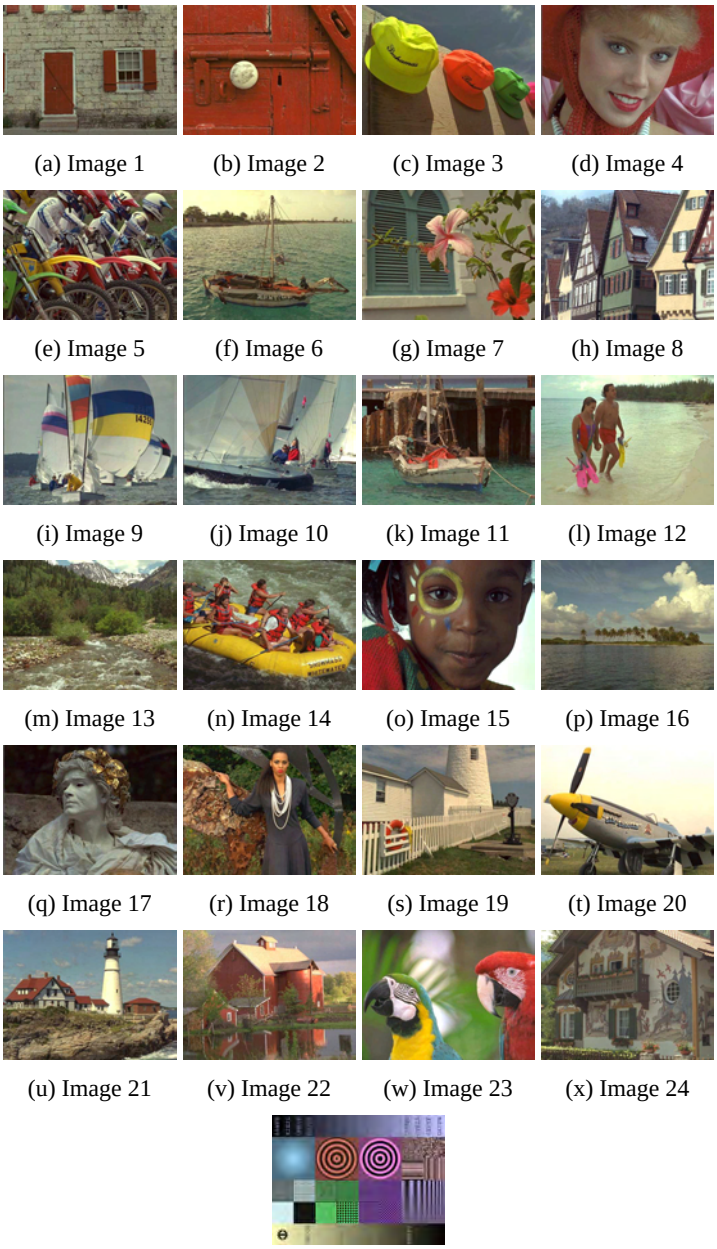
where N is the number of correlation samples. The upper and lower limits confidence interval limits for z' are found by using Equation (4.2). To translate from the z-space back in to the r-space, it is necessary to invert Equation (4.1). This results in the following equation:

$$r = \frac{e^{2z}-1}{e^{2z}+1}. \tag{4.4}$$

The upper and lower confidence interval limits for r can be found by using the upper and lower z values into Equation (4.4).

4.1 TID2008 Database, Ponomarenko et al.

The TID2008 database contains a total of 1,700 images, with 25 reference images and 17 types of distortions over four distortion levels (Figure 4.1 and Table 4.3) [131]. This database can be considered an extension of the well-known LIVE database [147]. The mean opinion scores (MOS) are the results of 654 observers performing the experiments. No viewing distance is stated in the TID database, so we have used a standard viewing distance (50 cm) for the metrics that require this setting. We have used version 1.0 of the TID2008 database, and



(a) Image 1 (b) Image 2 (c) Image 3 (d) Image 4

(e) Image 5 (f) Image 6 (g) Image 7 (h) Image 8

(i) Image 9 (j) Image 10 (k) Image 11 (l) Image 12

(m) Image 13 (n) Image 14 (o) Image 15 (p) Image 16

(q) Image 17 (r) Image 18 (s) Image 19 (t) Image 20

(u) Image 21 (v) Image 22 (w) Image 23 (x) Image 24

(y) Image 25

Fig. 4.1 The TID2008 database contains 25 reference images with 17 types of distortions over four levels.

Table 4.3. Overview of the distortions in the TID database and how they are related to the tested subsets. The database contains 17 types of distortions 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.

		Dataset							
	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	Non eccentricity pattern noise	−	−	−	+	−	+	+	+
15	Local block-wise distortion	−	−	−	−	−	+	+	+
16	Mean shift	−	−	−	−	−	+	+	+
17	Contrast change	−	−	−	−	−	+	+	+

then we have taken the log of the results of the metrics when calculating the correlation.

The correlations for the overall performance of the IQ metrics are listed in Table 4.4, and Pearson correlation is shown in Figure 4.2. These results show the correlation between all values from the IQ metrics and the subjective scores. The best metric for the full database measured by the Pearson correlation is the VIF, followed by VSNR, and then UIQ. These are metrics designed and optimized for the distortions found in the TID2008 database. The SSIM is not as good as its predecessor UIQ in terms of Pearson correlation, but it has a higher Spearman correlation, indicating a more correct ranking. It is worth noting that metrics based on color difference formulas do not perform well; S-CIELAB was the best, with a correlation of only 0.43. The simple color difference formulas are found in the bottom part of the list with a very low correlation for the entire database, as measured by both Pearson and Spearman correlations. The low correlation in many

Table 4.4. Comparison of the IQ metrics over all the images of the TID2008 database. The results are sorted from high to low for both Pearson and Spearman correlation. VIF, UIQ, and VSNR correlate rather well with subjective evaluation, with Pearson and Spearman correlations above 0.6. S_{DOG}-CIELAB is calculated with $R_c = 1, R_s = 2$, pyramid $= 1, \frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \ldots$, and equal weighting of the level. S_{DOG}-DEE is calculated with $R_c = 3, R_s = 4$, pyramid $= 1, \frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \ldots$, and the variance used as weighting of the level to give more importance to the high-resolution levels. S-CIELAB$_J$ includes only the improved CSF. For other metrics standard settings have been used, unless otherwise is stated. The total number of images (N) is 1700.

Metric	Pearson	Metric	Spearman
VIF	0.74	VIF	0.75
VSNR	0.71	VSNR	0.72
UIQ	0.62	SSIM	0.63
PSNR-HVS-M	0.59	PSNR-HVS-M	0.61
SSIM	0.55	UIQ	0.60
MSE	0.54	SSIM-IPT	0.57
Q_{COLOR}	0.52	MSE	0.56
PSNR	0.51	PSNR	0.56
SSIM-IPT	0.48	Q_{COLOR}	0.48
S-CIELAB	0.43	S-CIELAB	0.45
SHAME-II	0.41	SHAME-II	0.41
$S_{\mathrm{DOG}} - CIELAB$	0.38	ΔE_E	0.38
$S - CIELAB_J$	0.32	$S_{\mathrm{DOG}} - CIELAB$	0.37
SHAME	0.30	SHAME	0.35
$S - DEE$	0.29	$S - CIELAB_J$	0.31
ABF	0.28	Hue angle	0.29
ΔE_E	0.27	$S - DEE$	0.29
Hue angle	0.26	ΔE^*_{ab}	0.28
$S_{\mathrm{DOG}} - DEE$	0.26	ABF	0.26
ΔE^*_{ab}	0.23	$S_{\mathrm{DOG}} - DEE$	0.26
ΔE^*_{94}	−0.06	ΔE^*_{00}	0.24
ΔE^*_{00}	−0.06	ΔE^*_{94}	0.24

of the IQ metrics could be caused by the large number of different distortions in the TID2008 database, since there might be scale differences between the different distortions and scenes. This problem occurs when the observers rate images with different distortions as having the same quality, but the IQ metrics rate them as different. Looking at the same distortion, the metric might have a high correlation, but the measured correlation of two or more distortions might be low because of the scale differences.

When looking at specific distortions (Table 4.5), PSNR-HVS-M is the best for five of the seven datasets. However, in the two last datasets PSNR-HVS-M does not perform well, with a correlation of 0.26 in

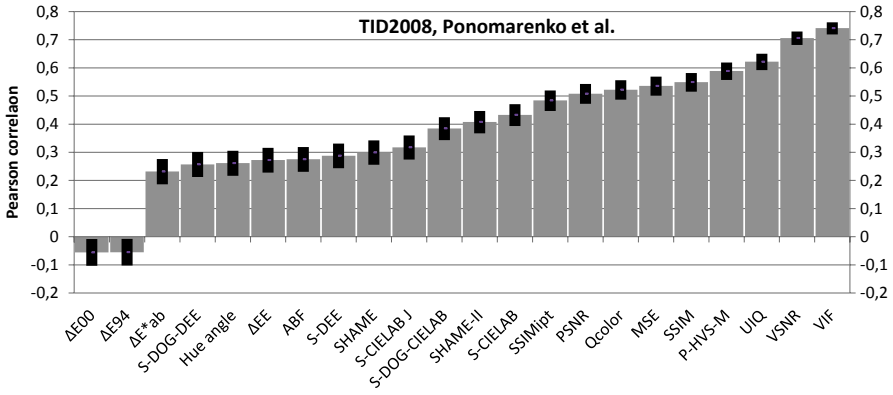


Fig. 4.2 Pearson correlation for TID2008 plotted with a 95% confidence interval. The correlation values are sorted from low to high.

both sets. The reason for this is that PSNR-HVS-M has a problem with the mean shift distortion. The PSNR-HVS-M incorporates a mean shift removal, and rates images where a mean shift has occurred as similar even though the observers rate them as different. VSNR and VIF do not have this problem, and therefore they are able to obtain a higher correlation for both the exotic and exotic2 dataset. SSIM has a low correlation in the exotic dataset, and a much higher correlation in exotic2. The reason for this is a couple of outliers in the exotic dataset; with the introduction of more images in exotic2 several points are placed between the outliers and the other data points, increasing the correlation. Some of the problem in the exotic dataset for SSIM stems from the contrast change distortion, but most of it stems from scale differences between the distortions (i.e., one distortion has an overall smaller difference than the other, but they are judged similarly by the observers). Within a specific distortion a fairly high correlation is found. Among the color difference-based metrics S-CIELAB is the most stable metric, with a high correlation in the noise, noise2, hard, and simple datasets. Since S-CIELAB incorporates spatial filtering, it will be able to modulate the frequencies that are less perceptible. Therefore S-CIELAB shows good results for the datasets with noise. S-CIELAB has problems in the exotic and exotic2 because of scale differences, as do many of the other metrics.

Table 4.5. Correlation of the IQ metrics for the seven different datasets in the TID2008 database. Each dataset is sorted descending from the highest Pearson correlation to the lowest Pearson correlation. N = 700,800,700,800,400,400, and 600 for the noise, noise2, safe, hard, simple, exotic and exotic2 datasets, respectively.

Dataset													
Noise		Noise2		Safe		Hard		Simple		Exotic		Exotic2	
PSNR-HVS-M	0.92	PSNR-HVS-M	0.87	PSNR-HVS-M	0.92	PSNR-HVS-M	0.81	PSNR-HVS-M	0.94	VSNR	0.57	VIF	0.66
VSNR	0.83	VIF	0.85	VSNR	0.84	VIF	0.79	VSNR	0.89	VIF	0.43	VSNR	0.59
VIF	0.78	VSNR	0.82	VIF	0.82	S-CIELAB	0.78	VIF	0.86	PSNR-HVS-M	0.26	SSIM	0.57
S-CIELAB	0.77	S-CIELAB	0.75	S-CIELAB	0.76	VSNR	0.75	MSE	0.83	$\mathrm{S_{DOG}}_{CIELAB}$	0.23	UIQ	0.56
MSE	0.74	MSE	0.73	MSE	0.74	SSIM-IPT	0.74	PSNR	0.81	MSE	0.19	Q_{COLOR}	0.50
PSNR	0.72	PSNR	0.72	PSNR	0.71	SSIM	0.73	S-CIELAB	0.78	UIQ	0.16	SSIM-IPT	0.49
SHAMEII	0.63	SHAMEII	0.59	UIQ	0.65	UIQ	0.71	UIQ	0.77	SSIM	0.16	PSNR-HVS-M	0.26
S-CIE LAB_J	0.59	UIQ	0.57	Q_{COLOR}	0.60	ΔE^*_{ab}	0.68	Q_{COLOR}	0.70	PSNR	0.15	MSE	0.25
ABF	0.57	S-CIE LAB_J	0.54	SHAMEII	0.58	Hue Angle	0.67	SSIM	0.67	SHAME	0.12	PSNR	0.22
UIQ	0.53	SHAME	0.51	S-CIE LAB_J	0.57	PSNR	0.67	ΔE^*_{ab}	0.65	$S-DEE$	0.04	SHAMEII	0.13
SHAME	0.51	$\mathrm{S}_{DOG_{DEE}}$	0.51	ABF	0.56	MSE	0.67	ABF	0.64	SHAMEII	0.03	$\mathrm{S_{DOG}}_{CIELAB}$	0.10
Q_{COLOR}	0.50	SSIM	0.50	SHAME	0.52	SHAMEII	0.66	Hue Angle	0.62	SSIM-IPT	0.03	$S-DEE$	0.09
$\mathrm{S}_{DOG_{DEE}}$	0.48	SSIM-IPT	0.47	$\mathrm{S_{DOG}}_{CIELAB}$	0.50	Q_{COLOR}	0.65	SSIM-IPT	0.61	Q_{COLOR}	−0.02	SHAME	0.08
ΔE_E	0.47	Q_{COLOR}	0.46	SSIM	0.48	S-CIELAB$_J$	0.65	S-CIELAB$_J$	0.56	ΔE_E	−0.09	ΔE_E	0.06
ΔE^*_{ab}	0.46	ABF	0.45	SSIM-IPT	0.44	ABF	0.56	SHAMEII	0.55	$\mathrm{S_{DOG}}_{DEE}$	−0.14	S-CIELAB	0.00
$\mathrm{S_{DOG}}_{CIELAB}$	0.45	ΔE_E	0.45	ΔE^*_{ab}	0.44	$\mathrm{S_{DOG}}_{DEE}$	0.49	ΔE^*_{94}	0.49	ABF	−0.20	Hue Angle	−0.01
Hue Angle	0.45	$S-DEE$	0.42	$\mathrm{S_{DOG}}_{DEE}$	0.43	$S-DEE$	0.49	$\mathrm{S_{DOG}}_{CIELAB}$	0.48	S-CIELAB	−0.21	$\mathrm{S_{DOG}}_{DEE}$	−0.03
ΔE^*_{94}	0.44	$\mathrm{S_{DOG}}_{CIELAB}$	0.40	Hue Angle	0.42	ΔE_E	0.45	ΔE^*_{00}	0.48	S-CIELAB$_J$	−0.22	S-CIELAB$_J$	−0.03
SSIM	0.43	ΔE^*_{94}	0.38	ΔE^*_{94}	0.42	$\mathrm{S_{DOG}}_{CIELAB}$	0.43	SHAME	0.48	Hue Angle	−0.25	ΔE^*_{ab}	−0.04
ΔE^*_{00}	0.43	ΔE^*_{00}	0.37	ΔE^*_{00}	0.41	SHAME	0.40	ΔE_E	0.39	ΔE^*_{ab}	−0.30	ABF	−0.13
SSIM-IPT	0.42	Hue Angle	0.36	ΔE_E	0.40	ΔE^*_{94}	0.25	$\mathrm{S_{DOG}}_{DEE}$	0.36	ΔE^*_{00}	−0.54	ΔE^*_{00}	−0.21
$S-DEE$	0.41	ΔE^*_{ab}	0.33	$S-DEE$	0.39	ΔE^*_{00}	0.25	$S-DEE$	0.34	ΔE^*_{94}	−0.54	ΔE^*_{94}	−0.21

4.2 Luminance Changed Images, Pedersen et al.

The database includes four original images (Figure 4.3) reproduced with different changes in lightness. Each scene has been altered in four ways globally and four ways locally [120, 127]. The globally changed images had an increase in luminance of 3 and 5 ΔE^*_{ab}, and a decrease in luminance of 3 and 5 ΔE^*_{ab}. The local regions had changes in luminance of 3 ΔE^*_{ab}. Twenty-five observers participated in the pair comparison experiment, which was carried out on a calibrated LaCIE electron 22 blue II CRT monitor, in a gray room. The viewing distance was set to 80 cm, and the ambient light measured as 17 lux. The images were judged in a pair comparison setup, with the original in the middle and a reproduction on each side.

For this database, a group of IQ metrics correlates reasonably well with subjective assessment (Table 4.6 and Figure 4.4). The original hue angle algorithm, SHAME, SHAME-II, S-CIELAB, and ABF exhibit the best Pearson correlation coefficients, all with a Pearson correlation above 0.8. It is also interesting to notice that all of these are based on color differences, using the ΔE^*_{ab} as a color difference formula. Because the changes in this database have been carried out according to ΔE^*_{ab}, it is not surprising that these metrics perform well. The metrics with a low correlation usually have a problem in the images where a local small change from the original is highly perceivable. It is also interesting to notice that the IQ metrics based on structural similarity, such as SSIM

(a) (b) (c) (d)

Fig. 4.3 Images in the experiment from Pedersen et al. [127]. These images were changed in lightness both globally and locally.

Table 4.6. Correlation between subjective scores and metric scores for the luminance changed images from Pedersen et al. [127]. We can see that the metrics based on color difference perform well compared to the metrics based on structural similarity. The results are sorted from high to low for both Pearson and Spearman correlation. $N = 32$.

Metric	Pearson	Metric	Spearman
SHAME-II	0.82	ABF	0.78
SHAME	0.81	S-CIELAB	0.77
Hue angle	0.81	$S - CIELAB_J$	0.76
ABF	0.81	ΔE^*_{ab}	0.74
S-CIELAB	0.80	Hue angle	0.73
$S - CIELAB_J$	0.78	SHAME-II	0.73
ΔE^*_{ab}	0.76	MSE	0.72
MSE	0.67	PSNR	0.72
PSNR	0.66	SHAME	0.71
ΔE^*_{94}	0.65	PSNR-HVS-M	0.68
ΔE_E	0.63	ΔE^*_{94}	0.68
PSNR-HVS-M	0.63	ΔE^*_{00}	0.68
ΔE^*_{00}	0.63	ΔE_E	0.64
UIQ	0.45	SSIM-IPT	0.51
$S - DEE$	0.39	$S - DEE$	0.51
VIF	0.39	UIQ	0.49
SSIM-IPT	0.30	SSIM	0.49
Q_{COLOR}	0.28	Q_{COLOR}	0.48
$S_{\text{DOG}} - DEE$	0.24	VIF	0.35
SSIM	0.22	$S_{\text{DOG}} - DEE$	0.32
$S_{\text{DOG}} - CIELAB$	0.20	$S_{\text{DOG}} - CIELAB$	0.32
VSNR	0.02	VSNR	0.02

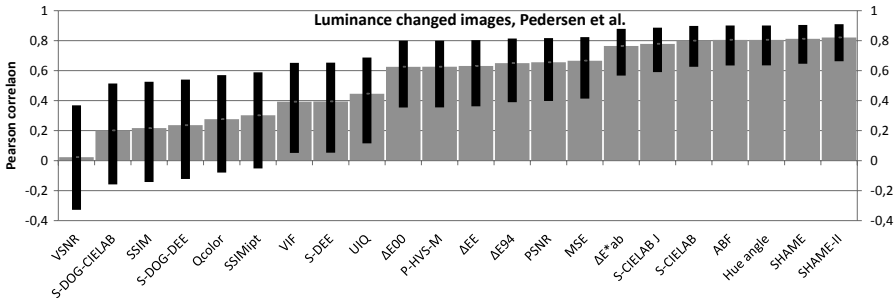


Fig. 4.4 Pearson correlation for the Pedersen et al. database plotted with a 95% confidence interval. The correlation values are sorted from low to high.

and UIQ, have lower correlation values than many other metrics. Simple metrics, such as MSE and PSNR, have fairly high correlation values. One should keep in mind that the number of images in this database is small (32), and therefore confidence intervals are large.

4.3 JPEG and JPEG2000 Compressed Images, Caracciolo

The 10 original images (Figure 4.5) of this database were corrupted by JPEG and JPEG2000 compression, generating a total of 80 degraded images [24]. The parameters for these distortions were randomly chosen with predefined ranges (Table 4.7). The images were used in a category judgment experiment, where the original was presented on one side and the compressed version on the other side. Each pair of images (original and compressed) was displayed on an Eizo ColorEdge CG241W digital LCD display. The monitor was calibrated and profiled using GretagMacbeth Eye-One Match 3. The settings on the monitor were sRGB with 40% of brightness and a resolution of 1600 × 1200 pixels. The white point was set to the D65 and the gamma was set to a value

(a) Balls (b) Barbara (c) Cafe (d) Flower

(e) House (f) Mandrill (g) Parrots

(h) Picnic (i) Poster (j) Sails

Fig. 4.5 The 10 images in the dataset from Caracciolo [24]. These images have have been compressed with JPEG and JPEG2000.

Table 4.7. Selected *bit per pixels* for each image for the dataset from Caracciolo [24].

Image scene	*bit per pixel* JPEG		*bit per pixel* JPEG2000	
Image 1: Balls	0.8783	0.9331	1.0074	1.1237
Image 2: Barbara	0.9614	1.0482	1.1003	1.2569
Image 3: Cafe	0.8293	0.9080	0.9869	1.0977
Image 4: Flower	0.6611	0.7226	0.7844	0.8726
Image 5: House	0.9817	1.0500	1.1445	1.2920
Image 6: Mandrian	1.4482	1.5910	1.7057	1.9366
Image 7: Parrots	0.6230	0.6891	0.7372	0.8169
Image 8: Picnic	1.0353	1.1338	1.2351	1.3765
Image 9: Poster	0.4148	0.4474	0.4875	0.5420
Image 10: Sails	0.7518	0.8198	0.8774	0.9927

of 2.2. The display was placed at a viewing distance of 70 cm. A total of 18 observers participated in the experiment.

We found that all the metrics perform poorly performance for this database (Table 4.8 and Figure 4.6), spanning a range below 0.4. This is probably because these particular images were initially selected in order to determine the just noticeable distortion (JND). Only small distortions were applied to the original images, making it arduous for the observers to assess IQ, and therefore also very difficult for the IQ metrics. PSNR-HVS-M and PSNR have the best Pearson correlation, but they do not perform well, with a Pearson correlation of only 0.31. With the given 95% confidence interval, all the evaluated metrics perform similarly. It is interesting to notice the higher ranking of the color difference-based IQ metrics, such as S-CIELAB and SHAME, although they are still not performing very well.

This dataset is similar to the simple dataset in the TID2008 dataset, which contains JPEG and JPEG2000 compressed images in addition to Gaussian blur and additive Gaussian noise (Table 4.3). However, there is a large difference in the performance of the metrics; for example, PSNR-HVS-M had a correlation of 0.94 for the simple dataset and only a correlation 0.31 for the images from Caracciolo [24]. The reason for this difference is the visual differences between the images in the dataset. In the TID2008 database the differences between the different levels of a distortion is high while in the images from Caracciolo [24] they are just noticeable. This results in a difficult task for the observers, and of course very difficult for an IQ metric.

Table 4.8. Correlation between subjective scores and metric scores for the JPEG and JPEG2000 compressed images from Caracciolo [24]. The results are sorted from high to low for both Pearson and Spearman correlation. All IQ metrics have a low correlation, both for Pearson and Spearman correlation. However, it is interesting to notice the higher ranking of the color difference-based IQ metrics, such as S-CIELAB and SHAME. $N = 80$.

Metric	Pearson	Metric	Spearman
PSNR	0.31	SSIM-IPT	0.37
PSNR-HVS-M	0.31	SHAME	0.36
SSIM	0.29	PSNR-HVS-M	0.36
MSE	0.26	$S - CIELAB_J$	0.36
SSIM-IPT	0.25	SSIM	0.35
$S_{\mathrm{DOG}} - CIELAB$	0.23	S-CIELAB	0.33
S-CIELAB	0.22	ABF	0.33
VIF	0.21	PSNR	0.32
$S - CIELAB_J$	0.17	MSE	0.30
ABF	0.16	ΔE^*_{ab}	0.28
ΔE^*_{ab}	0.16	VIF	0.24
Q_{COLOR}	0.15	ΔE_E	0.24
SHAME	0.15	SHAME-II	0.24
ΔE^*_{00}	0.13	Hue angle	0.23
$S - DEE$	0.13	Q_{COLOR}	0.21
ΔE_E	0.13	ΔE^*_{00}	0.20
ΔE^*_{94}	0.13	$S - DEE$	0.19
$S_{\mathrm{DOG}} - DEE$	0.11	ΔE^*_{94}	0.19
SHAME-II	0.09	$S_{\mathrm{DOG}} - CIELAB$	0.18
VSNR	0.08	$S_{\mathrm{DOG}} - DEE$	0.14
Hue angle	0.07	UIQ	0.11
UIQ	0.05	VSNR	0.11

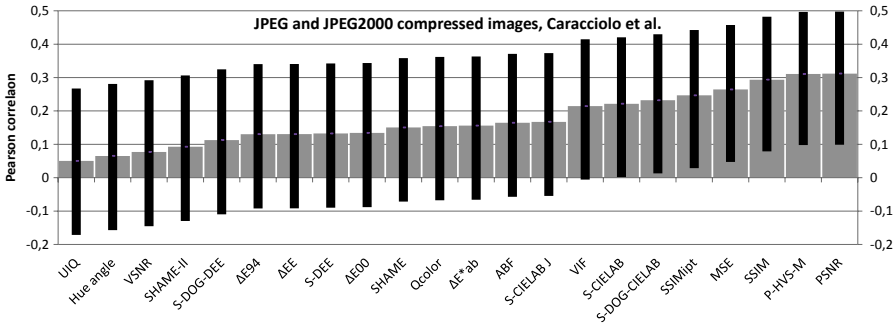


Fig. 4.6 Pearson correlation for the Caracciolo database plotted with a 95% confidence interval. The correlation values are sorted from low to high.

4.4 IVC Database, Le Callet et al.

The IVC database contains blurred images and images distorted by three types of lossy compression techniques: JPEG, JPEG2000, and locally adaptive resolution [88]. This database contains 10 different images (Figure 4.7). The viewing distance was set to 87 cm. Only the color images from this database have been used.

The most accurate IQ metrics are UIQ, VSNR, and VIF (Table 4.9 and Figure 4.8). It is interesting to note that these metrics are based on

Fig. 4.7 The images in the IVC database from Callet and Autrusseau [88]. These images have been compressed with three different compression techniques.

Table 4.9. Correlation between subjective scores and metric scores for the images from the IVC database. The results are sorted from high to low for both Pearson and Spearman correlation. $N = 185$.

Metric	Pearson	Metric	Spearman
VIF	0.88	VIF	0.90
UIQ	0.82	UIQ	0.83
VSNR	0.77	VSNR	0.78
PSNR-HVS-M	0.73	SSIM	0.78
SSIM-IPT	0.71	SSIM-IPT	0.77
SSIM	0.70	PSNR-HVS-M	0.77
PSNR	0.67	MSE	0.69
S-CIELAB	0.61	PSNR	0.69
Q_{COLOR}	0.61	SHAME	0.68
ΔE_E	0.56	ΔE^*_{ab}	0.66
SHAME	0.55	Q_{COLOR}	0.66
$S_{\mathrm{DOG}} - DEE$	0.55	S-CIELAB	0.66
ΔE^*_{ab}	0.54	ΔE_E	0.62
ABF	0.51	$S_{\mathrm{DOG}} - DEE$	0.58
MSE	0.51	ABF	0.56
$S - DEE$	0.46	Hue angle	0.52
$S - CIELAB_J$	0.42	ΔE^*_{00}	0.47
ΔE^*_{00}	0.38	ΔE^*_{94}	0.46
ΔE^*_{94}	0.37	$S - DEE$	0.46
SHAME-II	0.37	$S - CIELAB_J$	0.43
Hue angle	0.35	SHAME-II	0.42
$S_{\mathrm{DOG}} - CIELAB$	0.29	$S_{\mathrm{DOG}} - CIELAB$	0.32

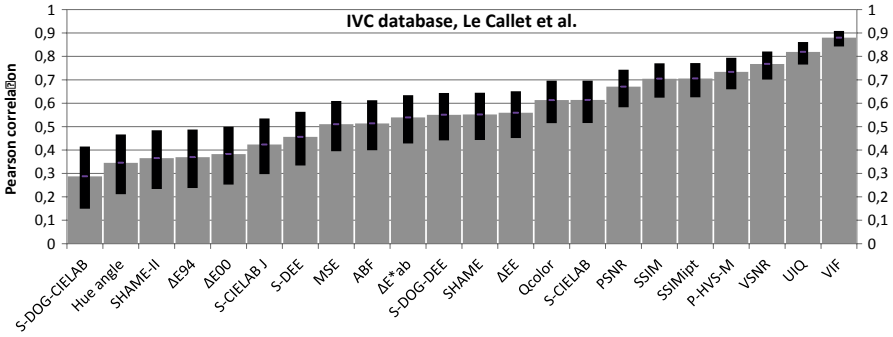


Fig. 4.8 Pearson correlation for the IVC database plotted with a 95% confidence interval. The correlation values are sorted from low to high.

different approaches: the VIF is based on statistics, UIQ on structural similarity, and VSNR on contrast thresholds. Many of the grayscale metrics have a higher performance than the color-based metrics; this

could be explained by the fact that the changes in this database are not directly related to color. The IVC database is very similar to the simple dataset from the TID2008 database: both contain blurring and compression, and we notice that the metrics with a high correlation for the simple dataset in TID2008 have a high correlation for the IVC database.

4.5 Images Altered in Contrast, Lightness, and Saturation, Ajagamelle et al.

This database contains 10 original images (Figure 4.9), covering a wide range of characteristics and scenes [2]. The images were modified on a global scale with separate and simultaneous variations of contrast, lightness, and saturation. The experiment was carried out as a category judgment experiment with 14 observers. The viewing distance was set to 70 cm, and the ambient illumination was 40 lux.

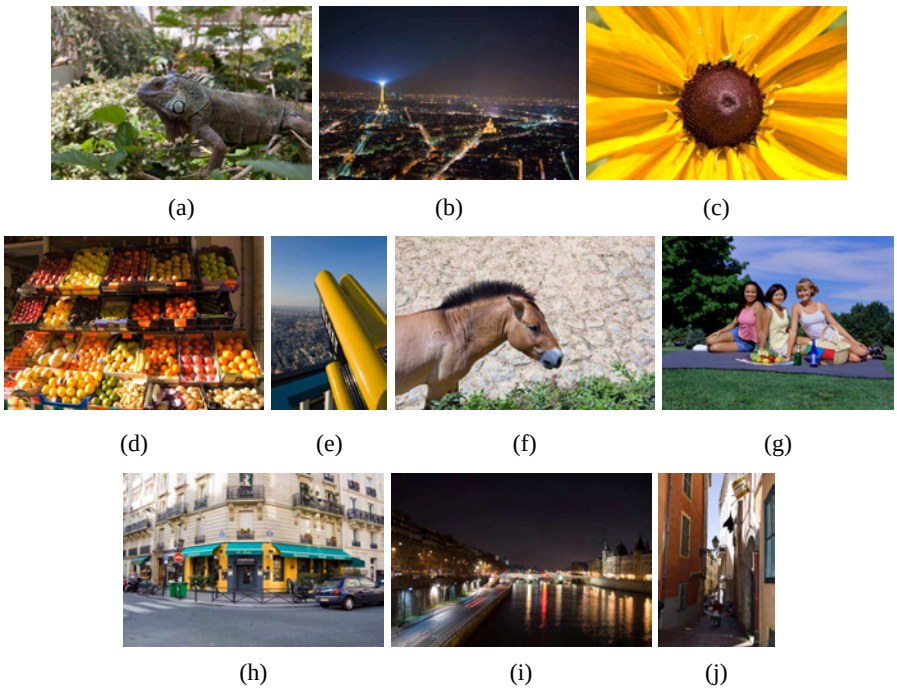


Fig. 4.9 Images altered in contrast, lightness, and saturation from Ajagamelle et al. [2].

Table 4.10. Correlation between subjective scores and metric scores for the images from Ajagamelle et al. [2]. The results are sorted from high to low for both Pearson and Spearman correlation. PSNR-HVS-M has the highest Pearson correlation, while SHAME-II has the best Spearman correlation. VSNR has the lowest correlation, which is caused by scale differences between the difference images. $N = 80$.

Metric	Pearson	Metric	Spearman
PSNR-HVS-M	0.79	SHAME-II	0.79
ΔE^*_{ab}	0.75	SSIM-IPT	0.76
ΔE^*_{94}	0.74	ABF	0.76
ΔE^*_{00}	0.74	PSNR-HVS-M	0.75
ABF	0.73	Q_{COLOR}	0.75
PSNR	0.72	UIQ	0.73
S-CIELAB	0.67	SHAME	0.73
SSIM-IPT	0.66	S-CIELAB	0.72
SSIM	0.64	Hue Angle	0.72
UIQ	0.63	ΔE^*_{ab}	0.72
Hue Angle	0.62	SSIM	0.72
SHAME-II	0.62	MSE	0.71
Q_{COLOR}	0.62	PSNR	0.71
MSE	0.61	ΔE^*_{00}	0.69
DEE	0.60	ΔE^*_{94}	0.69
$\mathrm{S_{DOG}}$-DEE	0.59	DEE	0.68
$S - CIELAB_J$	0.58	$S - CIELAB_J$	0.65
VIF	0.53	$\mathrm{S_{DOG}}$-DEE	0.61
SHAME	0.50	VIF	0.59
$S_{\mathrm{DOG}} - CIELAB$	0.41	$S_{\mathrm{DOG}} - CIELAB$	0.48
$S - DEE$	0.40	$S - DEE$	0.41
VSNR	0.16	VSNR	0.14

PSNR-HVS-M has the highest Pearson correlation, while SHAME-II has the best Spearman correlation (Table 4.10). We can see that the differences between the metrics are low (Figure 4.10), and many of them overlap. This is also true for the Spearman correlation, where 13 metrics have a correlation between 0.71 and 0.79. The color difference formulas from CIE (ΔE^*_{ab}, ΔE_{94}, and ΔE_{00}) all perform very well, as does PSNR. This is different from many of the other databases, where these simple pixelwise metrics do not correlate well with perceived IQ. The high correlation over some of the other databases is probably due to the global variations made by Ajagamelle et al., since the pixelwise IQ metrics weight all pixels equally. The results from this database are similar to the results from the lightness changed images from Pedersen et al. [127].

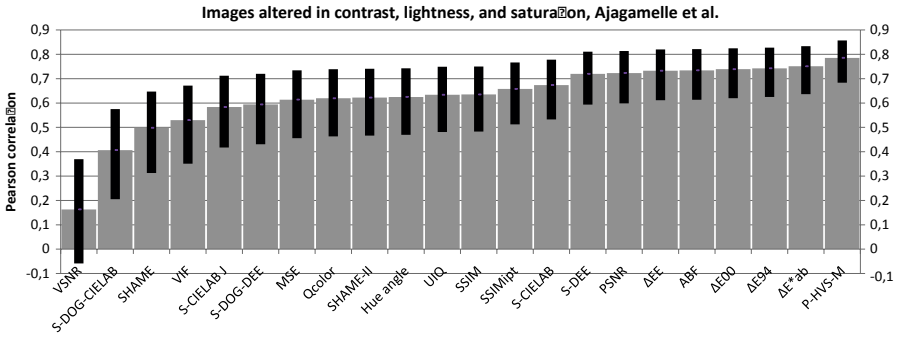


Fig. 4.10 Pearson correlation for the Ajagamelle et al. database plotted with a 95% confidence interval. The correlation values are sorted from low to high.

4.6 Gamut Mapped Images, Dugay et al.

In this dataset, 20 original images (Figure 4.11) were gamut-mapped with five different algorithms [44, 45]:

- HPminDE (hue preserving minimum ΔE^*_{ab} clipping), a baseline gamut-mapping algorithm proposed by the CIE [37]. The algorithm does not change in-gamut colors at all, while out-of-gamut colors are mapped to the closest color on the destination gamut while preserving the hue.
- SGCK [37] is an advanced spatially invariant sequential gamut compression algorithm. The lightness is first compressed by a chroma dependent sigmoidal scaling, resulting in high-chroma colors being compressed less than neutral ones. The resulting colors are then compressed along lines toward the cusp [105] of the destination gamut using a 90% knee scaling function. For the final compression the image gamut is used as the source gamut.
- Zolliker and Simon [192] proposed a spatial gamut mapping algorithm; its goal being to recover local contrast while preserving lightness, saturation and global contrast. The first step is a simple clipping; then, by using an edge-preserving high pass filter, the difference between the original and gamut clipped image is filtered. The filtered image is then added to

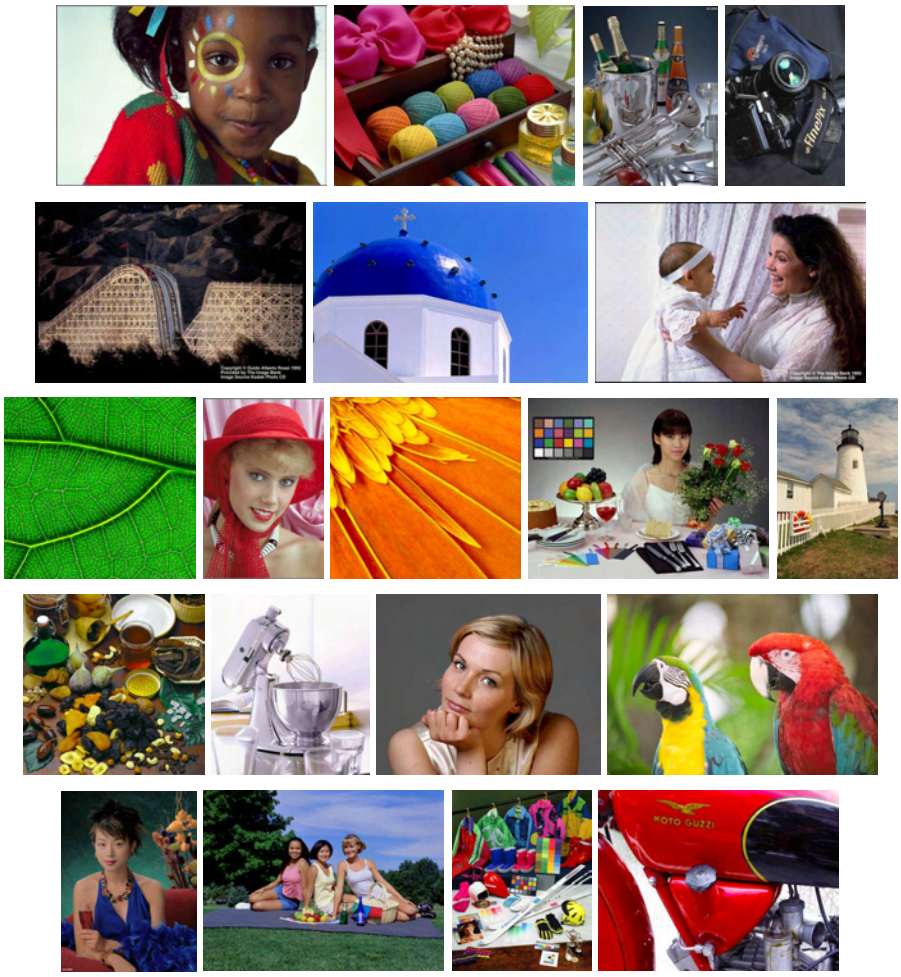


Fig. 4.11 Images from the experiment carried out Dugay et al. [45]. These images have been gamut-mapped with five different algorithms.

the gamut clipped image. The last step is to clip the image to make it in-gamut.

- Kolås and Farup [80] recently proposed a hue- and edge-preserving spatial color gamut mapping algorithm. The image is first gamut clipped along straight lines toward the center of the gamut. A relative compression map is then created from the original and clipped images. Using this

compression map, a new image can be constructed as a linear convex combination of the original image and neutral gray image. This image is in turn filtered by an edge-preserving smoothing minimum filter. As the final step the gamut mapped image is constructed as a linear convex combination of the original image and neutral gray using the filtered map.

- Farup et al. [52] proposed a multiscale algorithm that preserves hue and local relationship between closely related pixel colors. First a scale-space representation of the image is constructed. Then the lowest scale is gamut-clipped. The resulting gamut compression is applied to the image at the next smallest scale. Operators are used to reduce haloing. The process is repeated until all scales have been treated. The Fourier domain is used to speed up the process.

Table 4.11. Correlation between subjective scores and metric scores for the gamut-mapped images from Dugay [44]. The results are sorted from high to low for both Pearson and Spearman correlation. $N = 100$.

Metric	Pearson	Metric	Spearman
VIF	0.31	VIF	0.27
UIQ	0.31	UIQ	0.19
Q_{COLOR}	0.19	Q_{COLOR}	0.12
SSIM	0.16	VSNR	0.09
MSE	0.09	SSIM	0.05
VSNR	0.09	$S - CIELAB_J$	0.03
PSNR-HVS-M	0.07	PSNR-HVS-M	0.03
$S_{\mathrm{DOG}} - CIELAB$	0.05	ΔE^*_{00}	0.00
PSNR	0.04	ΔE^*_{94}	0.00
ΔE^*_{00}	0.03	SSIM-IPT	−0.02
$S_{\mathrm{DOG}} - DEE$	0.02	MSE	−0.02
$S - CIELAB_J$	0.02	$S_{\mathrm{DOG}} - DEE$	−0.02
ΔE^*_{94}	0.02	PSNR	−0.03
SSIM-IPT	0.01	ΔE_E	−0.06
Hue angle	0.00	ΔE^*_{ab}	−0.08
ΔE_E	−0.01	$S_{\mathrm{DOG}} - CIELAB$	−0.08
ΔE^*_{ab}	−0.03	Hue angle	−0.08
$S - DEE$	−0.03	$S - DEE$	−0.09
SHAME	−0.03	SHAME	−0.09
S-CIELAB	−0.06	S-CIELAB	−0.11
ABF	−0.08	ABF	−0.11
SHAME-II	−0.10	SHAME-II	−0.13

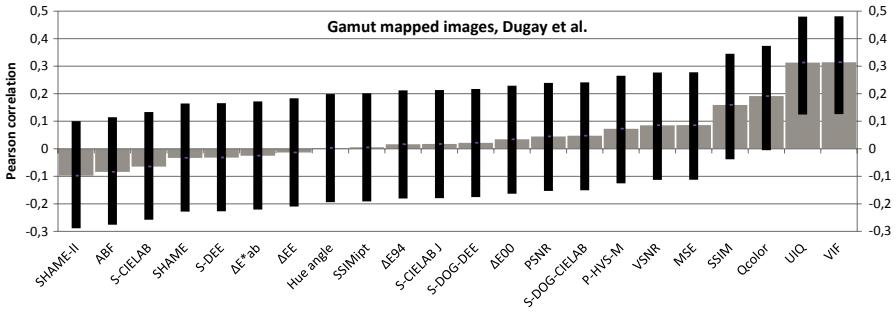


Fig. 4.12 Pearson correlation for the Dugay et al. database plotted with a 95% confidence interval. The correlation values are sorted from low to high.

The 20 different images were evaluated by 20 observers in a pair comparison experiment. The monitor used was a Dell 2407WFP LCD display calibrated with a D65 white point and a gamma of 2.2. The level of ambient illumination on the monitor was around 20 lux. The viewing distance for the observers was approximately 50 cm.

We see from the results in Table 4.11 and Figure 4.12 that most of the metrics do not correlate well with perceived quality. In gamut-mapping, many different attributes are changed simultaneously, and because of this the objective assessment is very complex. Previous research has also shown that IQ metrics have problems when multiple distortions occur simultaneously, as in gamut-mapping [63, 16]. This is not the case for TID2008 and some of the other databases evaluated here, since usually only one attribute at a time changes.

4.7 Overall Performance of the Metrics

None of the 22 evaluated metrics have a high correlation for all databases. If we discard the JPEG and JPEG2000 compressed images from Carraciolo et al. and the gamut-mapped images from Dugay et al., some metrics perform rather well, such as the PSNR-HVS-M and PSNR, which have a correlation of above 0.5 for the remaining databases. However, the performance seems to be database-dependent, and the metrics perform better for specific distortions, such as VIF for the IVC database with compression, and ABF for the luminance-changed images from Pedersen et al.

5

Conclusion

Today, the technological advances in the imaging field are rapid, and there is an increasing need to assess the quality of these advances. In this survey we have carried out a survey of full-reference image quality metrics. First, IQ metrics were classified into four groups: mathematically based metrics, low-level metrics, high-level metrics, and other metrics. Selected IQ metrics from each group have been explained in detail, and similarities and differences between the metrics have been discussed. This classification is helpful for understanding the foundation on which the metrics were created. IQ metrics from the different groups were selected in order to assess their performance, measured as the degree of correlation with the percept, against six state-of-the-art IQ databases. This evaluation is one of the most extensive carried out in the literature. The results of the evaluation show that none of the image metrics have a high performance over the six databases, but some metrics perform better than others for specific databases or subsets within a database.

A

Image Quality Metric Overview

In Table A.1 we show an overview of IQ metrics. The metrics are sorted in chronological order, then by the name of the metric if any name was given by the authors, and then by the author names in column three. In the type column the metrics are categorized by type: ID = image difference, IQ = image quality, IF = image fidelity, CD = color difference, HT = halftoning, DP = difference predictor, VQ = video quality, and IS = image similarity. For the column HVS, the metric must have an HVS model or CSF filter that simulates the HVS. Metrics like SSIM, which indirectly simulate the HVS, are assigned an HVS value of "no." MS indicates whether the metrics are multiscale. S/NS indicates whether the metric is spatial (FFS, FVS, VFS, VVS) or non-spatial (NS). We have divided the spatial metrics into four groups: FFS = fixed size of filter and fixed calculation; FVS = fixed size of filter and variable calculation; VFS = variable size of filter and fixed calculation; and VVS = variable size of filter and variable calculation. Fixed size indicates that the filter, block, or similar is fixed for the whole image; variable size indicates that the filter or block changes according to the image content. Fixed computation indicates that the same calculation is used within the filter or block; variable calculation indicates

that the calculation is dependent on the image content. C/G indicates whether the metric is for color or grayscale images. The test column indicates what kind of evaluation has been performed: either objective or subjective. Scenes indicates the number of scenes used in the original work where the metric was proposed, the same for modification and observers. For scenes, the first number is the number of scenes, while the second number, in parentheses, indicates the total number of images (originals $\times$ number of modifications). For observers, the total number of observers is stated, and inside the parentheses, the number of experts is stated if this information is given by the author, with a (A) in this column the observer was the author. A "—" indicates that this information is not available or not stated by the authors.

Table A.1. Overview of full-reference IQ metrics sorted according to year.

Year	Metric	Author(s)	Type	HVS	MS	S/NS	C/G	Test	Scenes	Modification	Observers	Comment
1976	ΔE^*_{ab}	CIE	CD	No	No	NS	Color	—	—	—	—	
1984	CMC	Clarke et al. [38]	CD	No	No	NS	Color	—	—	—	—	
1984		Näsänen [112]	HT	Yes	No	FFS	Gray	sub.	4 (128)	Dot profiles	3	1 observers in the main experiment, 2 in control experiment.
1986	SVF	Seim and Valberg[142]	CD	No	No	NS	Color	obj.	—		—	—
1987	BFD	Luo and Rigg [96]	CD	No	No	NS	Color	—	—	—	—	
1989		Safranek and Johnston [137]	IQ	No	No	FFS	Gray	obj.	30	—		
1990	SQRI	Barten [11]	IQ	No	No	FFS	Gray	sub.	5 (35)	Resolution	20	sub. data from Westerink and Roufs.
1991	LIPMSE	Bralilean et al. [20]	IF	Yes	No	FFS	Gray	sub.	1 (1)	Gaussian blur	1 (A)	Metric used to restore a blurred image.
1992		Mitsa and Varkur [102]	HT	Yes	No	VFS	Gray	sub.	11 (44)	Halftoning	12	3 metrics tested, same procedure but different CSFs.
1993	VDP	Daly [39]	DP	Yes	No	FFS	Gray	—	—	—	—	
1993		Watson [173]	IQ	No	No	FFS	Gray	obj.	2 (8)	—	—	
1993		Silverstein and Klein [152]	IQ	No	No	FFS	Gray	obj.	—	—	—	

(*Continued*)

Table A.1. (*Continued*)

Year	Metric	Author(s)	Type	HVS	MS	S/NS	C/G	Test	Scenes	Modification	Observers	Comment
1993		Lin [92]	HT	Yes	No	FFS	Gray	obj.	1 (5)	Halftoning	—	CSF filter from Mannos et al. and Nill et al. tested.
1993		Chaddha and Meng [26]	IQ	No	No	FFS	Gray	sub.	6 (48)	Compression artifacts	60	Also extended to video. Only the results from one scene presented.
1994		Mitsa and Alford [101]	HT	Yes	No	VFS	Gray	sub.	11 (44)	Halftoning		
1994	E_θ	Karunasekera and Kingsbury [76]	IQ	Yes	No	FFS	Gray	sub.	1 (8)	Lapped Orthogonal Transform	8	Observer expertise not stated.
1994		Teo and Heeger [159]	IQ	Yes	Yes	VFS	Color	both	1 (2)	—	(A)	
1995	ΔE_{94}	CIE	CD	No	No	NS	Color	—	—	—	—	Tested in various papers on various type of scenes.
1995	PDM	Heeger and Teo [64]	IF	Yes	Yes	VFS	Gray	sub.	1 (3)	JPEG	(A)	Extension of Teo and Heeger [159]
1995	PSPNR	Chou and Li [32]	IQ	Yes	No	FFS	Gray	obj.	—	—	—	
1995		Rushmeier et al. [136]	IS	Yes	No	FFS	Gray	obj.	Various	Synthetic images	—	Comparing real and synthetic images

(*Continued*)

Table A.1. (*Continued*)

Year	Metric	Author(s)	Type	HVS	MS	S/NS	C/G	Test	Scenes	Modification	Observers	Comment
1995	PEM	Westen et al. [174]	IQ	Yes	No	VFS	Gray	sub.	6(105)	PCM, DPCM, DCT and SBC coding at different bit rates	7 (5)	
1995	VDM	Lubin [93]	IQ	Yes	Yes	VFS	Gray	both	—	—	—	Various testing
1996	S-CIELAB	Zhang and Wandell [189]	ID, HT	Yes	No	FFS	Color	obj.	—	JPEG-DCT, halftoning and patterns	—	
1996	PQS	Miyahara et al. [103]	IQ	Yes	No	FFS	Gray	sub.	5(25)	global and local distortion	9 (9)	
1996	CMPSNR	Lambrecht and Farrell [84]	IQ	Yes	No	VFS	Color	sub.	1 (400)	JPEG	2	
1996		Scheermesser and Bryngdahl [140]	HT	No	No	VVS	Gray	obj.	2 (8)	Halftoning	—	Also tested on 2 test pattern.
1996	PQS	Miyahara et al. [103]	IQ	Yes	No	FFS	Gray	sub.	5 (25)	—	9 (9)	
1997	Sarnoff JND Vision Model	Lubin [94]	IQ	Yes	Yes	FFS	Color	sub.	5 (15)	MPEG-2 with different bit-rates	20	Also tested on JPEG data.
1997		Neumann et al. [113]	ID	No	No	VFS	Color	obj.	—	—	—	—
1997		Wong [176]	HT	No	No	FFS	Gray	—	—	—	—	
1997		Nijenhuis and Blommaert [114]	IQ	No	No	NS	Gray	sub.	2 (25)	Interpolation	6	
1997		Lai et al. [82]	IF	Yes	Yes	VFS	Color	sub.	1(1)	JPEG2000	(A)	
1997		Scheermesser and Bryngdahl [141]	HT	No	No	VVS	Gray	obj.	2 (—)	Halftoning	—	

(*Continued*)

Table A.1. (*Continued*)

Year	Metric	Author(s)	Type	HVS	MS	S/NS	C/G	Test	Scenes	Modification	Observers	Comment
1997	Δg	Wilson et al. [175]	IS	No	No	VFS	Gray	sub.	1 (5)	JPEG, different distortion types	(A)	
1998	IFA	Taylor et al. [158]	IF	Yes	Yes	FFS	Gray	obj.	—	—	—	—
1998	CVDM	Jin et al. [72]	ID	Yes	No	FFS	Color	obj.	—	—	—	
1998		Lai et al. [81]	IF	Yes	No	FFS	Color	sub.	1(1)	JPEG2000	(A)	
1998		Yu et al. [184]	HT	Yes	No	FFS	Color	sub.	6 (36)	Halftoning	8 (8)	
1998	KLK	Yu and Parker [183]	HT	No	No	FFS	Color	sub.	5 (20)	Halftoning	10	
1998		Veryovka et al. [163]	HT	Yes	Yes	VFS	Gray	obj.	1/1 (3/3)	Halftoning	—	—
1998	E	Yeh et al. [182]	VQ	Yes	No	FFS	Gray	sub.	1 (9)	Block artifacts	8	Sequence of 64 frames.
1998	CCETT visual metric	Pefferkorn and Blin [130]	IQ	Yes	No	FFS	Color	sub.	6 (30)	MPEG-2	16	
1998		Bolin and Meyer [14]	DP	Yes	No	FFS	Color	sub.	—	—	(A)	
1998		Franti [55]	IQ	No	No	FFS	Gray	sub.	3 (42)	Compression	15–39	Possiblilty for color images
1999	WVDP	Bradley [19]	DP	Yes	No	FFS	Gray	obj.	1 (3)	Noise	—	
1999	PDM	Avadhanam and Algazi [5]	IF	Yes	No	FFS	Color	sub.	5/5(50/75)	Compression	5(2)/5(2)	
1999	LSMB	Pappas and Neuhoff [118]	HT	Yes	No	FFS	Gray	—	—	—	—	Used for halftoning optimization
1999	QM_a	Nilsson [115]	HT	Yes	No	VVS	Gray	obj.	1(3)	Halftoning	—	
1999	PD	Nakauchi et al. [111]	ID	Yes	No	FFS	Color	sub.	8 (48)	Gamut mapping	10	Used to optimize gamut mapping

(*Continued*)

Table A.1. (*Continued*)

Year	Metric	Author(s)	Type	HVS	MS	S/NS	C/G	Test	Scenes	Modification	Observers	Comment
2000	DM and NQM	Damera-Venkata et al. [40]	IQ	No	No	FVS	Gray	sub.	—	Various	(A)	
2000		Farrugia and Peroche [51]	ID	Yes	No	FFS	Color	obj.	4(8)		(A)	
2001	ΔE_{00}	CIE [35]	CD	No	No	NS	Color	—	—	—	—	
2001	CIFA	Wu et al. [177]	ID	Yes	Yes	FFS	Color	sub.	1(1)	Hue	(A)	
2001	SCIELAB$_J$	Johnson and Fairchild [74]	IQ, HT	Yes	No	FFS	Color	sub.	1(72)	Sharpness	—	Extension of S-CIELAB Zhang and Wandell [189].
2001		Imai et al. [69]	CD	No	No	NS	Color	sub.	6(12)	—	—	
2001	SNR_W	Iordache and Beghdadi [70]	IS	No	No	FFS	Gray	sub.	1(3)	Salt-and-pepper noise, blurring, JPEG	5 (0)	
2002	iCAM	Fairchild and Johnson [50]	ID	Yes	No	FFS	Color	—	—	—	—	
2002		Hong and Luo [66]	ID	No	No	NS	Color	obj.	2	Local color change	—	
2002	UIQ	Wang and Bovik [165]	IQ	No	No	FFS	Gray	sub.	1 (8)	Different distortion types	22	
2002		Feng et al. [53]	IQ	Yes	No	FFS	Color	sub.	2 (14)	Halftoning		Extension of CVDM
2002	ΔI_{cm}	Morovic and Sun [106]	ID	No	No	FFS	Color	sub.	7 (32)	Gamut mapping	—	
2002	WMSE	Ayed et al. [7]	IQ	No	No	FFS	Gray	obj.	2 (14)	Noise	—	—
2002	HVS REAL	Avcibas et al. [6]	IQ	No	No	FFS	Gray	obj.	30 (420)	JPEG, SPHIT, noise, and blur	17	Metric build on work by Frese et al. [56].

(*Continued*)

Table A.1. (*Continued*)

Year	Metric	Author(s)	Type	HVS	MS	S/NS	C/G	Test	Scenes	Modification	Observers	Comment
2003	Q_{color}	Toet and Lucassen [161]	IF	No	No	FFS	Color	sub.	2 (21)	Quantization	4–16	
2003	FDP	Ferwerda and Pellacini [54]	DP	Yes	No	NS	Gray	sub.	8 (24)	Computer generated images	18	
2003	MSSIM	Wang et al. [172]	IQ	No	Yes	FFS	Gray	sub.	29 (344)	JPEG and JPEG2k	—	Scenes from LIVE
2003	M-SVD	Shnayderman et al. [149]	IQ	No	No	FFS	Gray	sub.	5 (30)	JPEG, JPEG2k, Gaussian noise, Gaussian blur, sharpening, DC shift	10 (5/5)	Color extension possible
2003	SNR_{WAV}	Beghdadi and Pesquet-Popescu [12]	IQ	No	No	FFS	Gray	sub.	1 (3)	Gaussian noise, JPEG and grid pattern	>25	
2003	DQM and CQM	de Freitas Zampolo and Seara [41]	IQ	No	No	FFS	Gray	sub.	4 (45)	Frequency distortion	7	
2003	Quality Assesser	Carnec et al. [25]	IQ	Yes	No	VFS	Color	sub.	— (90)	JPEG and JPEG2k	—	Also tested on LIVE
2004	B-CQM	de Freitas Zampolo and Seara [42]	IQ	No	No	NS	Gray	sub.	1 (81)	Frequency distortion and noise injection		
2004	SSIM	Wang et al. [167]	IQ	No	No	FFS	Gray	sub.	29 (344)	JPEG and JPEG2k	—	Scenes from LIVE
2004		Ivkovic and Sankar [71]	IQ	Yes	No	FFS	Gray	obj.	5 (20)	Contrast streching, white noise, blur, JPEG2k	—	
2004	NNET	Bouzerdoum et al. [17]	IQ	No	No	FFS	Gray	sub.	—	JPEG and JPEG2k	—	Scenes from LIVE
2004	I^*	McCormick-Goodhart et al. [98]	IQ	No	No	FFS	Color	sub.	—	2 printsystems	—	

(*Continued*)

Table A.1. (*Continued*)

Year	Metric	Author(s)	Type	HVS	MS	S/NS	C/G	Test	Scenes	Modification	Observers	Comment
2004	HDR-VDP	Mantiuk et al. [97]	DP	No	Yes	VFS	Gray	sub.	—	Quantization, noise	(A)	
2004	pdiff	Yee [181]	ID	Yes	Yes	FFS	Color	—	—	Film production	—	
2004	IFC	Sheikh et al. [146]	IF	No	No	FFS	Gray	sub.	29 (344)	JPEG, JPEG200, noise, blur	20–25	Scenes from LIVE
2005	CBM	Gao et al. [57]	IQ	No	No	FFS	Gray	sub.	29 (344)	JPEG and JPEG2k	—	Scenes from LIVE
2005	CWSSIM	Wang and Simoncelli [171]	IS, IQ	No	No	FFS	Gray	obj.	1 (12)	Distortion as JPEG, noise etc.	—	
2005	PQSI	Guarneri et al. [61]	IQ	No	No	—	Color	sub.	—	5 interpolation algorithms	—	
2005	M-DWT	Gayle et al. [58]	IQ	No	No	NS	Color	sub.	5 (30)	JPEG, JPEG2k, blur, noise, sharpening and DC-shift	14	Stated as a color metric, but only operates on luminance
2005	VQM_ESC	Yao et al. [180]	IQ	No	No	FFS	Gray	sub.	— (344)	JPEG and JPEG2k	—	Scenes from LIVE
2005	MHI	An et al. [4]	IQ	No	No	FFS	Gray	sub.	1 (4)	JPEG, JPEG2k, Gaussian noise and speckled noise	(A)	
2005		Kimmel et al. [79]	IS	No	Yes	FFS	Color	sub.	3 (3)	Gamut mapping	(A)	Used for gamut mapping optimization
2005		Xu et al. [178]	ID	No	No	NS	Color	sub.	—	Compression	(A)	

(*Continued*)

Table A.1. (*Continued*)

Year	Metric	Author(s)	Type	HVS	MS	S/NS	C/G	Test	Scenes	Modification	Observers	Comment
2005	NwMSE	Samet et al. [138]	IQ	Yes	No	FFS	Gray	sub.	—	JPEG2k, JPEG, and blurring	—	Scenes from LIVE
2006	VIF	Sheikh and Bovik [145]	IF	Yes	No	FFS	Gray	sub.	29 (344)	JPEG and JPEG2k	—	Scenes from LIVE
2006	WCWSSIM	Brooks and Pappas [22]	VQ	Yes	Yes	FFS	Color	sub.	3 (5)	Video compression and transmission distortion	—	Various testing of the metric.
2006	DTWT-SSIM	Lee et al. [91]	IS	No	No	FFS	Gray	obj.	10 (4860)	Blurring, scaling ,rotation and shift	—	Tested on handwritten data as a similarity measure.
2006	ESSIM	Chen et al. [29]	IQ	No	No	FFS	Gray	sub.	— (489)	JPEG2k, JPEG, and blurring	—	
2006	GSSIM	Chen et al. [30]	IQ	No	No	FFS	Gray	sub.	— (489)	JPEG2k, JPEG, and blurring	—	
2006	UQI-HVS	Egiazarian et al. [47]	IQ	Yes	No	FFS	Gray	sub.	2 (44)	Noise, blur, JPEG and JPEG2000	56	
2006	PSNR-HVS	Egiazarian et al. [47]	IQ	Yes	No	FFS	Gray	sub.	2 (44)	Noise, blur, JPEG and JPEG2000	56	
2006	PSNR-HVS-M	Egiazarian et al. [47]	IQ	Yes	No	FFS	Gray	sub.	2 (44)	Noise, blur, JPEG and JPEG2000	56	
2006		Mindru and Jung [100]	IQ	Yes	No	FFS	Color	sub.	1 (3)	Halftoning	(A)	

(*Continued*)

Table A.1. (*Continued*)

Year	Metric	Author(s)	Type	HVS	MS	S/NS	C/G	Test	Scenes	Modification	Observers	Comment
2006	$SSIM_{IPT}$	Bonnier et al. [16]	ID	No	No	FFS	Color	sub.	15 (90)	Gamut mapping	22	
2006	mPSNR	Munkberg et al. [109]	IQ	No	No	NS	Color	sub.	16 (—)	HDR	(A)	
2007	RCBM	Dong et al. [43]	IQ	No	No	FFS	Gray	obj.	29 (204)	JPEG		Scenes from LIVE
2007	SEME	Silva et al. [151]	IS	No	No	FFS	Gray	sub.	(233)	JPEG	—	Scenes from LIVE
2007	CISM	Lee et al. [90]	HT	Yes	No	FFS	Color	obj.	1 (28)	Halftoning	—	
2007	VSNR	Chandler and Hemami [28]	IF	Yes	No	FFS	Gray	sub.	29 (779)	—	—	Scenes from LIVE
2007	P-CIELAB (ΔPE)	Chou and Liu [33]	IF	No	No	VFS	Color	sub.	3 (6)	JND profile and JPEG	(A)	
2007	QMCS	Yao et al. [179]	IQ	No	No	FFS	Gray	sub.	29 (344)	JPEG and JPEG2k	—	Scenes from LIVE
2007	DÉCOR-WSNR	Wan et al. [164]	HT	Yes	No	FFS	Gray	obj.	1 (3)	Error diffusion	—	
2007	PSNR-HVS-M	Ponomarenko et al. [132]	IQ	YES	No	FFS	Gray	Sub.	1 (18)	Noise		
2008	DP	Granger [60]	CD	No	No	NS	Color	obj.	—			
2008		Lam and Loo [83]	IQ	No	No	NS	Gray	sub.	2 (6)	Noise		
2008	SBLC	Gorley and Holliman [59]	IQ	No	No	NS	Gray	sub.	3 (54)	JPEG compression	20	
2008	Busyness	Orfanidou et al. [117]	IQ	No	No	FFS	Gray	sub.	10 (80)	JPEG and JPEG2k compression	10	Psychophysical data from Allen et al. [3]
2008	Spatial ΔE_{00}	Chen et al. [31]	IQ	Yes	No	FFS	Color	obj.	1 (1)	Blurring	—	

(*Continued*)

Table A.1. (*Continued*)

Year	Metric	Author(s)	Type	HVS	MS	S/NS	C/G	Test	Scenes	Modification	Observers	Comment
2008	CED	Naguib et al. [110]	IQ	No	No	FFS	Gray	sub.	(227)	Various	—	—
2009	ΔE_E	Oleari et al. [116]	CD	No	No	FFS	Color	—	—	—	—	Various testing
2009	S-DEE	Simone et al. [155]	ID	Yes	No	FFS	Color	sub.	25 (1700)	—	—	Scenes from TID
2009	SHAME	Pedersen and Hardeberg [126]	IQ	Yes	No	FFS	Color	sub.	—	—	—	Various testing
2009	SHAME-II	Pedersen and Hardeberg [126]	IQ	Yes	No	FFS	Color	sub.	—	—	—	Various testing
2009	COMB	Bianco et al. [13]	IQ	Yes	No	FFS	Color	sub.	29 (779)	Various	—	LIVE database
2009	D	Rajashekar et al. [134]	IQ	No	No	FFS	Color	sub.	1(10)	Various	(A)	—
2009	ABF	Wang and Hardeberg [168]	ID	Yes	No	FFS	Color	sub.	10 (420)	Various	10	—
2010	PSNR_V	Zhao and Deng [191]	IQ	Yes	Yes	FFS	Gray	sub.	(169)	Various	—	Scenes from LIVE
2010	S_{DOG}-CIELAB	Ajagamelle et al. [1]	IQ	Yes	Yes	FFS	Color	sub.	—	Various	—	Various evaluation
2010	S_{DOG}-DEE	Ajagamelle et al. [1]	IQ	Yes	Yes	FFS	Color	sub.	—	Various	—	Various evaluation
2010		Cao et al.[23]	IQ	No	No	FFS	Color	sub.	19(38)	Gamut mapping	12	Expert observers
2010	PSNR-HVS-S	Tong et al. [162]	IQ	Yes	No	FFS	Color	sub.	—	Various	—	TID2008 and LIVE database

(*Continued*)

Table A.1. (*Continued*)

Year	Metric	Author(s)	Type	HVS	MS	S/NS	C/G	Test	Scenes	Modification	Observers	Comment
2010	PSNR-HVS-M-S	Tong et al. [162]	IQ	Yes	No	FFS	Color	sub.	—	Various	—	TID2008 and LIVE database
2010	RFSIM	Zhang et al. [186]	IQ	Yes	No	FFS	Gray	sub.	25 (1700)	—	838	TID2008
2010	M_{DOG}-DEE	Simone et al. [153]	ID	Yes	YES	FFS	Color	—	—	—	—	Various
2010	Simple	Zhang et al. [185]	IQ	No	YES	FFS	Gray	—	—	—	—	Five image quality databases
2011	TVD	Pedersen et al. [128]	IQ	Yes	No	FFS	Color	sub.	—	Various printers	—	Printed images from Pedersen et al. [129] and Pedersen et al. [122]
2011	FSIM	Zhang et al. [187]	IQ	Yes	No	FFS	Gray	sub.	—	Various	—	Six image quality databases
2011	$FSIM_C$	Zhang et al. [187]	IQ	Yes	No	FFS	Color	sub.	—	Various	—	Six image quality databases
2011	NSER	Mou et al. [107]	IQ	No	Yes	FFS	Gray	sub.	—	Various	—	Six image quality databases
2011	BBCQ	Shoham et al. [150]	IQ	No	No	FFS	Gray	sub.	—	Various	—	Two image sets with compression

Acknowledgments

We would like to thank Fritz Albregtsen, Nicolas Bonnier, Gabriele Simone, and Alessandro Rizzi for their advice and enlightening discussions regarding this survey.

References

[1] S. A. Ajagamelle, M. Pedersen, and G. Simone, "Analysis of the difference of gaussians model in image difference metrics," in *European Conference on Colour in Graphics, Imaging, and Vision (CGIV)*, pp. 489–496, Joensuu, Finland, June 2010.

[2] S. A. Ajagamelle, G. Simone, and M. Pedersen, "Performance of the difference of gaussian model in image difference metrics," in *Gjøvik Color Imaging Symposium*, number 4 in Høgskolen i Gjøviks rapportserie, (G. Simone, A. Rizzi, and J. Y. Hardeberg, eds.), pp. 27–30, Gjøvik, Norway, June 2009.

[3] E. Allen, S. Triantaphillidou, and R. E. Jacobson, "Image quality comparison between JPEG and JPEG2000. i. psychophysical investigation," *The Journal of Imaging Science and Technology*, vol. 3, no. 51, pp. 248–258, May/June 2007.

[4] K. An, J. Sun, and W. Du, "Homogeneity based image objective quality metric," in *Proceedings of International Symposium on Intelligent Signal Processing and Communication Systems*, pp. 649–652, Hong Kong, December 2005.

[5] N. Avadhanam and V. R. Algazi, "Evaluation of a human-vision-system-based image fidelity metric for image compressio," in *Applications of Digital Image Processing XXII*, vol. 3808 of *Proceedings of SPIE*, (A. G. Tescher, ed.), pp. 569–579, San Jose, CA, USA, 1999.

[6] I. Avcibas, B. Sankur, and K. Sayood, "Statistical evaluation of image quality measures," *Journal of Electronic Imaging*, vol. 11, pp. 206–223, 2002.

[7] M. A. B. Ayed, A. Samet, M. Loulou, and N. Masmoudi, "A new perceptual quality assessment for image coding," in *Proceedings of Visualization, Imaging, and Image Processing (VIIP2002)*, Malaga, Spain, September 2002.

[8] J. S. Babcock, J. B. Pelz, and M. D. Fairchild, "Eye tracking observers during rank order, paired comparison, and graphical rating tasks," in *Image Processing, Image Quality, Image Capture Systems Conference (PICS)*, pp. 10–15, Rochester, NY, May 2003.

[9] J. Bai, T. Nakaguchi, N. Tsumura, and Y. Miyake, "Evaluation of image corrected by retinex method based on S-CIELAB and gazing information," *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, vol. E89-A, no. 11, pp. 2955–2961, 2006.

[10] E. Bando, J. Y. Hardeberg, and D. Connah, "Can gamut mapping quality be predicted by color image difference formulae?," in *Human Vision and Electronic Imaging X*, vol. 5666 of *SPIE Proceedings*, (B. Rogowitz, T. N. Pappas, and S. Daly, eds.), pp. 180–191, San Jose, CA, USA, January 2005.

[11] P. G. J. Barten, "Evaluation of subjective image quality with the square-root integral method," *Journal of the Optical Society of America A*, vol. 7, pp. 2024–2031, October 1990.

[12] A. Beghdadi and B. Pesquet-Popescu, "A new image distortion measure based on wavelet decomposition," in *International Symposium on Signal Processing and Its Applications*, vol. 1, pp. 485–488, Paris, France, July 2003.

[13] S. Bianco, G. Ciocca, F. Marini, and R. Schettini, "Image quality assessment by preprocessing and full reference model combination," in *Image Quality and System Performance VI*, vol. 7242, (S. P. Farnand and F. Gaykema, eds.), p. 72420O, San Jose, CA, January 2009.

[14] M. Bolin and G. Meyer, "A perceptually based adaptive sampling algorithm," in *SIGGRAPH '98 Conference Proceedings*, pp. 409–418, Orlando, FL, July 1998.

[15] N. Bonnier, C. Leynadier, and F. Schmitt, "Improvements in spatial and color adaptive gamut mapping algorithms," in *European Conference on Colour in Graphics, Imaging, and Vision (CGIV)*, pp. 341–346, June 2008.

[16] N. Bonnier, F. Schmitt, H. Brettel, and S. Berche, "Evaluation of spatial gamut mapping algorithms," in *Color Imaging Conference*, pp. 56–61, Scottsdale, AZ, November 2006.

[17] A. Bouzerdoum, A. Havstad, and A. Beghdadi, "Image quality assessment using a neural network approach," in *Proceedings of the IEEE International Symposium on Signal Processing and Information Technology*, pp. 330–333, Rome, Italy, December 2004.

[18] S. Bouzit and L. MacDonald, "Colour difference metrics and image sharpness," in *Color Imaging Conference*, pp. 262–267, Scottsdale, AZ, November 2000.

[19] A. P. Bradley, "A wavelet visible difference predictor," *IEEE Transactions on Image Processing*, vol. 8, pp. 717–730, 1999.

[20] J. C. Brailean, B. J. Sullivan, C. T. Chen, and M. L. Giger, "Evaluating the EM algorithm for image processing using a human visual fidelity criterion," in *International Conference on Acoustics, Speech and Signal Processing (ICASSP'91)*, vol. 4, pp. 2957–2960, Toronto, Ontario, Canada, April 1991. doi: 10.1109/ICASSP.1991.151023.

[21] G. J. Braun, M. D. Fairchild, C. F. Carlson, and F. Ebner, "Color gamut mapping in a hue-linearized CIELAB color space," in *Color Imaging Conference*, pp. 163–168, Scottsdale, AZ, USA, November 1998.

[22] A. C. Brooks and T. N. Pappas, "Structural similarity quality metrics in a coding context: Exploring the space of realistic distortions," in *Human Vision and Electronic Imaging XI*, vol. 6057 of *SPIE Proceedings*, (B. E. Rogowitz, T. N. Pappas, and S. J. Daly, eds.), pp. 299–310, San Jose, CA, USA, February 2006. doi: 10.1117/12.660611.

[23] G. Cao, M. Pedersen, and Z. Barańczuk, "Saliency models as gamut-mapping artifact detectors," in *European Conference on Colour in Graphics, Imaging, and Vision (CGIV)*, pp. 437–443, Joensuu, Finland, June 2010.

[24] V. Caracciolo, "Just noticeable distortion evaluation in color images," Master's thesis, Gjøvik University College and Roma Tre University, 18/07/11: `http://colorlab.no/content/download/25901/274793/file/Caracciolo 2009_Master_Thesis.pdf`, 2009.

[25] M. Carnec, P. Le Callet, and D. Barba, "An image quality assessment method based on perception of structural information," in *IEEE International Conference on Image Processing (ICIP2003)*, pp. 185–188, Barcelona, Spain, September 2003.

[26] N. Chaddha and T. H. Y. Meng, "Psycho-visual based distortion measures for monochrome image and video compression," in *Proceedings of the Asilomar Conference on Signals, Systems and Computers*, vol. 2, pp. 841–845, Pacific Grove, CA, November 1993. doi: 10.1109/ACSSC.1993.342451.

[27] D. M. Chandler and S. S. Hemami, Online supplement to "VSNR: A visual signal-to-noise ratio for natural images based on near-threshold and suprathreshold vision," 18/02/11: `http://foulard.ece.cornell.edu/dmc27/vsnr/vsnr_a57.pdf`, 2007.

[28] D. M. Chandler and S. S. Hemami, "VSNR: A wavelet-based visual signal-to-noise ratio for natural images," *IEEE Transactions on Image Processing*, vol. 16, no. 9, pp. 2284–2298, September 2007.

[29] G. H. Chen, C. L. Yang, L. M. Po, and S. L. Xie, "Edge-based structural similarity for image quality assessment," in *Proceedings of International Conference in Acoustics, Speech and Signal Processing*, vol. 2, pp. 933–936, Toulouse, France, May 2006.

[30] G. H. Chen, C. L. Yang, and S. L. Xie, "Gradient-based structural similarity for image quality assessment," in *IEEE International Conference on Image Processing*, pp. 2929–2932, Atlanta, GA, USA, October 2006.

[31] S. Chen, A. Beghdadi, and A. Chetouani, "Color image assessment using spatial extension to CIE DE2000," in *The International Conference on Consumer Electronics (ICCE)*, pp. 1–2, Las Vegas, CA, USA, January 2008. ISBN: 978-1-4244-1458-1.

[32] C. Chou and Y. Li, "A perceptually tuned subband image coder based on the measure of just-noticeable-distortion profile," *IEEE Tranactions on Circuits and Systems for Video Technology*, vol. 5, pp. 467–476, 1995.

[33] C. Chou and K. Liu, "A fidelity metric for assessing visual quality of color images," in *Proceedings of International Conference on Computer Communications and Networks (ICCCN)*, pp. 1154–1159, Honolulu, HI, August 2007.

[34] CIE, "Industrial colour-difference evaluation," publication CIE 116-95, bureau central de la CIE, 1995.

[35] CIE, "CIE 142-2001: Improvement to industrial colour-difference evaluation," Technical Report, ISBN 978 3 901906 08 4, 2001.

[36] CIE, "CIE 15:2004: Colorimetry," Technical Report, 3rd edition, 2004.

[37] CIE, "Guidelines for the evaluation of gamut mapping algorithms," Technical Report ISBN: 3-901-906-26-6, CIE TC8-03, 2004.

[38] F. J. J. Clarke, R. McDonald, and B. Rigg, "Modification to the JPC 79 colour difference formula," *Journal of the Society of Dyers and Colourists*, vol. 100, pp. 128–132, April 1984.

[39] S. Daly, "The visible differences predictor: An algorithm for the assessment of image fidelity," in *Digital Images and Human Vision*, pp. 179–206, Cambridge, MA, USA: MIT Press, 1993. ISBN 0-262-23171-9.

[40] N. Damera-Venkata, T. D. Kite, W. S. Geisler, B. L. Evans, and A. C. Bovik, "Image quality assessment based on a degradation model," *IEEE Transactions on Image Processing*, vol. 9, pp. 636–650, 2000.

[41] R. de Freitas Zampolo and R. Seara, "A measure for perceptual image quality assessment," in *International Conference on Image Processing (ICIP)*, vol. 1, pp. 433–436, Barcelona, Spain, September 2003.

[42] R. de Freitas Zampolo and R. Seara, "Perceptual image quality assessment based on bayesian networks," in *International Conference on Image processing (ICIP2004)*, pp. 329–332, Singapore, October 2004.

[43] W. Dong, Q. Yu, C. N. Zhang, and H. Li, "Image quality assessment using rough fuzzy integrals," in *Proceedings of the International Conference on Distributed Computing Systems Workshops (ICDCSW'07)*, pp. 1–5, Washington, DC, USA, 2007. ISBN 0-7695-2881-3. doi: http://dx.doi.org/10.1109/ICDCSW.2007.114.

[44] F. Dugay, "Perceptual evaluation of colour gamut mapping algorithms," Master Thesis, Perceptual Evaluation of colour gamut mapping algorithms. Gjøvik University College and Grenoble Institute of Technology, 2007.

[45] F. Dugay, I. Farup, and J. Y. Hardeberg, "Perceptual evaluation of color gamut mapping algorithms," *Color Research and Application*, vol. 33, no. 6, pp. 470–476, December 2008.

[46] F. Ebner and M. D. Fairchild, "Development and testing of a color space (IPT) with improved hue uniformity," in *Information Science and Technology/SID Color Imaging Conference: Color Science, Systems and Applications*, vol. 6, pp. 8–13, November 1998. ISBN/ISSN: 0-89208-213-5.

[47] K. Egiazarian, J. Astola, N. Ponomarenko, V. Lukin, F. Battisti, and M. Carli, "Two new full-reference quality metrics based on HVS," in *Proceedings of the International Workshop on Video Processing and Quality Metrics*, pp. 22–24, Scottsdale, USA, January 2006.

[48] U. Engelke, M. Kusuma, H.-J. Zepernick, and M. Caldera, "Reduced-reference metric design for objective perceptual quality assessment in wireless imaging.," *Image Communications*, vol. 24, pp. 525–547, August 2009. ISSN 0923-5965.

[49] M. D. Fairchild, "Still photography throwdown: Silver halide vs. silicon," in *Color and Imaging Conference*, pp. 154–159, San Antonio, TX, November 2010.

[50] M. D. Fairchild and G. M. Johnson, "Meet iCAM: A next-generation color appearance model," in *Color Imaging Conference*, pp. 33–38, Scottsdale, AZ, November 2002.

[51] J. P. Farrugia and B. Peroche, "A perceptual image metric in computer graphics," in *International Conference on Color in Graphics and Image Processing (CGIP2000)*, pp. 13–17, Saint-Etienne, France, October 2000.

[52] I. Farup, C. Gatta, and A. Rizzi, "A multiscale framework for spatial gamut mapping," *IEEE Transactions on Image Processing*, vol. 16, no. 10, pp. 2423–2435, 2007.

[53] X.-F. Feng, J. Speigle, and A. Morimoto, "Halftone quality evaluation using color visual models," in *Image Processing, Image Quality, Image Capture, Systems Conference (PICS)*, pp. 5–10, Portland, Oregon, USA, April 2002.

[54] J. A. Ferwerda and F. Pellacini, "Functional difference predictors (FDPs): Measuring meaningful image differences," in *Conference Record of the Asilomar Conference on Signals, Systems and Computers*, vol. 2, pp. 1388–1392, Pacific Groove, CA, November 2003.

[55] P. Franti, "Blockwise distortion measure for statistical and structural errors in digital images," *Signal Processing: Image Communication*, vol. 13, pp. 89–98, 1998.

[56] T. Frese, C. A. Bouman, and J. P. Allebach, "A methodology for designing image similarity metrics based on human visual system models," Technical Report TR-ECE 97-2, Purdue University, West Lafayette, IN, USA., 1997.

[57] X. Gao, T. Wang, and J. Li, "A content-based image quality metric," in *Rough Sets, Fuzzy Sets, Data Mining, and Granular Computing*, vol. 3642 of *Lecture Notes in Computer Science*, pp. 231–240, Springer Berlin/Heidelberg, 2005.

[58] D. Gayle, H. Mahlab, Y. Ucar, and A. M. Eskicioglu, "A full-reference color image quality measure in the DWT domain," in *European Signal Processing Conference, EUSIPCO*, p. 4, Antalya, Turkey, September 2005.

[59] P. Gorley and N. Holliman, "Stereoscopic image quality metrics and compression," in *Stereoscopic Displays and Applications XIX*, vol. 6803, (A. J. Woods, N. S. Holliman, and J. O. Merritt, eds.), p. 680305, San Jose, CA, USA, January 2008.

[60] E. M. Granger, "A comparison of color difference data and formulas," in *TAGA Annual Technical Conference*, pp. 191–202, San Francisco, CA, USA, March 2008.

[61] I. Guarneri, M. Guarnera, A. Bosco, and G. Santoro, "A perceptual quality metric for color-interpolated images," in *Image Quality and System Performance II*, vol. 5668 of *SPIE Proceedings*, (R. Rasmussen and Y. Miyake, eds.), pp. 61–69, San Jose, CA, USA, January 2005.

[62] R. Halonen, M. Nuutinen, R. Asikainen, and P. Oittinen, "Development and measurement of the goodness of test images for visual print quality evaluation," in *Image Quality and System Performance VII*, vol. 7529, (S. P. Farnand and F. Gaykema, eds.), pp. 752909–1–10, San Jose, CA, USA, January 2010.

[63] J. Y. Hardeberg, E. Bando, and M. Pedersen, "Evaluating colour image difference metrics for gamut-mapped images," *Coloration Technology*, vol. 124, no. 4, pp. 243–253, August 2008.

[64] D. J. Heeger and P. C. Teo, "A model of perceptual image fidelity," in *Proceedings of the International Conference on Image Processing (ICIP)*, vol. 2, p. 2343, Washington, DC, USA, October 1995. ISBN 0-8186-7310-9.

[65] D. W. Hertel, "Exploring s-CIELAB as a scanner metric for print uniformity," in *Image Quality and System Performance II*, vol. 5668 of *SPIE Proceedings*, (R. Rasmussen and Y. Miyake, eds.), pp. 51–60, San Jose, CA, January 2005.

[66] G. Hong and M. R. Luo, "Perceptually based colour difference for complex images," in *Congress of the International Colour Association*, vol. 4421 of *SPIE Proceedings*, (R. Chung and A. Rodrigues, eds.), pp. 618–621, Rochester, NY, USA, June 2002.

[67] G. Hong and M. R. Luo, "New algorithm for calculating perceived colour difference of images," *Imaging Science Journal*, vol. 54, no. 2, pp. 86–91, 2006.

[68] Y. Horita, K. Shibata, and Y. Kawayoke, "Mict image quality evaluation database," 18/02/2011: `http://mict.eng.u-toyama.ac.jp/mictdb.html`.

[69] F. H. Imai, N. Tsumura, and Y. Miyake, "Perceptual color difference metric for complex images based on mahalanobis distance," *Journal of Electronic Imaging*, vol. 10, pp. 385–393, April 2001.

[70] R. Iordache and A. Beghdadi, "A wigner-ville distribution-based image dissimilarity measure," in *Proceedings of the International Symposium on Signal Processing and Its Applications (ISSPA '01)*, vol. 2, pp. 430–433, Kuala Lumpur, Malaysia, August 2001.

[71] G. Ivkovic and R. Sankar, "An algorithm for image quality assessment," in *International Conference on Acoustics, Speech, and Signal Processing, (ICASSP '04)*, pp. 713–716, Montreal, Canada, May 2004.

[72] E. W. Jin, X.-F. Feng, and J. Newell, "The development of a color visual difference model (CVDM)," in *Image Processing, Image Quality, Image Capture, Systems Conference*, pp. 154–158, Portland, Oregon, May 1998.

[73] E. W. Jin and S. Field, "A groundtruth database for testing objective metrics for image difference," in *Image Processing, Image Quality, Image Capture, Systems Conference (PICS)*, pp. 114–119, Rochester, NY, May 2003.

[74] G. M. Johnson and M. D. Fairchild, "Darwinism of color image difference models," in *Color Imaging Conference*, pp. 108–112, Scottsdale, AZ, November 2001.

[75] G. M. Johnson and M. D. Fairchild, "On contrast sensitivity in an image difference model," in *Image Processing, Image Quality, Image Capture, Systems Conference (PICS)*, pp. 18–23, Portland, OR, April 2002.

[76] S. A. Karunasekera and N. G. Kingsbury, "A distortion measure for image artifacts based on human visual sensitivity," in *Proceedings of the International Conference on Acoustics, Speech and Signal Processing*, vol. 5, pp. 117–120, Adelaide, SA, Australia, April 1994.

[77] J.-S. Kim, M.-S. Cho, and B.-K. Koo, "Experimental approach for human perception based image quality assessment," in *International Conference on Entertainment Computing*, vol. 4161 of *Lecture Notes in Computer Science*, (R. Harper, M. Rauterberg, and M. Combetto, eds.), pp. 59–68, Cambridge, UK, September 2006. Springer Berlin/Heidelberg.

[78] J.-S. Kim, M.-S. Cho, S. Westland, and M. R. Luo, "Image quality assessment for photographic images," in *AIC Colour Congress of the International Colour Association*, pp. 1095–1098, Granada, Spain, May 2005.

[79] R. Kimmel, D. Shaked, M. Elad, and I. Sobel, "Space-dependent color gamut mapping: A variational approach," *IEEE Transactions on Image Processing*, vol. 14, no. 6, pp. 796–803, 2005.

[80] Ø. Kolås and I. Farup, "Efficient hue-preserving and edge-preserving spatial gamut mapping," in *Color Imaging Conference*, pp. 207–212, Albuquerque, New Mexico, November 2007.

[81] Y.-K. Lai, J. Guo, and C.-C. J. Kuo, "Perceptual fidelity measure of digital color images," in *Human Vision and Electronic Imaging III*, vol. 3299 of *Proceedings of SPIE*, (B. E. Rogowitz and T. N. Pappas, eds.), pp. 221–231, San Jose, CA, July 1998.

[82] Y.-K. Lai, C.-C. J. Kuo, and J. Li, "New image compression artifact measure using wavelets," in *Visual Communications and Image Processing*, vol. 3024 of *Proceedings of SPIE*, (J. Biemond and E. J. Delp, eds.), pp. 897–908, San Jose, CA, January 1997.

[83] E. P. Lam and K. C. Loo, "An image similarity measure using homogeneity regions and structure," in *Image Quality and System Performance V*, vol. 6808 of *Proceedings of SPIE*, (S. P. Farnand and F. Gaykema, eds.), pp. 680811–680811–9, San Jose, CA, January 2008.

[84] C. Lambrecht and J. E. Farrell, "Perceptual quality metric for digitally coded color images," in *VIII European Signal Processing Conference EUSIPCO,* IEEE Proceedings, pp. 1175–1178, Trieste, Italy, September 1996.

[85] E. C. Larson and D. M. Chandler, "Unveiling relationships between regions of interest and image fidelity metrics," in *Visual Communications and Image Processing*, vol. 6822 of *SPIE Proceedings*, (W. A. Pearlman, J. W. Woods, and L. Lu, eds.), pp. 68222A–68222A–16, San Jose, CA, January 2008.

[86] E. C. Larson and D. M. Chandler, "Most apparent distortion: A dual strategy for full-reference image quality assessment," in *Image Quality and System Performance VI*, vol. 7242 of *Proceedings of SPIE*, (S. Farnand and F. Gaykema, eds.), p. 72420S, San Jose, CA, January 2009.

[87] E. C. Larson and D. M. Chandler, "Most apparent distortion: Full-reference image quality assessment and the role of strategy," *Journal of Electronic Imaging*, vol. 19, no. 1, p. 011006, March 2010.

[88] P. Le Callet and F. Autrusseau, "Subjective quality assessment IRCCyN/IVC database," 18/07/11: `http://www.irccyn.ec-nantes.fr/ivcdb/`, 2005.

[89] P. Le Callet and D. Barba, "A robust quality metric for color image quality assessment," in *International Conference on Image Processing (ICIP)*, vol. 1, pp. 437–440, Barcelona, Spain, September 2003.

[90] J. Lee, T. Horiuchi, R. Saito, and H. Kotera, "Digital color image halftone: Hybrid error diffusion using the mask perturbation and quality verification," *The Journal of Imaging Science and Technology*, vol. 51, no. 5, pp. 391–401, September/October 2007.

[91] M.-S. Lee, L.-Y. Liu, and F.-S. Lin, "Image similarity comparison using dual-tree wavelet transform," in *Pacific Rim Symposium Advances in Image*

and Video Technology, vol. 4319 of *Lecture Notes in Computer Science*, pp. 189–197, Hsinchu, Taiwan, December 2006.

[92] Q. Lin, "Halftone image quality analysis based on a human vision model," in *Human Vision, Visual Processing, and Digital Display IV*, vol. 1913 of *Proceedings of SPIE*, (J. P. Allebach and B. E. Rogowitz, eds.), pp. 378–389, San Jose, CA, September 1993.

[93] J. Lubin, "Vision models for target detection and recognition," in *Chapter A Visual Discrimination Model for Imaging Systems Design and Evaluation*, pp. 245–283, River Edge, NJ, USA: World Scientific Publishing Co, Inc., 1995. (Singapore: World Scientific).

[94] J. Lubin, "Sarnoff JND vision model: Algorithm description and testing," Technical Report, Sarnoff Corporation, 5/10/2007: `ftp://ftp.its.bldrdoc.gov/dist/ituvidq/old2/jrg003.rtf`, 1997.

[95] M. R. Luo, G. Cui, and B. Rigg, "The development of the CIE 2000 colour-difference formula: CIEDE2000," *Color Research and Application*, vol. 26, no. 5, pp. 340–350, 2001.

[96] M. R. Luo and B. Rigg, "BFD(l:c) colour-difference formula: Part 1 — development of the formula," *Journal of the Society of Dyers and Colourists*, vol. 103, pp. 86–94, 1987.

[97] R. Mantiuk, K. Myszkowski, and H.-P. Seidel, "Visible difference predicator for high dynamic range images," in *International Conference on Systems, Man and Cybernetics*, vol. 3, pp. 2763–2769, Hague, Netherlands, October 2004.

[98] M. McCormick-Goodhart, H. Wilhelm, and D. Shklyarov, "A 'retained image apparance' metric for full tonal scale, colorimetric evaluation of photographic image standard stability," in *NIP20: International Conference on Digital Printing Technologies*, pp. 680–688, Salt Lake City, UT, October 2004.

[99] G. Menegaz, A. Le Troter, J. Sequeira, and J. M. Boi, "A discrete model for color naming," *EURASIP Journal on Advances in Signal Processing*, vol. 1, p. 10, 2007.

[100] F. Mindru and J. Jung, "Structure and color based comparison of halftone images," in *International Congress of Imaging Science (ICIS06)*, pp. 629–632, Rochester, New York, May 2006.

[101] T. Mitsa and J. R. Alford, "Single-channel versus multiple-channel visual models for the formulation of image quality measures in digital halftoning," in *Recent Progress in Digital Halftoning*, vol. 1, (R. Eschbach, ed.), pp. 14–16, 1994.

[102] T. Mitsa and K. L. Varkur, "Evaluation of contrast sensitivity functions for the formulation of quality measures incorporated in halftoning algorithms," in *International Conference on Acostics, Speech, Signal Processing: Image and Multidimensional Signal Processing*, vol. 5, pp. 301–304, Minneapolis, MN, USA, April 1993.

[103] M. Miyahara, K. Kotani, and R. Algazi, "Objective picture quality scale (PQS) for image coding," *IEEE Transactions on Communications*, vol. 46, no. 9, pp. 1215–1226, September 1996.

[104] N. Moroney, "A radial sampling of the OSA uniform color scales," in *Color Imaging Conference: Color Science and Engineering Systems, Technologies, and Applications*, pp. 175–180, Scottsdale, AZ, November 2003.

[105] J. Morovic and M. R. Luo, "The fundamentals of gamut mapping: A survey," *Journal of Imaging Science and Technology*, vol. 45, no. 3, pp. 283–290, 2001.

[106] J. Morovic and P. Sun, "Visual differences in colour reproduction and their colorimetric correlates," in *Color Imaging Conference*, pp. 292–297, Scottsdale, AZ, 2002.

[107] X. Mou, M. Zhang, W. Xue, and L. Zhang, "Image quality assessment based on edge," in *Digital Photography VII*, vol. 7876 of *Proceedings of SPIE*, (F. H. Imai, F. Xiao, J. M. DiCarlo, N. Sampat, and S. Battiato, eds.), p. 78760N, San Francisco, CA, January 2011.

[108] J. A. Movshon and L. Kiorpes, "Analysis of the development of spatial sensitivity in monkey and human infants," *Journal of Optical Society of America A*, vol. 5, pp. 2166–2172, December 1988.

[109] J. Munkberg, P. Clarberg, J. Hasselgren, and T. Akenine-Möller, "High dynamic range texture compression for graphics hardware," *ACM Transactions on Graphics*, vol. 25, no. 3, pp. 698–706, July 2006.

[110] N. A. Naguib, A. E. Hussein, H. A. Keshk, and M. I. El-Adawy, "Contrast error distribution measurement for full reference image quality assessment," in *International Conference on Computer Theory and Applications*, p. 4, Alexandria, Egypt, October 2008.

[111] S. Nakauchi, S. Hatanaka, and S. Usui, "Color gamut mapping based on a perceptual image difference measure," *Color Research and Application*, vol. 24, pp. 280–291, 1999.

[112] R. Näsänen, "Visibility of halftone dot textures," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 14, pp. 920–924, 1984.

[113] L. Neumann, K. Matkovic, and W. Purgathofer, "Perception based color image difference," Technical Report TR-186-2-97-21, Institute of Computer Graphics and Algorithms, Vienna University of Technology, Vienna, Austria, December 1997.

[114] M. R. M. Nijenhuis and F. J. J. Blommaert, "Perceptual-error measure and its application to sampled and interpolated single-edged images," *Journal of the Optical Society of America A*, vol. 14, pp. 2111–2127, September 1997.

[115] F. Nilsson, "Objective quality measures for halftoned images," *Journal of the Optical Society of America A*, vol. 16, pp. 2151–2162, 1999.

[116] C. Oleari, M. Melgosa, and R. Huertas, "Euclidean color-difference formula for small-medium color differences in log-compressed OSA-UCS space," *Journal of the Optical Society of America A*, vol. 26, no. 1, pp. 121–134, 2009.

[117] M. Orfanidou, S. Triantaphillidou, and E. Allen, "Predicting image quality using a modular image difference model," in *Image Quality and System Performance V*, vol. 6808 of *SPIE Proceedings*, (S. P. Farnand and F. Gaykema, eds.), pp. 68080F–68080F–12, San Jose, CA, January 2008.

[118] T. N. Pappas and D. L. Neuhoff, "Least-squares model-based halftoning," *IEEE Transactions on Image Processing*, vol. 8, no. 8, pp. 1102–1116, August 1999.

[119] D. Pascale, "AN-7 the optical society of america uniform color scales (OSA UCS)," Technical Report, Babelcolor, 21/01/10: `http://www.babelcolor.com/download/AN-7%20The%20OSA%20UCS.pdf`, September 2009.

[120] M. Pedersen, "Importance of region-of-interest on image difference metrics," Master's Thesis, Gjøvik University College, 2007.

[121] M. Pedersen, "111 full-reference image quality metrics and still not good enough?," in *Proceedings from Gjøvik Color Imaging Symposium 2009*, number 4 in Høgskolen i Gjøviks rapportserie, (G. Simone, A. Rizzi, and J. Y. Hardeberg, eds.), p. 4, Gjøvik, Norway, June 2009.

[122] M. Pedersen, N. Bonnier, J. Y. Hardeberg, and F. Albregtsen, "Estimating print quality attributes by image quality metrics," in *Color and Imaging Conference*, pp. 68–73, San Antonio, TX, USA, November 2010.

[123] M. Pedersen and J. Y. Hardeberg, "Rank order and image difference metrics," in *European Conference on Colour in Graphics, Imaging, and Vision (CGIV)*, pp. 120–125, Terrassa, Spain, June 2008.

[124] M. Pedersen and J. Y. Hardeberg, "SHAME: A new spatial hue angle metric for perceptual image difference," *Journal of Vision*, vol. 9, no. 8, pp. 343, 8, 2009. ISSN 1534-7362.

[125] M. Pedersen and J. Y. Hardeberg, "Survey of full-reference image quality metrics," Høgskolen i Gjøviks rapportserie 5, The Norwegian Color Research Laboratory (Gjøvik University College), ISSN: 1890-520X, June 2009.

[126] M. Pedersen and J. Y. Hardeberg, "A new spatial hue angle metric for perceptual image difference," in *Computational Color Imaging*, vol. 5646 of *Lecture Notes in Computer Science*, pp. 81–90, Saint Etienne, France, March 2009.

[127] M. Pedersen, J. Y. Hardeberg, and P. Nussbaum, "Using gaze information to improve image difference metrics," in *Human Vision and Electronic Imaging VIII*, vol. 6806 of *SPIE Proceedings*, (B. Rogowitz and T. Pappas, eds.), p. 680611, San Jose, CA, USA, January 2008.

[128] M. Pedersen, G. Simone, M. Gong, and I. Farup, "A total variation based color image quality metric with perceptual contrast filtering," in *International Conference on Pervasive Computing, Signal Processing and Applications*, Gjøvik, Norway, September 2011.

[129] M. Pedersen, Y. Zheng, and J. Y. Hardeberg, "Evaluation of image quality metrics for color prints," in *Scandinavian Conference on Image Analysis*, vol. 6688 of *Lecture Notes in Computer Science*, (A. Heyden and F. Kahl, eds.), pp. 317–326, Ystad Saltsjöbad, Sweden, May 2011.

[130] S. Pefferkorn and J.-L. Blin, "Perceptual quality metric of color quantization errors on still images," in *Human Vision and Electronic Imaging III*, vol. 3299 of *SPIE Proceedings*, (B. E. Rogowitz and T. N. Pappas, eds.), pp. 210–220, San Jose, CA, January 1998.

[131] N. Ponomarenko, V. Lukin, K. Egiazarian, J. Astola, M. Carli, and F. Battisti, "Color image database for evaluation of image quality metrics," in *International Workshop on Multimedia Signal Processing*, pp. 403–408, Cairns, Queensland, Australia, October 2008.

[132] N. Ponomarenko, F. Silvestri, K. Egiazarian, M. Carli, J. Astola, and V. Lukin, "On between-coefficient contrast masking of DCT basis functions," in *International Workshop on Video Processing and Quality Metrics for Consumer Electronics VPQM-07*, pp. 1–4, Scottsdale, Arizona, USA, January 2007.

[133] G. Qiu and A. Kheiri, "Social image quality," in *Image Quality and System Performance VIII*, vol. 7867 of *Proceedings of SPIE*, (S. P. Farnand and F. Gaykema, eds.), p. 78670S, San Francisco, CA, January 2011.

[134] U. Rajashekar, Z. Wang, and E. P. Simoncelli, "Quantifying color image distortions based on adaptive spatio-chromatic signal decompositions," in *Internatial Conference on Image Processing*, pp. 2213–2216, Cairo, Egypt, November 2009.

[135] D. L. Ruderman, T. W. Cronin, and C.-C. Chiao, "Statistics of cone responses to natural images: Implications for visual coding," *Journal of the Optical Society of America A*, vol. 15, no. 8, pp. 2036–2045, 1998.

[136] H. Rushmeier, G. Ward, C. Piatko, P. Sanders, and B. Rust, "Comparing real and synthetic images: Some ideas about metrics," in *Eurographics Workshop on Rendering Techniques*, pp. 82–91, Dublin, Ireland, June 1995. Springer-Verlag, London, UK.

[137] R. J. Safranek and J. D. Johnston, "A perceptually tuned sub-band image coder with image dependent quantization and post-quantization data compression," in *International Conference on Acoustics, Speech, and Signal Processing*, vol. 3, pp. 1945–1948, Glasgow, UK, May 1989.

[138] A. Samet, M. A. B. Ayed, N. Masmoudi, and L. Khriji, "New perceptual image quality assessment metric," *Asian Journal of Information Technology*, vol. 4, no. 11, pp. 996–1000, 2005.

[139] C. Sano, T. Song, and M. R. Luo, "Colour differences for complex images," in *Color Imaging Conference: Color Science and Engineering Systems, Technologies, Applications*, pp. 121–126, Scottsdale, Arizona, November 2003. ISBN/ISSN: 0-89208-248-8.

[140] T. Scheermesser and O. Bryngdahl, "Texture metric of halftone images," *Journal of the Optical Society of America A*, vol. 13, pp. 18–24, 1996.

[141] T. Scheermesser and O. Bryngdahl, "Spatially dependent texture analysis and control in digital halftoning," *Journal of the Optical Society of America A*, vol. 14, pp. 827–835, 1997.

[142] T. Seim and A. Valberg, "Towards a uniform color space: A better formula to describe the munsell and OSA color scales," *Color Research and Application*, vol. 11, no. 1, pp. 11–24, 1986.

[143] K. Seshadrinathan and A. C. Bovik, "The encyclopedia of multimedia," in *Chapter Image and Video Quality Assessment*, pp. 8–17, Springer-Verlag, 2009.

[144] G. Sharma, *Digital Color Imaging Handbook*. Boca Raton, FL, USA: CRC Press, Inc., 2002.

[145] H. R. Sheikh and A. C. Bovik, "Image information and visual quality," *IEEE Transactions on Image Processing*, vol. 15, no. 2, pp. 430–444, 2006.

[146] H. R. Sheikh, A. C. Bovik, and G. de Veciana, "An information fidelity criterion for image quality assessment using natural scene statistics," *IEEE Transactions on Image Processing*, vol. 14, no. 12, pp. 2117–2128, December 2004.

[147] H. R. Sheikh, M. F. Sabir, and A. C. Bovik, "A statistical evaluation of recent full reference image quality assessment algorithms," *IEEE Transactions on Image Processing*, vol. 15, no. 11, pp. 3440–3451, 2006.

[148] H. R. Sheikh, Z. Wang, L. K. Cormack, and A. C. Bovik, "LIVE image quality assessment database release 2," 18/07/11: `http://live.ece.utexas.edu/research/quality`, 2007.

[149] A. Shnayderman, A. Gusev, and A. M. Eskicioglu, "A multidimensional image quality measure using singular value decomposition," in *Image Quality and System Performance*, vol. 5294 of *Proceedings of SPIE*, (Y. Miyake and D. R. Rasmussen, eds.), pp. 82–92, San Jose, CA, USA, December 2004. doi: 10.1117/12.530554.

[150] T. Shoham, D. Gill, and S. Carmel, "A novel perceptual image quality measure for block based image compression," in *Image Quality and System Performance VIII*, vol. 7867 of *Proceedings of SPIE*, (S. P. Farnand and F. Gaykema, eds.), p. 786709, San Francisco, CA, January 2011.

[151] E. A. Silva, K. Panetta, and S. S. Agaian, "Quantifying image similarity using measure of enhancement by entropy," in *Mobile Multimedia/Image Processing for Military and Security Applications*, vol. 6579, (S. S. Agaian and S. A. Jassim, eds.), pp. 65790U.1–65790U.12, San Jose, CA, USA, January 2007.

[152] D. A. Silverstein and S. A. Klein, "DCT image fidelity metric and its application to a text-based scheme for image display," in *Human Vision, Visual Processing, and Digital Display IV*, vol. 1913 of *Proceedings of SPIE*, (J. P. Allebach and B. E. Rogowitz, eds.), pp. 229–239, San Jose, CA, September 1993.

[153] G. Simone, V. Caracciolo, M. Pedersen, and F. A. Cheikh, "Evaluation of a difference of gaussians based image difference metric in relation to perceived compression artifacts," in *Advances in Visual Computing — International Symposium, Lecture Notes in Computer Science*, pp. 491–500, Las Vegas, NV, November 2010.

[154] G. Simone, C. Oleari, and I. Farup, "An alternative color difference formula for computing image difference," in *Proceedings from Gjøvik Color Imaging Symposium 2009*, number 4 in Høgskolen i Gjøviks rapportserie, (G. Simone, A. Rizzi, and J. Y. Hardeberg, eds.), pp. 8–11, Gjøvik, Norway, June 2009. URL `http://brage.bibsys.no/hig/bitstream/URN:NBN:no-bibsys_brage_9313/3/sammensatt_elektronisk.pdf`.

[155] 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," in *Congress of the International Colour Association (AIC)*, Sydney, Australia, September/October 2009.

[156] T. Song and M. R. Luo, "Testing color-difference formulae on complex images using a CRT monitor," in *Color Imaging Conference*, pp. 44–48, Scottsdale, AZ, November 2000.

[157] Y. Tadmor and D. J. Tolhurst, "Calculating the contrasts that retinal ganglion cells and lgn neurones encounter in natural scenes," *Vision Research*, vol. 40, pp. 3145–3157, 2000.

[158] C. C. Taylor, J. P. Allebach, and Z. Pizlo, "The image fidelity assessor," in *Image Processing, Image Quality, Image Capture, Systems Conference (PICS)*, pp. 237–241, Portland, Oregon, USA, May 1998.

[159] P. C. Teo and D. J. Heeger, "Perceptual image distortion," in *International Conference Image Processing*, vol. 2, pp. 982–986, Austin, TX, November 1994.

[160] K.-H. Thung and P. Raveendran, "A survey of image quality measures," in *International Conference for Technical Postgraduates (TECHPOS)*, pp. 1–4, Kuala Lumpur, Malaysia, December 2009.

[161] A. Toet and M. P. Lucassen, "A new universal colour image fidelity metric," *Displays*, vol. 24, pp. 197–204, 2003.

[162] Y. Tong, H. Konik, F. A. Cheikh, and A. Tremeau, "Full reference image quality assessment based on saliency map analysis," *Journal of Imaging Science and Technology*, vol. 54, no. 3, p. 030503, 2010.

[163] O. Veryovka, A. Fournier, and J. W. Buchanan, "Multiscale edge analysis of halftoned images," in *Human Vision and Electronic Imaging III*, vol. 3299 of *SPIE Proceedings*, (B. E. Rogowitz and T. N. Pappas, eds.), pp. 461–472, San Jose, CA, USA, January 1998.

[164] X. Wan, D. Xie, and J. Xu, "Quality evaluation of the halftone by halftoning algorithm-based methods and adaptive method," in *Image Quality and System Performance IV*, vol. 6494 of *SPIE Proceedings*, (L. C. Cui and Y. Miyake, eds.), p. 64940U, San Jose, CA, USA, January 2007.

[165] Z. Wang and A. C. Bovik, "A universal image quality index," *IEEE Signal Processing Letters*, vol. 9, pp. 81–84, 2002.

[166] Z. Wang and A. C. Bovik, *Modern Image Quality Assessment*. Morgan & Claypool Publishers, 2006.

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

[168] Z. Wang and J. Y. Hardeberg, "An adaptive bilateral filter for predicting color image difference," in *Color Imaging Conference*, pp. 27–31, Albuquerque, NM, USA, November 2009.

[169] Z. Wang and M. R. Luo, "Experimental filters for estimating image differences," in *CGIV 2008 — European Conference on Color in Graphics, Imaging and Vision*, pp. 112–115, Terrassa, Barcelona, Spain, June 2008.

[170] Z. Wang and E. P. Simoncelli, "Reduced-reference image quality assessment using a wavelet-domain natural image statistic model," in *Human Vision and Electronic Imaging X*, vol. 5666 of *Proceedings of SPIE*, (B. Rogowitz, T. N. Pappas, and S. J. Daly, eds.), pp. 149–159, San Jose, CA, USA, January 2005. doi: 10.1117/12.597306.

[171] Z. Wang and E. P. Simoncelli, "Translation insensitive image similarity in complex wavelet domain," in *International Conference on Acoustics, Speech and Signal Processing*, vol. 2, pp. 573–576, Philadelphia, PA, March 2005.

[172] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multi-scale structural similarity for image quality assessment," in *Asilomar Conference on Signals, Systems and Computers*, vol. 2, pp. 1398–1402, Pacific Grove, CA, November 2003.

[173] A. B. Watson, "DCT quantization matrices visually optimized for individual images," in *Human Vision, Visual Processing, and Digital Display IV*, vol. 1913 of *SPIE Proceedings*, (J. P. Allebach and B. E. Rogowitz, eds.), pp. 202–216, San Jose, CA, USA, February 1993.

[174] S. J. P. Westen, R. L. Lagendijk, and J. Biemond, "Perceptual image quality based on a multiple channel HVS model," in *International Conference on Acoustics, Speech and Signal Processing*, vol. 4, pp. 2351–2354, Detroit, MI, May 1995.

[175] D. L. Wilson, A. J. Baddeley, and R. A. Owens, "A new metric for grey-scale image comparison," *International Journal of Computer Vision*, vol. 24, pp. 5–17, 1997.

[176] P. W. Wong, "A mixture distortion criterion for halftones," Technical Report, HP Labs, 18/07/11: `http://www.hpl.hp.com/techreports/97/HPL-97-49.html`, March 1997.

[177] W. Wu, Z. Pizlo, and J. P. Allebach, "Color image fidelity assessor," in *Image Processing, Image Quality, Image Capture, Systems Conference (PICS)*, pp. 148–152, Monreal, Quebec, Canada, April 2001.

[178] R. Xu, S. N. Pattanaik, and C. E. Hughes, "High-dynamic-range still-image encoding in JPEG2000," *IEEE Computer Graphics and Applications*, vol. 25, no. 6, pp. 57–64, 2005.

[179] S. Yao, W. Lin, Z. K. Lu, E. P. Ong, M. H. Locke, and S. Q. Wu, "Image quality measure using curvature similarity," in *International Conference on Image Processing*, vol. 3, pp. 437–440, San Antonio, TX, September 2007. doi: 10.1109/ICIP.2007.4379340.

[180] S. Yao, W. Lin, S. Rahardja, X. Lin, E. P. Ong, Z. K. Lu, and X. K. Yang, "Perceived visual quality metric based on error spread and contrast," in *International Symposium on Circuits and Systems*, vol. 4, pp. 3793–3796, Kobe, Japan, May 2005.

[181] H. Yee, "A perceptual metric for production testing," *Journal of Graphics, GPU, and Game Tools*, vol. 9, no. 4, pp. 33–40, 2004.

[182] E. M. Yeh, A. C. Kokaram, and N. G. Kingsbury, "A perceptual distortion measure for edge-like artifacts in image sequences," in *Human Vision and Electronic Imaging III*, vol. 3299 of *SPIE Proceedings*, (T. N. Pappas and B. E. Rogowitz, eds.), pp. 160–172, San Jose, CA, January 1998.

[183] Q. Yu and K. J. Parker, "Quality issues in blue noise halftoning," in *Color Imaging: Device-Independent Color, Color Hardcopy, and Graphic Arts III*, vol. 3300 of *SPIE Proceedings*, (G. B. Beretta and R. Eschbach, eds.), pp. 376–385, San Jose, CA, January 1998.

[184] Q. Yu, K. J. Parker, R. Buckley, and V. Klassen, "A new metric for color halftone visibility," in *Image Processing, Image Quality, Image Capture, Systems Conference (PICS)*, pp. 226–230, Portland, OR, USA, May 1998.

[185] F. Zhang, S. Li, L. Ma, and K. N. Ngan, "Limitation and challenges of image quality measurement," in *Visual Communications and Image Processing Conference*, vol. 7744 of *Proceedings of SPIE*, (P. Frossard, H. Li, F. Wu, B. Girod, S. Li, and G. Wei, eds.), pp. 774402–774402–8, Huang Shan, China, July 2010.

[186] L. Zhang, L. Zhang, and X. Mou, "RFSIM: A feature based image quality assessment metric using riesz transforms," in *Internatonal Conference on Image Processing*, pp. 321–324, Hong Kong, September 2010.

[187] L. Zhang, L. Zhang, X. Mou, and D. Zhang, "FSIM: A feature similarity index for image quality assessment," *IEEE Transactions on Image Processing*, 2011.

[188] 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," in *COMPCON97 Digest of Papers*, pp. 44–48, Washington, DC, USA, 1997.

[189] X. Zhang and B. A. Wandell, "A spatial extension of CIELAB for digital color image reproduction," in *Soc. Inform. Display 96 Digest*, pp. 731–734, San Diego,CA, 1996.

[190] 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*, vol. 70, pp. 201–214, 1998.

[191] B. Zhao and C. Deng, "Image quality evaluation method based on human visual system," *Chinese Journal of Electronics*, vol. 19, no. 1, pp. 129–132, January 2010.

[192] P. Zolliker and K. Simon, "Adding local contrast to global gamut mapping algorithms," in *European Conference on Colour in Graphics, Imaging, and Vision (CGIV)*, pp. 257–261, Leeds, UK, June 2006.