Høgskolen i Gjøviks rapportserie, 2009 nr. 5

Survey of full-reference image quality metrics

Marius Pedersen and Jon Yngve Hardeberg

Gjøvik 2009

ISSN: 1890-520X ISBN: 978-82-91313-20-7

Survey of full-reference image quality metrics

Marius Pedersen and Jon Yngve Hardeberg
The Norwegian Color Research Laboratory
E-mail: marius.pedersen@hig.no
`www.colorlab.no` `www.hig.no`

June 24, 2009

Abstract

This paper aims to give a survey of full-reference image quality metrics, including metrics specifically designed to evaluate image quality, but also metrics for image difference, image fidelity, and more. These metrics have in common that they try to predict the perceived difference between an original image and a modified version of it, this modification can typically be compression, halftoning and blurring. They output one numerical value and/or an image difference map. More than 100 image quality metrics have been reviewed and categorized, and short descriptions and analysis of all metrics and their relationships are given. This should prove valuable to researchers in various fields, to find the most appropriate metric for their application, and to give a better understanding of the state of the art of the field of image quality metrics.

Contents

1 Introduction

The need for an objective method to evaluate image difference has become greater as the number of image reproduction methods increase. Many would like to find the best image among several reproductions from a computable metric without asking many observers, both due to the time and resources needed to do the latter. To eliminate psychophysical experiments to find the best reproduction, image quality metrics have been developed. These metrics have the common goal to predict image quality that reflects the perceived image quality, but none have yet been successful on a global scale, because of this more and more metrics appear, either for overall quality or for specific reproduction methods (for example banding in prints, compression artifacts, noise, etc.). A survey is therefore needed. We will limit the survey to full-reference image quality metrics, the goal of this survey is to help researchers find the most appropriate metrics for their field, and to give a better understanding of the state of the art of the field. This should also result in a good tool for further improvements in full-reference image quality metrics. The purpose is to give a survey, rather than giving a critical review of individual papers.

First we explain some terminology and what kind information that can be found in this article. Then a survey of the full-reference metrics are given, where each section is given by the name of the author, and the name of the metric if stated. If independent testing of the metrics has been done this is also mentioned. A discussion chapter is given before concluding remarks. A table with all metrics in chronological order based on year of publication is also given together with a map showing the connection between a number of the image quality metrics.

1.1 Terminology

The terminology used by different authors is quite different. A metric in mathematics can be defined as a function which defines a distance between elements of a set. A strict metric ($\rho(x,y)$) is essentially an abstract distance, with the following properties: $\rho(x,y) = 0$ if $x = y$, symmetry, triangle inequality and non-negativity. The metrics in this survey do not necessarily follow this, and can differ for one or more properties.

Metrics have been developed for image quality, image difference, image fidelity, image similarity, color difference, halftoning, difference predictors and video quality. Even though they are developed for different fields they all have in common that they are full-reference, meaning that they calculate the difference between a complete original and a reproduction (Figure 1). Reduced-reference (where parts of the reference is available) and no-reference metrics (no reference available) have also been proposed, but these are not covered in this paper.

Some metrics are hybrids, they can be used for both image quality, halftoning and color difference, such as S-CIELAB [Zhang and Wandell, 1996], or in both image and video quality. We have tried to adopt the authors terminology in the different publications, and therefore many terms will be used throughout this survey. Performance is a term often used by researchers, this term will also be used throughout this survey to indicate correlation between metric score and mean opinion score, z-score or similar observer score.

Figure 1: Full-reference image quality assessment, where an original is available in the quality calculation. The final quality measure is based on information extracted from both the original image and the distorted image.

1.2 Mapping of important components

A number of components have been registered in order to get an overview of the metrics. The main objective for what kind of reproduction method they were intended for, this can be image quality, halftoning, color difference etc. We have adopted the formulations and terminology of the authors as close as possible. We have also noted whether they use any model of the human visual system (HVS), if they work on different scales (i.e. multiscale), if they are spatial or non-spatial, and if they are color or grayscale metrics. The type of evaluation performed in the original publication has also been registered, with the number of scenes and reproductions. In the case of subjective evaluation the number of observers has been noted. The experience of the observers has also been registered, since several researchers have found differences in preference for experts and non-experts. [Deffner et al., 1994; Dugay, 2007; Dugay et al., 2008; Engeldrum, 2000; Heynderickx and Bech, 2002]. The expertise of the observers will affect the evaluation of the metric, and therefore it is important to have this information when choosing an image quality metric. Some suggestions on publications where the metrics have been evaluated are also made, in order to show which metrics that are widely evaluated by other researchers and which metrics that need more evaluation.

1.3 Review of image difference metrics

Eckert and Bradley [1998] reviewed a number of image quality metrics, and discussed how they incorporate different visual factors in their design. This review was directed toward still image compression.

Ahumada Jr [1993] did a review on image quality metrics for monochrome images, with the intention of helping researchers in search of a suitable metric. The metrics were divided in groups for halftoning, image quality and image compression. Ahumada also proposed a framework for image quality measurements.

Eskicioglu [2000] reviews criteria for monochrome compressed image quality from 1974 to 1999. Eskicioglu states that simple HVS incorporated in quality measures will improve their performance.

Beaton [1983] reports an evaluation of 14 image quality measures for both soft-copy and hard-copy images. The degradations on the images were noise and blur. A database of 10 scenes, and a total of 250 images were used in the evaluation. The result indicated that several measures correlated with subjective scores.

Dosselmann and Yang [2005] reviewed and summarized a set of 10 image quality metrics. The metrics were evaluated on 60 images with 8 different distortion types. The conclusion by the authors was that no metric was superior to any other.

Ouni et al. [2008] gave a review of subjective and objective image and video quality metrics, including full-reference, reduced reference and no-reference metrics. They give a brief introduction to some metrics, and they conclude that there is still much work left.

Most of these reviews are rather short (less than 10 pages), and many of them are more than 10 years old. This contributes to the need of a new and better review of image quality metrics.

2 Full-reference metrics

This section gives a short introduction to the different full-reference image quality metrics.

2.1 Mean square error

The mean square error (MSE) is the cumulative squared error between the compressed and the original image.

$$MSE = \frac{1}{MN} \sum_{y=0}^{M-1} \sum_{x=0}^{N-1} [R(x,y) - S(x,y)]^2$$

where R(x,y) is the original image, S(x,y) is the reproduced version of the original, M and N are the dimensions of the images. The MSE has been used due to its easy calculation and analytical tractability [Teo and Heeger, 1994].

2.1.1 RMSE

The root mean squared error (RMSE) is the square root of MSE.

$$RMSE = \sqrt{MSE}$$

It is well known that MSE and RMSE is inaccurate in predicting perceived distortion [Teo and Heeger, 1994], and therefore many extensions of these measures have been proposed.

2.1.2 SNR

The signal to noise ratio (SNR) of an image is usually defined as the ratio of the mean pixel value to the standard deviation of the pixel values. Due to it's simple calculation SNR has been widely used in objective quality measurement.

$$SNR = 10 \cdot log10 \left(\frac{\sum_{y=0}^{M} \sum_{x=0}^{M} S(x,y)^2}{MN \cdot MSE} \right)$$

2.1.3 PSNR

The Peak Signal to Noise Ratio is a measure of the peak error between the compressed and the original image.

$$PSNR = 20 \cdot log10 \left(\frac{255}{RMSE} \right)$$

The higher the PSNR, the better the quality of the reproduction. SNR and PSNR have usually been used to measure the quality of a compressed or distorted image.

2.1.4 Mitsa and Varkur, CSF weighted MSE

Mitsa and Varkur [1992] investigated the effect of contrast sensitivity function (CSF) on quality measures incorporated in halftoning algorithms. They used a CSF weighted MSE with different types of CSFs. They used both band pass and low pass CSFs from Mannos and Sakrison [1974] and Barten [1990]. 4 different halftoning algorithms were used, and these were evaluated by 12 observers. The results indicate that a low pass CSF performs better than a band pass when compared against psychovisual data. The CSF by Barten also performs slightly better than the one from Mannos and Sakrison. This could be because Barten's model incorporates information about viewing conditions and the display device. The authors conclude that factors that significantly improves the performance of the metrics are:

- Information about the printing device and the viewing conditions.
- Information about the angular sensitivity of the visual system.
- Utilization of a lowpass CSF instead of a bandpass one.

2.1.5 Lin

Lin [1993] proposed a metric for halftone image quality based on the root mean square error (RMSE). This metric applies a CSF on the Fourier transformed original grayscale image and the halftoned image before using the RMSE on the images. The metric also incorporates the viewing distance, this is important because the visibility of the halftone pattern varies according to the distance. The metric is evaluated objectively, with 2 different CSFs from Mannos and Sakrison [1974] and Nill and Bouzas [1992].

2.1.6 Ayed et al., WMSE

Ayed et al. [2002] proposed a quality metric based on MSE called weighted mean squared error (WMSE). Because this is a mathematically defined metric, it does not depend on the images being tested, the viewing conditions or the individual observers. The MSE is calculated with a weighting function that is a relative measure of the spatial frequency activity. From this measure we can easily get WPSNR. The metric was tested on two original images distorted with seven different corruptions; white uniform noise, gaussian noise, impulsive salt-pepper noise and multiplicative speckle noise. The metric was evaluated by the authors, and the conclusion was that WMSE outperforms MSE.

Evaluation of WMSE is found in [Ayed et al., 2002; Samet et al., 2005].

2.1.7 Samet et al., NwMSE

Samet et al. [2005] proposed an image quality metric based on MSE, called New weighted Mean Square Error (NwMSE). It takes into account the pixels neighbourhood information, it also accounts for that the HVS is less sensitive to contrast areas with high spatial frequency activity, and that the variance in low spatial frequency activity regions is small.

2.1.8 Brailean et al., LIPMSE

Brailean et al. [1991] proposed a metric based on the logarithmic image processing (LIP) model by Jourlin and Pinoli [1988]. This model (LIPMSE) accommodates spatial masking, by calculating the contrast at one point depending on the surrounding of that point. Once the original and reproduction are transformed into the contrast domain, MSE is used to calculate the fidelity of the images. The metric is evaluated on one image blurred by a Gaussian MTF. The results indicate that the LIPMSE outperforms the traditional MSE. This metric was used by the authors to find the number of iterations in an image enhancing algorithm for restoring blurred images.

2.1.9 Rushmeier et al., CSF weighted MSE

Rushmeier et al. [1995] presented image quality metrics based on different CSF models. The first one uses the CSF from Mannos and Sakrison [1974] where the similarity between the images is computed in the Fourier space. The second is after Gervais et al. [1984] where the effect of phase as well as magnitude in the frequency domain representation of the image is included. The third is adapted from Daly [1993], where the effects of adaptation and non-linearity are combined in one transformation, the overall distance between the images is computed as MSE. The metrics are evaluated objectively on synthetic images against a real scene.

2.1.10 Mitsa and Alford

Mitsa and Alford [1994] proposed two image quality metrics for digital halftoning, one for single-channel visual models and one for multiple-channel. The single-channel metric is the mean square error between the contrast sensitivity filtered halftone image and contrast sensitivity filtered original image. The multiple-channel metric is based in the frequency-domain halftoning error and is formulated as a vector-magnitude.

2.1.11 Iordache and Beghdadi, SNR_W

Iordache and Beghdadi [2001] proposed an image dissimilarity measure based on a joint spatial/spatial-frequency representation using Wigner-Ville distribution. The measure is built on the fact that structured distortions are more annoying than unstructured distortions. Results show that SNR_W outperforms SNR. A PSNR version of this metric ($PSNR_W$) is found in [Beghdadi and Iordache, 2006].

Evaluation of SNR_W is found in [Beghdadi and Pesquet-Popescu, 2003; Iordache and Beghdadi, 2001].

2.1.12 Beghdadi and Pesquet-Popescu, SNR_{WAV}

Beghdadi and Pesquet-Popescu [2003] proposed an image quality metric based on a non-redundant wavelet decomposition. The metric is inspired by the HVS, and this is also incorporated into the metric. The metric was evaluated by 25 observers on one image reproduced with 3 different distortions. Results indicate that SNR_{WAV} outperform PSNR and SNR_W. A PSNR version of this metric ($PSNR_{WAV}$) is found in [Beghdadi and Iordache, 2006].

2.1.13 Wan et al., *DÉCOR-WSNR*

Wan et al. [2007] introduced the decorrelated weighted signal-to-noise-ratio (*DÉCOR-WSNR*). This metric uses a Point Spread Function (PSF) to mimic the human eye and it also incorporates a constant for the relationship between the halftoned image and original.

2.1.14 Egiazarian et al, PSNR-HVS

Egiazarian et al. [2006] proposed PSNR-HVS based on the HVS and PSNR. The metric uses a scanning window to remove mean shift and contrast stretching similar to UIQ [Wang and Bovik, 2002]. PSNR-HVS is then calculated on the scanned images by using PSNR, where MSE is calculated as described by Nill [1985].

2.1.15 Munkberg et al., mPSNR

Munkberg et al. [2006] proposed an image quality metric for HDR images. The HDR image was divided into LDR images at different exposures, then computing an average of the peak signal-to-noise ratios (PSNR) of each individual exposure.

2.1.16 Chou and Li, PSPNR

Chou and Li [1995] incorporated the properties of the HVS into the estimation of the just-noticeable-difference (JND) profile. The estimation of JND profiles are done in the spatial domain through analyzing local properties of image signals. The JND profiles are further used in the calculation of a fidelity criterion called peak signal-to-perceptible-noise ratio (PSPNR). This measure is based on the PSNR calculation. The criterion was used in order to compress images, and could reach higher perceptual quality at lower bit rates than other methods.

Evaluation found in [Chou and Li, 1995; Mayache et al., 1998]

2.1.17 Lambrecht and Farrell, CMPSNR

Lambrecht and Farrell [1996] proposed a distortion measure (CMPSNR) built on the opponent-colors theory and on multi-channel model of spatial vision. The opponent-color space chosen was the same as proposed by Poirson and Wandell [1993]. The metric transforms the original and reproduction to the opponent color space, then a Gabor filterbank is applied before masking and eventually pooling of the errors in the image. The metric was evaluated by two observers on one scene compressed with 400 different JPEG levels. The results indicate that CMPSNR is better than RMS, and also more consistent.

2.2 Color difference

In this section color difference metrics and formulas are reviewed.

2.2.1 CIELAB ΔE^*_{ab}

The CIE [1986] published the CIELAB ($L^*a^*b^*$) color space specification, with 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. A sample color with CIELAB values L^*_s, a^*_s, b^*_s and a reference color L^*_r, a^*_r, b^*_r. The distance is given by

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

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

The most common way of using ΔE^*_{ab} as an image difference metric is by calculating the color difference in each pixel and finding the mean of these values.

$$\Delta E^*_{ab} = \frac{\sum_{x=0}^{m} \sum_{y=0}^{n} \Delta E^*_{ab_{(x,y)}}}{m \cdot n}, \tag{2}$$

where m is the width of the image and n is the height of the image. Other measures of the ΔE^*_{ab} can be the minimum value or the maximum value in the computed difference.

The CIELAB metric has served as a satisfactory tool for measuring perceptual difference between uniform patches of colors. The human visual system is not as sensitive to color differences in fine details as compared to large patches, yet the CIELAB color metric will predict the same visual difference between the two cases since there is no spatial variable in the CIELAB color metric [Zhang and Wandell, 1998].

Wang and Luo [2008] proposed weighting filters for the CIELAB ΔE^*_{ab} based on how observers assess differences, results from this research indicate that these filters do not improve the performance of the formula.

There are also other versions of this using the CIELUV color space.

$$\Delta E^*_{uv} = \sqrt{(\Delta L^*)^2 + (\Delta u^*)^2 + (\Delta v^*)^2}. \tag{3}$$

Evaluations of ΔE^*_{ab} are found in [Bando et al., 2005; Bonnier et al., 2006; Chou and Liu, 2007; Hardeberg et al., 2008; Kim et al., 2005, 2006; Pedersen, 2007; Pedersen and Hardeberg, 2008, 2009b; Pedersen et al., 2008; Sano et al., 2003; Song and Luo, 2000; Wang and Luo, 2008; Zhang and Wandell, 1998; Zhang et al., 1997b]

2.2.2 CMC

The CMC color difference (ΔE_{CMC}) formula [Clarke et al., 1984] is based on the colorimetric principles of the CIE 1976 system. The CMC formula has acceptance in industrial color control applications. ΔE_{CMC} is a modification of CIE $L^*C^*h^*$ color difference [Sharma, 2002].

Evaluations of CMC can be found in [Sano et al., 2003; Song and Luo, 2000].

2.2.3 Luo and Rigg, BFD

The BFD colour-difference formula was introduced in 1987 [Luo and Rigg, 1987], and it provided a correction for the CMC in the blue region [Imai et al., 2001].

Evaluation of BFD can be found in [Song and Luo, 2000].

2.2.4 CIE ΔE_{94}

The CIE ΔE_{94} [Commission Internationale de l'Eclairage, 1995] was developed as it became clear that the CIELAB ΔE^*_{ab} did not correlate with the perceptual color difference. This formula is based on CIE lightness ΔL^*, chroma ΔC^*, and hue ΔH^* differences.

$$\Delta E^*_{94} = \sqrt{\left(\frac{\Delta L^*}{k_L S_L}\right)^2 + \left(\frac{\Delta C^*}{k_C S_C}\right)^2 + \left(\frac{\Delta H^*_{ab}}{k_H S_H}\right)^2}, \tag{4}$$

where k_L, k_C, k_H are scaling parameters, S_L, S_C, S_H are lightness, chroma and hue scaling functions [Sharma, 2002]. ΔL^*, ΔC^* and ΔH^* are referred to lightness, chroma and hue differences.

Evaluation of ΔE_{94} can be found in [Guarneri et al., 2005; Sano et al., 2003; Song and Luo, 2000; Zhang and Wandell, 1998].

2.2.5 CIE ΔE_{00}

The CIE ΔE_{00} [Commission Internationale de l'Eclairage, 2001; Luo et al., 2001] was published because of the same problems as CIE ΔE_{94} [Sharma, 2002].

$$\Delta E_{00} = \sqrt{\left(\frac{\Delta L'}{k_L S_L}\right)^2 + \left(\frac{\Delta C'_{ab}}{k_C S_C}\right)^2 + \left(\frac{\Delta H'_{ab}}{k_H S_H}\right)^2 + R_T \left(\frac{\Delta C'_{ab}}{k_C S_C}\right)\left(\frac{\Delta H'_{ab}}{k_H S_H}\right)}, \qquad (5)$$

where k_L, k_C, k_H, S_L, S_C, S_H are scaling parameters as in ΔE_{94} and R_T is an additional scaling function depending on chroma and hue [Sharma, 2002]. $\Delta L'$, $\Delta C'$ and $\Delta H'$ are differences in lightness, chroma and hue. Features from both the CMC and BFD have been incorporated in ΔE_{00}.

Evaluations of CIE ΔE_{00} can be found in [Chen et al., 2008; Jin and Field, 2003; Sano et al., 2003; Song and Luo, 2000].

2.2.6 Granger, DP

Granger [2008] proposed a metric for the ATD color space [Granger, 2001; Guth and Lodge, 1973] named Delta Perception (DP). This metric is based on the principle that the final perception of color difference is a linear function of each vision's channel. It also uses the opponent physiology of the HVS. The final value of color difference is calculated using a City block metric on luminance, saturation and hue.

2.2.7 Seim and Valberg, SVF

Seim and Valberg [1986] proposed the SVF formula, based on 3 stages: (1) the absorption of quanta in the three cone pigments (S_i, i=1,3) is a linear, (2) the relative sensitivies of the three cone types are determined by the achromatic adapting stimulus using von Kries coefficient rule and saturating hyperbolic intensity functions (V_i, i=1,3) account of the nonlinear transformations of the resulting cone excitations, (3) linear opponent combinations (F_i, i=1,2) of the hyperbolic functions can be used to approximate chromatic signal processing. The color difference is calculated using the Euclidean distance of the S, V and F.

2.3 Spatial color difference metrics

This section handles spatial metrics, both new and extensions of other non-spatial metrics.

2.3.1 Zhang and Wandell, S-CIELAB

Zhang and Wandell [1996] proposed a spatial extension to the CIELAB color metric (Figure 2). This metric should fulfill 2 goals, a spatial filtering to simulate the blurring of the HVS and a consistency with the basic CIELAB calculation for large uniform areas. The image is separated into an opponent-color space, and each opponent color image is convolved with a kernel determined by the visual spatial sensitivity of that color dimension. Finally the filtered image is transformed into CIE-XYZ, and this representation is transformed using the CIELAB formulae.

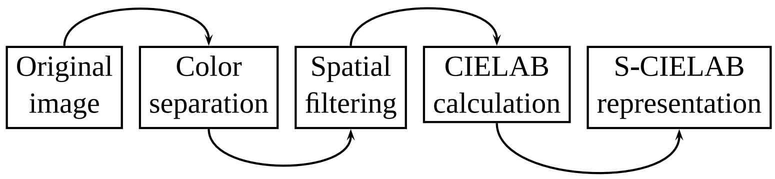


Figure 2: S-CIELAB workflow. Spatial filtering is done in the opponent color space.

Evaluation of S-CIELAB can be found in [Bai et al., 2006; Bando et al., 2005; Bonnier et al., 2006; Bouzit and MacDonald, 2000; Feng et al., 2002; Hardeberg et al., 2008; Hertel, 2005; Jin and Field, 2003; Kim et al., 2005, 2006; Pedersen, 2007; Pedersen and Hardeberg, 2008, 2009a,b; Pedersen et al., 2008; Yu et al., 1998; Zhang and Wandell, 1998; Zhang et al., 1997b]

2.3.2 Johnson and Fairchild, modified S-CIELAB

Johnson and Fairchild [2001] describe CSFs that generally serve to decrease the perceived differences for high frequency image information, such as halftone dots. The CSF removes information about edges in scenes since these mostly contain high frequencies. To avoid this a simple edge-enchancing kernel can be applied, as a convolution with a common Sobel kernel. The results indicate that adding a more precise CSF will improve the performance of S-CIELAB, and it can further be improved by accounting for localization using edge-enhancing filters.

Evaluation of the modified S-CIELAB is found in [Johnson and Fairchild, 2001].

2.3.3 Spatial CIE ΔE_{00}

Chen et al. [2008] proposed a spatial extension of CIE ΔE_{00} with cortex transform decomposition. A multi-channel decomposition is applied using cortex transform [Daly, 1993] in order to simulate HVS selectivites to different spatial frequencies and orientations. Then the sub-images are weighted by a luminance CSF [Daly, 1993] and a chroma CSF [Johnson and Fairchild, 2003]. Finally a pixel level comparison between the orignal and distorted image is carried out using CIE ΔE_{00}.

2.3.4 Nakauchi et al., PD

Nakauchi et al. [1999] proposed a gamut mapping algorithm based on a perceptual image difference measure (PD). The image difference measure used is similar to S-CIELAB, it uses the ΔE^*_{ab} with a spatial filtering. But there are differences, in S-CIELAB the spatial filtering is performed in the opponent color space, while in the proposed measure this is done directly on the CIELAB representation. This is done to calculate the optimal gamut mapping more efficiently. The measure has options to tune weighting parameters using a 3 channel structure, with tunable peak gains for each channel. The input to the model is a difference map between an original and a reproduction, the results are 9 planes (L, a and b for the original, reproduction and difference map), and these planes are spatially filtered.

The measure was used to find the minimum distance in the gamut mapping algorithm. This optimization is iterative, and stops when the difference is lower than a threshold. The results indicate that the proposed gamut mapping algorithm performs better than other tested algorithms (clipping, minimum Δ L, minimum Δ C, minimum Δ H, norm L), and that the proposed image difference metric outperforms ΔE^*_{ab} and ΔL. The authors also state that the number of iterations needed is too high for any practical use.

2.3.5 Kimmel et al.

Kimmel et al. [2005] proposed a metric between 2 color images. This metric is closely related to a variational framework for Retinex [Kimmel et al., 2003], S-CIELAB [Zhang and Wandell, 1996] and PD [Nakauchi et al., 1999]. The metric is a similarity measure in the Sobolev space, where the proximity of the derivatives capture detailed information and the small scale differences between the original and the reproduction. The metric was intended as a variational approach to space-dependent color gamut mapping.

2.3.6 Fairchild and Johnson, iCAM

Fairchild and Johnson [2002] proposed the iCAM model. This model incorporates an image difference metric and it is influenced by S-CIELAB and research in many different fields. The iCAM adds several preprocessing steps, in addition to a spatial filtering. The spatial filtering used serves to modulate spatial frequencies that are not perceptible, and enhance frequencies that are most perceptible. In addition a module of spatial localization is incorporated, to account for the HVS sensitivity to edges. Also a contrast detection is found within iCAM to account for local and global contrast. Finally a color difference formula is used to calculate a color difference map, which can be further analyzed using different statistical methods [Fairchild and Johnson, 2004], such as mean, maximum, and minimum. An analysis of the different statistical methods for iCAM is done by Fernandez et al. [2003].

Evaluation of iCAM can be found in [Bonnier et al., 2006; Fairchild and Johnson, 2002, 2004; Fernandez et al., 2003; Hardeberg et al., 2008; Liu et al., 2005; Orfanidou et al., 2008; Pedersen, 2007; Pedersen and Hardeberg, 2008; Pedersen et al., 2008]

2.3.7 Morovic and Sun, ΔI_{cm}

This image difference metric was proposed by Morovic and Sun [2002]. This metric is based on previous work by the same authors [Sun and Morovic, 2002], where they try to understand what factors contribute to judgments made by observers in experiments where they judge the quality of color reproduction. They proposed a seven step image difference metric. These seven steps includes the use of a 99th percentile in ΔE_{97s} from CIECAM97s2. A CSF is also applied, the same as in S-CIELAB. A weighting of the ΔE_{97s} with the ratio 1:2:1 for ΔJ, ΔC and ΔH. A proportion of unacceptable differences are also taken into account. The distribution of lightness differences is included. The lightness and chroma from the original are also incorporated and at last how the spatial details have changed from the original to the reproduction.

2.3.8 Jin et al., CVDM

Jin et al. [1998] proposed the Color Visual Difference Model (CVDM). The four steps in CVDM is a color space conversion, where the image is transformed in the opponent color space similar to the process in S-CIELAB [Zhang and Wandell, 1996]. The CSF used is adopted from VDP [Daly, 1993] for the luminance channel, while for the chroma channels the CSF is derived from Mullen [1985]. A cortical transform is performed on the three color channels before a visual masking. After this a CSF weighted difference between the two images is calculated. The overall visual difference is added to the filtered reference image, creating a new image. These two images are tranformed back to CIELAB and then the CIELAB color difference formulae is used to compute a visible color difference map.

In addition to the reference image and reproduction image, parameters such as viewing distance, resolution of the images and white point must be given. The model is created to detect the visibility of blur, noise, grating and compression artifacts.

Evaluation of CVDM can be found in [Feng et al., 2002; Jin et al., 1998]

2.3.9 Feng et al., CVDM extension

An extension of CVDM is proposed by Feng et al. [2002]. The extension made was to apply the CSF and the cortex transform in the CIELAB space instead of in the opponent color space, after identifying two problems with CVDM: filtering in the opponent color space sometimes produce out-of-gamut colors, and the CSFs used are low pass filters with a peak sensitivity of 1. The luminance CSF is derived from Barten's CSF [Barten, 1999], resulting in better sensitivity. A pair comparison experiment was carried out to obtain perceived halftone pattern visibility. The results indicate that the modified CVDM predicts texture visibility better than S-CIELAB, and the overall correlation is a bit higher for the modified CVDM than for S-CIELAB.

2.3.10 Taylor et al., IFA

Taylor et al. [1998] proposed an image fidelity assessor (IFA), which accepts two grayscale images as input and generates a probability map as output. The input images are run through a multiresolution decomposition to generate a number of channels, each containing the response of a particular receptive field. These receptive fields are modelled by Gabor functions. A local contrast calculation produces contrast images that describe the response of an ensemble of neurons tuned to a particular spatial frequency and orientation. A psychometric Look Up Table (LUT) is used in the psychometric selector to select the most appropriate psychometric function. The difference between the contrast images for each channel is then applied to the appropriate psychometric function to produce a probability map, these maps are then summarized using a method called limited memory probability summation to create an overall probability map. This model is influenced by the VDP [Daly, 1993] and the VDM [Lubin, 1995].

2.3.11 Wu et al., CIFA

Wu et al. [2001] proposed a color extension of IFA [Taylor et al., 1998] called color image fidelity assessor (CIFA). This extension involves adding two opponent channels (Red-Green and Blue-Yellow), and can be linked to the CVDM [Jin et al., 1998]. The main differences are related to the input and output mismatch. CIFA employs a novel spatial opponent feature, it characterizes the spatial interaction of colors, and CIFA uses normalization of chromatic responses to remove dependency on the luminance level. The metric consists of 3 components, an achromatic IFA and two chromatic IFAs. The difference between these are found in the signal decomposition stage and the psychometric LUTs. In the chromatic IFAs, the luminance contrast is replaced by a chromatic difference. The authors conclude that CIFA provides good predictions over a wide range of distortion types.

2.3.12 Neumann et al

Neumann et al. [1997, 1998] proposed a perception based image metric. This metric takes into account a CSF based on the work by Mannos and Sakrison [1974] and the CIELUV color space. A number of rectangles are placed in the image, in each rectangle the CIE color difference is calculated and summed up to an overall difference between the images.

2.3.13 Yee, pdiff

Yee [2004] proposed a perceptual metric for production testing. The images are transformed from RGB to CIELAB, where a CSF Barten [1990] is applied and the visual masking from Daly Daly [1993] is used. The difference between the original and reproduction is then computed, and these values are compared to a threshold, one for luminance and one for color. If above the threshold value the difference is perceivable, if below it is not perceived.

2.3.14 Pefferkorn and Blin, CCETT visual metric

Pefferkorn and Blin [1998] proposed the CCETT visual metric for color quantization errors on still images. This metric incorporates a merging of a chromatic model and an achromatic model. The achromatic errors are estimated from the original and reproduction images, where a pre-processing of the image is done. Then a retinal processing where a logarithmic response is used, a processing of the mean local luminance for both images is also carried out before an isotropic spatial filtering. After the retinal processing a cortical processing is done, where a splitting of the frequency against orientation is done and a neuronal contrast response is applied. The modelling of chromatic errors is done by a pereptual color representation and a perceptual color-difference. The images are transformed to the CIELAB colorspace, where lightness, chroma and hue can be defined. The color difference is calculated using ΔE^*_{94}. The third stage of this chromatic model is the comparison of color difference, where a local perception of color differences are calculated. A comparison of color differences of each pixel according to its neighbourhood is calculated in both the reference and reproduction. A variation of the neighbourhoods of the images is also taken into account. The metric is calculated by summing the errors of the achromatic and chromatic model. The chromatic error is weighted, because of the lower sensitivity to local chromatic errors. The metric is evaluated with a subjective test, where 6 images were encoded with MPEG-2. One test with chromatic images and one with achromatic images. The CCETT visual metric has a strong correlation with the MOS.

2.3.15 Hong and Luo

This algorithm for color difference is based on the known fact that systematic errors over the entire image are quite noticeable and unacceptable. The algorithm proposed by Hong and Luo [2006, 2002] is based on some conjectures, these are:

- Pixels or areas of high significance can be identified and a suitable weight allocation can be found.
- Larger areas of the same color should be weighted higher.
- Larger color difference between the pixels should get higher weights.
- Hue is an important color perception for discriminating colors within the context.

The first step is to transfer each pixel in the image from L^*, a^*, b^* to L^*, C^*_{ab}, h^*_{ab}. Then a histogram based on the hue angle is computed, and sorted ascending so weights can be applied to 4 different quartiles of the histogram. The overall color difference is calculated by multiplying the weighted hue angle for every pixel with the color difference pixel-by-pixel.

Evaluation of the hue angle metric can be found in [Hardeberg et al., 2008; Hong and Luo, 2006, 2002; Pedersen, 2007; Pedersen and Hardeberg, 2008, 2009a,b; Pedersen et al., 2008]

2.3.16 Pedersen and Hardeberg, SHAME

Pedersen and Hardeberg [2009a,b] proposed an extension of the measure by Hong and Luo [2002]. The spatial hue angle measure (SHAME) used a spatial filtering similar to S-CIELAB [Zhang and Wandell, 1996]. The filtered images (original and reproduction) are used as input to the measure by Hong and Luo [2002]. The results indicate that the spatial filtering combined with the weight allocation in the hue angle measure improve the prediction of perceived image quality.

2.3.17 Pedersen and Hardeberg, SHAME-II

Pedersen and Hardeberg [2009a,b] also proposed another extension of the measure by Hong and Luo [2002]. The spatial filtering is performed in the frequency domain, similar to Johnson and Fairchild [2001], this results in a more precise filtering. The filtered images are then used as input to the hue angle measue by Hong and Luo [2002]. The results indicate that a precise spatial filtering is important for image quality measures.

2.3.18 Chou and Liu, P-CIELAB

Chou and Liu [2007] proposed an image fidelity measure for color images named P-CIELAB (ΔPE). This measure is based on ΔE^*_{ab}, and only the errors exceeding a visibility threshold are taken into account. The visibility threshold is estimated by a proposed visual model that varies from pixel to pixel with local properties of luminance, chroma and background uniformity. The results indicate a better performance by the proposed metric than for ΔE^*_{ab} and PSNR.

2.3.19 Farrugia and Peroche

Farrugia and Peroche [2000] proposed an image metric for computer graphics. The metric treats contrast with a CSF decomposition, a local contrast calculation done for the short, medium and long cone receptors and contrast masking is done in the AC1C2 color space. The color difference is computed using the CIELAB ΔE^*_{ab}, while the contrast maps are computed using the Minkowski metric. The color difference map and contrast map is calibrated in JND, and the sum of these is used to obtain the final difference map. Another variant of this metric is found in Farrugia et al. [2004].

2.4 Structural similarity

This section reviews metrics based on structural similarity, both for grayscale and color images.

2.4.1 Wang and Bovik, UIQ

The Universal Image Quality Index (UIQ) was proposed by Wang and Bovik [2002]. This is a mathematically defined image quality 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 index models any distortion as a combination of loss of correlation, luminance distortion and contrast distortion. 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 UIQ outperform MSE significantly for different types of distortion. The test images had 7 types of distortion; salt-and-pepper noise, mean shift, JPEG compression, additive gaussian noise, multiplicative speckle noise, contrast stretching and blurring. The authors also stress the point that the ranking done by UIQ is the same as by the observers, and the correlation between Q values and mean subjective rank (MSR) is high.

Evaluation of UIQ can be found in [Egiazarian et al., 2006; Gayle et al., 2005; Hardeberg et al., 2008; Pedersen and Hardeberg, 2009b; Samet et al., 2005; Shnayderman et al., 2004, 2006; Wang and Bovik, 2002; Wang et al., 2004]

2.4.2 Toet and Lucassen, Q_{color}

Toet and Lucassen [2003] introduced a color image fidelity metric based on the UIQ by Wang and Bovik [2002]. The UIQ is performed on each channel in the l, α and β channels, these channels are calcualted by a transformation from the LMS space. The final color metric is defined as

$$Q_{color} = \sqrt{w_l(Q_1)^2 + w_\alpha(Q_\alpha)^2 + w_\beta(Q_\beta)^2}, \tag{6}$$

where Q is the UIQ calculation. The w indicate weights that can be set according to the distortion in each channel. The results indicate a correlation between the ranking of image fidelity by the observers and the ranking based on the Q_{color} values.

2.4.3 Egiazarian et al., UQI-HVS

Egiazarian et al. [2006] proposed an extension of UIQ by Wang and Bovik [2002]. This metric takes into account the human visual system by using an one-level discrete wavelet transform similar to the on in JPEG2000. As a result of this the image is divided into 4 frequency subbands, and the UIQ values are calculated for each subband. The final value (UQI-HVS) is obtained by summing the 4 weighted subband values.

2.4.4 Wang et al., SSIM

The SSIM (structural similarity) index proposed by Wang et al. [2004] attempts to quantify the visible difference between a distorted image and a reference image. This index is based on the UIQ [Wang and Bovik, 2002]. 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, the overall image quality is the mean of all these local windows.

$$MSSIM(X,Y) = \frac{1}{M}\sum_{j=1}^{M} SSIM(x_j, y_j), \tag{7}$$

where X and Y is the reference and distorted images, x_j and y_j are image content in local window j and M indicates the total number of local windows. Figure 3 shows the SSIM flowchart, where signal x or signal y has perfect quality and the other is the distorted image.

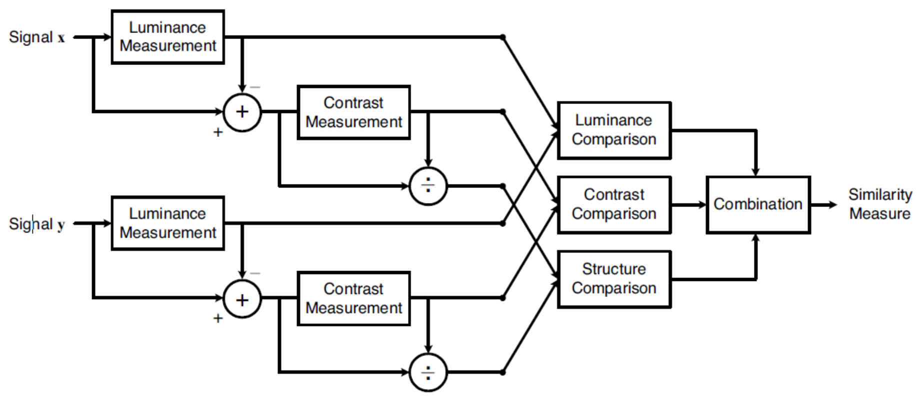


Figure 3: SSIM flowchart, signal x and y goes through a luminance and contrast measurement before comparison of luminance, contrast and structure. A combination of these results in the final similarity measure. Picture from Wang et al. [2004].

Evaluation of SSIM can be found in [Beghdadi and Iordache, 2006; Bouzerdoum et al., 2004; Brooks and Pappas, 2006; Chen et al., 2006b; Dosselmann and Yang, 2005; Egiazarian et al., 2006; Gao et al., 2005; Gayle et al., 2005; Hardeberg et al., 2008; Larson and Chandler, 2008; Lee and Horiuchi, 2008; Lee et al., 2006; Pedersen, 2007; Pedersen and Hardeberg, 2008, 2009b; Pedersen et al., 2008; Samet et al., 2005; Sheikh and Bovik, 2006; Sheikh et al., 2004; Shnayderman et al., 2006; Silva et al., 2007; Wang and Simoncelli, 2005; Wang et al., 2003b, 2004; Yao et al., 2005]

2.4.5 Wang et al., multiscale SSIM

A multiscale version of SSIM was proposed by Wang et al. [2003b]. The original and reproduction is run through the SSIM, where the contrast and structure is computed for each subsampled level. The images are low-passed filtered and downsampled by 2. The lightness (l) is only computed in the final step, contrast (c) and structure (s) for each step. The overall values are obtained by multiplying the lightness value with the sum of contrast and structure for all subsampled levels. Weighting parameters for l, c and s are suggested based on experimental results.

Evaluation of multiscale SSIM can be found in [Sheikh et al., 2006; Wang et al., 2003b]

2.4.6 Chen et al., ESSIM

Chen et al. [2006b] proposed an edge-based SSIM (ESSIM). This metric was developed due to the findings that SSIM fails to predict perceived difference in badly blurred images. This measure replaces the structural comparision with an edge-based structural comparision based on the Sobel operator and an edge direction histogram. The metric is evaluated on the LIVE database, and the results indicate that ESSIM perform better than SSIM and PSNR.

2.4.7 Chen et al., GSSIM

Chen et al. [2006a] proposed a gradient-based SSIM (GSSIM). This metric is very similar to the ESSIM by Chen et al. [2006b], it uses a Sobel filter in horizontal and vertical direction just as ESSIM. The difference is that both the structure and contrast comparision are based on the Sobel filtered images. The metric is evaluated on the LIVE database, and the results indicate that GSSIM performs better than SSIM and PSNR, both on an overall basis and for blurred images.

2.4.8 Bonnier et al., $SSIM_{ipt}$

A color extension of SSIM was developed and tested by Bonnier et al. [2006], where each SSIM for each channel in the IPT color space were performed. After the transformation all three channels were combined with a geometrical mean. This implementation is similar to the one used by Toet and Lucassen [2003] on UIQ.

2.4.9 Gao et al., CBM

Gao et al. [2005] proposed an extension to SSIM, based on a fuzzy Sugeno integral. The content-based metric (CBM) has l(x,y) and c(x,y) similar to SSIM [Wang et al., 2004] but the s(x,y) is modified from containing values between -1 and 1 to contain values between 0 and 1 due to the Sugeno integral. By analyzing the content of the original and reproduction the image is partitioned into 3 parts;

edge, texture and flat regions. Then the similarity measure of each part is calculated by synthesizing the SSIMs of all the pixels in the corresponding regions with the Sugeno integral. Finally the overall image quality is evaluated with the weighting average of the three regions. The results indicate a higher performance by the CBM than SSIM, PSNR and fuzzy integral (FE) on a set of JPEG and JPEG2k compressed images.

2.4.10 Dong et al., RCBM

Dong et al. [2007] proposed another extension to SSIM using rough fuzzy integrals (RCBM). The CBM by Gao et al. [2005] is always based on pixelwise integral, making the analysis less flexible, by using the rough fuzzy integral this should become more flexible. The image is partitioned into 3 parts as in CBM; edges, texture and flat regions. The rough fuzzy integral is then applied instead of the fuzzy integral. Finally the lower and upper measurements of the three parts are calculated using the weighting average. This method results in an upper and lower limit, and therefore being more flexible than the CBM.

2.4.11 Wang and Simoncelli, CWSSIM

Wang and Simoncelli [2005] address the problem SSIM has with translation, scaling and rotation. The solution for this is to extend SSIM to the complex wavelet domain. In order to apply this metric (CWSSIM) for comparing images, the images are decomposed using a complex version of the steerable pyramid transform. The CWSSIM is computed with a sliding window, and the overall similarity is estimated as the average of all local CWSSIM values. From the objective test done, CWSSIM outperform SSIM and MSE. The authors also tested the metric as a similiarity measure on 2430 images.

Evaluation of CWSSIM can be found in [Brooks and Pappas, 2006; Wang and Simoncelli, 2005].

2.4.12 Brooks and Pappas, WCWSSIM

Brooks and Pappas [2006] presented a multi-scale weighted variant of the complex wavelet SSIM (WCWSSIM). This extension of the CWSSIM have weights based on the CSF to handle local mean shift distortions. They use SSIM, CWSSIM and WCWSSIM to evaluate video quality. They also show examples with local mean shifts where CWSSIM have problems, and they propose the WCWSSIM to account for these problems. The authors conclude that WCWSSIM is superior to SSIM and CWSSIM.

2.4.13 Lee et al., DTWT-SSIM

Lee et al. [2006] proposed a dual-tree wavelet transform extension of SSIM (DTWT-SSIM), this extension should be less sensitive to translation, scaling and rotation than SSIM. The mean intensity of luminance signals (μ_x and μ_y) are replaced by a sum of the six sub-bands of the DTWT coefficents. This is a similarity measure used for classification, and has been tested on the MNIST handwritten digit database [Lecun et al., 1998]. 4860 images were created with scaling, shift, blurring and rotating. The results show a higher correct identification rate for DTWT-SSIM than for SSIM and MSE. The authors also indicate that DTWT-SSIM can be used in other fields, as face recognition or content-based image retrieval.

2.4.14 Mindru and Jung

Mindru and Jung [2006] proposed a similarity metric for image quality based on SSIM [Wang et al., 2004] and I^*_{color} [McCormick-Goodhart et al., 2004]. The original and reproduction are transformed into XYZ, after this the images are filtered with a spatial human visual observer model before going back to XYZ. The final measure is a weighted combination of SSIM and I^*_{color}.

2.4.15 Lee et al., CISM

Lee et al. [2007] and Lee and Horiuchi [2008] proposed a hybrid error diffusion algorithm which uses an internal pseudorandom number and a 4x4 mask. To verify the quality of the algorithm a structural similarity measure for color images is proposed. This measure builds on SSIM, but differs on several points. The RGB image is transformed to CIEXYZ and further to CIELAB, then a DFT is performed on the transformed image, both original and reproduction. A HVS filter is applied on both images. For the luminance channel the model proposed by Sullivan et al. [1991] and Näsänen [1984] is used, and for the chrominance a filter based on results by Mullen [1985] is used. Then the HVS filtered images are transformed to RGB again after an inverse DCT. The input in the similarity part is the HVS filtered RGB values. A comparison of the mean RGB, variance RGB and structure is performed, and a combination of these is done in order to obtain the similarity measure. This is done for all three channels, and the authors propose to weight each channel with 1/3. This results in a final measure (CISM) where the mean SSIM for each channel is summed. The results indicate that the proposed algorithm performs better than other error diffusion algorithms, and that CISM outperforms SSIM.

2.4.16 Lam and Loo

Lam and Loo [2008] proposed an image quality metric based on SSIM, but the structure of the image is extracted using a quadtree decomposition. These quadtree segments during decomposition are used to examine contrast and luminance.

2.4.17 Silva et al., SEME

Silva et al. [2007] proposed an image similarity measure based on a modified version of the measurement of enchancement by entropy. This measure is a similarity measure, and can therefore also be used as a quality measure. The similarity enchancement by entropy (SEME) is defined with 2 variants, one with the a number of horizontal and vertical blocks in the image, and with maximum and minimum luminance in each block. The other is similar, but it incorporates a MSE weighting function. The metric has been compared to PSNR, SSIM and MSE. 4 different SEME are tested, with block size 3 and 4 for both variants of the measure. The results indicate that SSIM is better than SEME, but with only minor differences. The SEME is though twice as fast as SSIM , but slower than MSE and PSNR. The dataset contained 233 images from the LIVE database, distorted with JPEG compression.

2.4.18 Franti

Franti [1998] proposed a distortion mesure for statistical and structural errors in digital images. The method consist of 3 quality factors, detecting contrast errors, structural errors and quantization error. Contrast masking is performed on the contrast errors and on the structural errors, but not on the quantization errors. The final measure is a sum of the weighted factors. 3 different scenes with 14 distorted versions were used to evaluate the proposed measure, 15-39 observers were used. The results indicate that the proposed measure perform better than PQS and MSE.

2.5 Difference predictors

Difference predictors are reviews in this section, these metrics are used to predict the difference between two images.

2.5.1 Daly, Visible Differences Predictor

This is an algorithm proposed by Daly [1993]. The goal of the Visible Differences Predictor (VDP) is to determine the degree to which physical differences become visible differences. The author states that this is not an image quality metric, but it addresses the problem of describing the differences between two images. The output from this algorithm is an image containing the visible differences between the images. Two different visualization techniques are proposed for the output VDP, the free-field difference map optimized for compression, and the in-context difference showing the output probabilites in color on the reference image. The VDP can be used for all image distortions including blur, noise, algorithm artifacts, banding, blocking, pixellation and tone-scale changes.

Evaluation of VDP is found in [Daly, 1993; Li et al., 1998]

2.5.2 Fewerda and Pellacini, FDP

Another predictor, Functional Difference Predictors (FDPs) [Ferwerda and Pellacini, 2003], has been built on the same principles as VDP [Daly, 1993]. VDPs predict whether images will be visibly different, but FDPs predict whether they are functionally different, affecting the user's ability to perform a task [Ferwerda and Pellacini, 2003].

2.5.3 Bradley, WVDP

Bradley [1999] proposed a wavelet extension of the VDP [Daly, 1993]. The original and reproduction are transformed in to the wavelet domain. The differences between these are tested against a contrast masking. A psychometric function is used to estimate the probability of error detection for each wavelet coefficient. These detection probabilities are combined to get a visible difference map. A CSF is not explicity found in WVDP, but is incorporated in the contrast masking.

2.5.4 Mantiuk, HDR-VDP

Mantiuk et al. [2004] proposed an extension of VDP for HDR images. The extensions done improve the model's prediction of perceivable differences in the full visible range of luminance and under the adaptation condition. HDR-VDP takes into account aspects of high contrast vision in order to predict perceived differences. This model does not take into account chromatic changes, only luminance.

Evaluation of HDR-VDP is found in [Mantiuk et al., 2004, 2005].

2.6 Discrimination models

2.6.1 Lubin, VDM

Lubin [1995] proposed the visual discrimination model (VDM). This model is based on the just-noticiable-differences (JND) model by Carlson and Cohen [1980]. The model design was motivated by speed and accuracy. Input to the model is a reference image and a distorted version, both grayscale. A set of parameters must be defined based on the viewing conditions. The first step includes a simulation of the optics of the eye before sampling the retina cone mosaic. The sampling is done by a gaussian convolution and point sampling. The next stage converts the raw luminance signal into units of local contrast based on a method similar to Peli [1990]. After this each pyramid level is convolved with 4 pairs of spatially oriented filters. Then on the 4 pairs of filters an energy response is computed. Each energy measure is normalized and each of these values are as input to a non-linear sigmoid function. The distance between the vectors can be calculated and results in a JND map as output, but the values across this map can be used to calculate mean, maximum or other statistical measure of the similarity between the images. This single value can further be converted into probability values.

2.6.2 Lubin, Sarnoff JND Vision Model

The Sarnoff JND Vision Model [Lubin, 1997a,b] is a method of predicting the perceptual ratings that observers will assign to a degraded color-image sequence relative to its nondegraded counterpart. The model takes two images, an original and a degraded image, and produces an estimate of the perceptual difference between them. The model does a front end processing to transform the input signals to light outputs (YC_bC_r to Yuv), and then the light output is transformed to psychophysically defined quantities that seperately characterize luma and chroma. A luma JND map and a chroma JND map are created. The JND maps are then used for a correlation summary, resulting in a measure of difference between the original and the degraded image. It should be noted that the metric was developed as a video quality metric, showing a high correlation between predicted quality and perveiced quality. The model has also been tested on JPEG data, where a high correlation also was found. Lubin concludes that the model has wide applicability as an objective picture quality measurement tool.

Evaluation of Sarnoff can be found in [Brill et al., 1999; Li et al., 1998; Lubin, 1997b; Sheikh and Bovik, 2006; Sheikh et al., 2004, 2006; Wang et al., 2003b, 2004]

2.6.3 Bolin and Meyer, Simplified color extension of VDM

Bolin and Meyer [1998, 1999] proposed a simpliflied VDM [Lubin, 1995] for realistic image synthesis by using the Haar wavelet and a simpler spatial pooling operation with a smaller filter than the one found in VDM. This version was extended to handle color and it includes the effects of chromatic aberration.

Evaluation of this metric is found in [Bolin and Meyer, 1998, 1999].

2.7 DCT based metrics

2.7.1 Watson

Watson [1993a,b] proposed a metric based on DCT, the method treats each DCT coefficient as an approximation to the local response of a visual channel. The metric is built to optimize individual images made from JPEG, MPEG etc. The model incorporates contrast masking, luminance masking, perceptual quantization, JND, spatial error pooling, frequency error pooling, an optimization method and bit rate optimization.

Evaluation of this metric is found in [Mayache et al., 1998; Watson, 1993a,b].

2.7.2 Silverstein and Klein

Silverstein and Klein [1993] proposed a DCT image fidelity metric for displayed text. The closest fit for of ASCII symbols to rectangular segments of a gray-scale image. Each segment was converted into a DCT coefficient matrix which was compared to the coefficient matrix of each ASCII symbol. The image segment was replaced with the symbol that had the least weighted Euclidean distance.

2.8 Wavelet based metrics

2.8.1 Lai et al.

Lai et al. [1997] proposed a fidelity measure based on the Haar wavelet. This measure was more refined by Lai et al. [1998] where they still use the Haar wavelet to simulate the HVS in luminance and chrominance. Based on the wavelet representation the contrast at each pixel is computed, the contrast is adjusted by the masking effect and the threshold curve is truncated. The suprathreshold at each resolution is computed, and the error measure is based upon this. The errors are computed in luminance, red-green and yellow-blue dimensions with respect to their individual contrast threshold curves. The final measure is a geometrical mean of these.

2.8.2 Veryovka et al.

Veryovka et al. [1998] used a multiscale approach to analyze edges, this measure was created for halftone images. Two types of edges were identified: reproduced edges and egde artifacts. Reproduced edges are those edges that are present in the original images, while edge artifacts are those introduced by the halftoning algorithm. The authors use the wavelet tranform proposed by Mallat and Hwang

[1992] abd Mallat and Zhong [1992], and applied this to the halftoned image. Extrema points are identified and linked together to form edge contours. Edges that correspond to high frequency noise are removed together with weak edges. Short edges are also removed, short edges are defined as edges smaller then 7 pixels. The strength of an edge artifact is calculated as the average of the magnitudes of edge points at a given scale. The metric is tested on clustered dot dither, ordered dispersed dither and Floyd-Steinberg error diffusion.

2.8.3 Sheikh et al., IFC

Sheikh et al. [2004] and Sheikh [2004] proposed an image quality assessment method based on natural scene statistics (NSS). The NSS model used is Gaussian scale mixtures (GSM) in the wavelet domain. The distortion model used is also described in the wavelet domain. This model captures two important, and complementary, distortion types: blur and additive noise. The Information Fidelity Criterion (IFC) is the mutual information between the source and the distorted images. The IFC is not a distortion metric, but a fidelity criterion. It theoretically ranges from zero (no fidelity) to infinity (perfect fidelity). The IFC was evaluated on the LIVE database, and the results show that the IFC outperform the Sarnoff JNDmetrix, and a vector version of IFC even outperform SSIM.

2.8.4 Yao et al., *QMCS*

Yao et al. [2007] proposed an image quality measure based on curvature similarity, and builds on the ideas of Wang et al. [2004]. The metric attempts to exploit structural similarity in wavelet bands using differential geometric information. The reference image and distorted image are decomposed into a 4 level structure with a total of 13 subbands using 9/7 biorthogonal wavelet filter banks. In each subband mean surface curvature maps are obtained, and the similarity between two curvature maps are found with a correlation formula. The overall quality measure is computed by summing the values for all subbands. The results indicate a better performance by *QMCS* than SSIM, Sarnoff and PSNR on the LIVE database.

2.8.5 Gayle et al., M-DWT

Gayle et al. [2005] proposed a full-reference image quality measure in the Discrete Wavelet Transform (DWT) domain. The measure applies DWT to the luminance layer of the orignal and degraded image, and for each band the standard deviation of differences are calculated. The measure was evaluated by 14 observers on 30 full color scenes. The proposed measure (M-DWT) is better correlated with MOS (mean opinion score) than PSNR, UIQ and SSIM.

2.9 Various metrics

2.9.1 Damera-Venkata et al., *DM* and *NQM*

Damera-Venkata et al. [2000] proposed a distortion measure (*DM*) and noise quality measure (*NQM*). For the *DM* the frequency distortion in the reproduction is found, then the deviation of this from an allpass response. Finally a weighting of the deviation by a HVS model, and integration of the visible frequencies. The *NQM* is based on Peli's contrast pyramid [Peli, 1990], and it takes into account variations in distance, variations in local luminance mean, interaction between spatial frequencies and contrast masking effects.

Evaluation of these metrics are found in [Damera-Venkata et al., 2000; Sheikh et al., 2006]. NQM is also found in [de Freitas Zampolo and Seara, 2004; de Freitas Zampolo and Seara, 2005], while DM is also found in [de Freitas Zampolo and Seara, 2003].

2.9.2 de Freitas Zampolo and Seara, *DQM* and *CQM*

de Freitas Zampolo and Seara [2003] proposed measures based on the *NQM* by Damera-Venkata et al. [2000]. The distortion quality measure (*DQM*) is based on *NQM*, and is done in the frequency distortion assessment block. The results show superior performance of the *DQM* over the *NQM*. The authors also propose the *CQM*, a composit quality measure, that combines frequency distortion with noise injection. This measure was derived from *NQM* and *DQM*, and a high correlation against the subjective score is found.

Evaluation of *DQM* and *CQM* is found in [de Freitas Zampolo and Seara, 2003; de Freitas Zampolo and Seara, 2004].

2.9.3 de Freitas Zampolo and Seara, *B-CQM*

de Freitas Zampolo and Seara [2004] proposed an extension of the *CQM* proposed by the same authors [de Freitas Zampolo and Seara, 2003]. This extension uses the Bayesian networks, and is based on the experimental data from [de Freitas Zampolo and Seara, 2003]. Results indicate increased performance for *B-CQM* over *NQM* and *DQM*.

2.9.4 Yu et al.

Yu et al. [1998] proposed a metric for color halftone visibility based on Weber's law [Weber, 1965]. The first step is to transform the RGB values of the patch to XYZ, the Fourier transform is applied to find the luminance in the frequency domain. A cut-off radial frequency is defined, then a new image is defined in the frequency domain according to the cut-off radial frequency. The inverse Fourier transform is applied to the previous image, and the standard deviation of this is found. This value is compared to the limit given by the mean luminance and a factor. If the value is higher than the limit the cut-off radial frequency is increased, and a new check against the limit is done. When the the value is lower than the limit, the cut-off radial freqeuency is reported as the frequency threshold. The metric is evaluated with a psychovisual experiment with 8 observers and 6 scenes and 5 different halftone algorithms. The results indicate that the proposed metric performs better than S-CIELAB on the same dataset.

2.9.5 Yu and Parker, KLK

Yu and Parker [1998] evaluated and proposed quality metrics for individual blue-noise binary patterns; the HWMSE (HVS weighted mean square error), LF (lowest frequency component), AMD (average distance beween nearest neighboring minority-pixel pairs) and KLK (kurtosis local kurtosis). A metric similar to the AMD has earlier been proposed by Wong [1997]. The KLK is proposed by the authors, where for each minority pixel a lowpass gaussian is applied to its neighbourhood. The kurtosis of the filter output in this neighbourhood is calculated and then the kurtosis of all kurtosis distributions to get a single numerical value for the quality. Based on these quality metrics an optimization of blue-noise binary patterns is performed. They conclude that by using quality metrics to choose the best blue-noise binary pattern they reflect the visually best blue-noise binary pattern.

2.9.6 Ivkovic and Sankar

Ivkovic and Sankar [2004] proposed an algorithm for image quality assessment built on the linear relationship between blocks of pixels. The first step in this model is to process both the original image and the reproduction with a simple HVS model. This HVS model consists of a brightness perception function and a CSF. The CSF contains an user defined parameter for changing with the reference image content. The correlation coefficient is then computed on the different blocks on the image. At last the overall quality measure is computed as the average correlation coefficient. The metric is evaluated on three images reproduced with contrast stretching, additive white noise, blurring and JPEG2000 coding. The results were evaluated by the authors, and the proposed algorithm was found to perform better than MSE.

2.9.7 Chaddha and Meng

Chaddha and Meng [1993] introduced an image distortion measure for monochrome images. This measure splits the error signal in 2 parts. One for the orthogonal-space, where parts of the error signal is correlated with local features, and one for the sub-space error signal, where parts of the error signal is uncorrelated with local features. This splitting is done by a 2-D adaptive filter, where a sliding filter of 9$\times$9 is used. The second stage of the measure is a full wave rectifier, used for outputting absolute values. Then a logarithmic non-linearity is performed, and as a fourth stage where the image is raised to a power of 2. The fifth stage consists of a weighted summer. Then a relative weighting of the two distortion measures is done to obtain the final distortion measure. The result from the metric is compared against perceived distortion, and the proposed measure outperform MSE both in correlation and correct ranking. This metric is also extended to video [Chaddha and Meng, 1993].

2.9.8 Näsänen

Näsänen [1984] examined the visibility of halftone dot textures, and a method based on the human visual contrast sensitivity was proposed. The estimated visibility is expressed as the visual resolution frequency, that is the highest fundamental spatial frequency where a periodic dot texture is visible for the observers. The results from this method correspond with those from a psychophysical experiment.

2.9.9 Sheikh and Bovik, VIF

Sheikh and Bovik [2006] proposed the visual information fidelity (VIF) metric. This metric quantifies the Shannon information present in the reproduction relative to the information present in the original. The natural scene model used is a Gaussian scale model (GSM) in the wavelet domain, and as a HVS model they use an additive white Gaussian noise model. The VIF metric showed a higher correlation with perceived quality from the LIVE database than Sarnoff, PSNR and SSIM, and with a lower RMSE.

Evaluation of VIF is found in [Larson and Chandler, 2008; Sheikh and Bovik, 2006; Sheikh et al., 2006].

2.9.10 McCormick-Goodhart et al., I^*

McCormick-Goodhart et al. [2004] proposed a computational model for ”retained image appearance”. The I^* metric is a combination of an image appearance function for the lightness and contrast ($I^*_{B\&W}$) and retained image appearance function for the color information (I^*_{color}). The totalt error is weighted with a factor ω.

$$I^* = \frac{I^*_{color} + (I^*_{B\&W} \times \omega)}{1+\omega}. \tag{8}$$

The results indicate that I^* is consistent with visual results of print aging, in terms of contrast and color loss.

2.9.11 Yeh et al., E

Yeh et al. [1998] presented a perceptual distortion measure (E) for edge-like artifacts in image sequences. The measure does a spatial processing, then a temporal processing before a distortion pooling. The spatial processing is divided into 4 parts. The first part being a distortion detection, where a lowpass filter is added along a direction θ, and a bandpass filter along $\theta + 90$. Then an absolute value for the current position is calculated. The spatial masking is done on the background luminance and on the background spatial activity. The spatial nonlinearity and summation are then carried out. The results indicate that E is more correlated with the HVS than MSE.

2.9.12 Shnayderman et al., $M-SVD$

Shnayderman et al. [2004, 2006] proposed an image quality metric using singular value decomposition ($M-SVD$), according to the authors it is reliable across different distortion methods. The measure works in blocks, and calculates the difference between the singular values of these blocks. A total value can be found by a simple summation of the blocks. It was tested using 6 different distortion types (JPEG, JPEG2000, gaussian blur, gaussian noise, sharpening and DC-shifting) on 5 different scenes with 10 observers. The results show better performance by $M-SVD$ than UIQ and MSE. Because this metric does not require a HVS model, it is simple and does not have any assumptions about viewing distance or the distortion type. It can be used for both local and global measurements.

2.9.13 Pappas and Neuhoff, LSMB

Pappas and Neuhoff [1999] proposed the least-squares model-based (LSMB) approach to digital halftoning that exploits both a printer model and a visual model to create high quality images. The squared error between the output of the printer model and the visual model is used for optimization of halftoning. Results indicate that the LSMB error metric agrees well with visual evaluations of image quality. LSMB image quality metric can be computed over the whole image or over a small segment of the image. The evaluation of performance of the LSMB techinque is done in terms of 1) spatial resolution (sharpness) 2) texture (visibility of halftone pattern) 3) gray-scale resolution (number of perceived gray levels) 4) gray scale distortion of halftoned images.

2.9.14 Scheermesser and Bryngdahl

Scheermesser and Bryngdahl [1996] introduced a texture metric for halftone images. It is a numerical metric for the occurrence or absence of specific textures in quantized images. The metric conforms to the visual impression of the image, it permits a judgement over visually unrecognizable texture features, and it results in a number for easy interpretation. The authors also state that this metric could be used as an optimization algorithm for halftoning.

2.9.15 Scheermesser and Bryngdahl

Scheermesser and Bryngdahl [1997] proposed a space-variant texture metric for halftone images that allows the identification and quantification of spatially dependent texture characteristics. 2 approaches are given for investigation of the image, one on segmentation and on a continuous distribution.

2.9.16 Barten, SQRI

Barten [1990] proposed the square-root integral (SQRI), in this metric a fixed mathematical expression for the contrast sensitivity of the eye is used. This results in the ability to detect various phenomenon as resolution, addressability, contrast, luminance, display size and viewing distance. The results indicate a linear correlation with perceived image quality.

Evaluation of SQRI is found in [Barten, 1990; Bouzit and MacDonald, 2000].

2.9.17 Nijenhuis and Blommaert

Nijenhuis and Blommaert [1997] presented a framework for an alternative metric that uses the distance in a perceptual space to predict the perceived impairment of reproduced images. This metric include attributes induced by Michelson's contrast and average luminance.

2.9.18 Avadhanam and Algazi, PDM

Avadhanam and Algazi [1999] proposed an image fidelity metric, called the picture distortion metric (PDM), based on visual masking. The main components of the metric are perceptual nonlinearity, CSF, cortex bands, visual masking and error summation. The CSF used is adopted from the VDP [Daly, 1993], together with the cortex bands. They introduce a new approach to masking, where the response is normalized for each band. The metric is evaluated with 5 different scenes, coded with JPEG and a wavelet coder. The results indicate a high correlation and high correct ranking for PDM, a high scene dependency is also found.

2.9.19 Karunasekera and Kingsbury, E_θ

Karunasekera and Kingsbury [1994, 1995] proposed a measure (E_θ), which is divided into 3 parts. Edge detection, where the original image and the reproduction are filtered with a directional filter. Masking of artifacts due to surrounding activity and brightness. Accounting for the nonlinearity in the human visual system is done before error calculation. The error calculation is the average of transformed error, either of a part of the image or the entire image. The ranking by the metric match the subjective ranking well.

Evaluation of E_θ is found in [Karunasekera and Kingsbury, 1994, 1995; Mayache et al., 1998]

2.9.20 Miyahara et al., PQS

Miyahara et al. [1996] introduced the Picture Quality Scale (PQS). The measure is based on luminance coding error, spatial frequency weighting of errors, random errors and disturbances, structured and localized errors and disturbances, and principal component analysis. The PQS is calculated as a linear combination of the different principal components. The metric was evaluated by a psychophysical experiment with 5 scenes and 9 expert observers. A strong correlation is found for PQS, and the performance of the proposed measure is better than WMSE.

Evaluation of PQS is found in [Miyahara et al., 1996; Sheikh et al., 2006].

2.9.21 Safranek and Johnston

Safranek and Johnston [1989] introduced a simple metric based on the PIC (perceptual image coder), where an error pooling is done on the subband indexes. The coder incorporates a model that ensures that the most sensitive areas are not overcoded. A texture masking adjusment model is incorporated in order to code areas with different frequency content at different levels.

2.9.22 Westen et al., PEM

Westen et al. [1995] proposed the Perceptual Error Measure (PEM), which incorporates the HVS light sensitivity, spatial frequency and orientation sensitivity, and masking effects. The model is based on local band-limited contrast in oriented spatial frequency bands. The PEM calculation is based on a simple vector norm over frequency bands and positions, and weights can be given to the different frequency bands, positions and orientations. The results indicate a superior performance by PEM over PSNR.

2.9.23 Nilsson, QM_a

The halftoning quality metric proposed by Nilsson [1999] incorporates models of both the printer and the observers. This model is based on earlier work by Nilsson and Kruse [1997]. The print model consist of 2 parts, one with the mechanical distortion of the halftone and one with the optical effects of the paper. The observer model has been adopted from Mannos and Sakrison [1974] for the HVS, and for the modulation-transfer function (MTF) values and data from Sullivan et al. [1993] is used. An adaptive filter is used to separate the halftone characteristics from the original image, and Fourier analysis of the separations using weight functions to derive measurements on different aspects of quality. The square root of the averaged energy in each frequency band (H_r) is used as a descriptor of the signal. The final measure (QM_a) is computed as a sum of the H_r and a weighting function.

2.9.24 Kipman, ImageXpert metrics

Kipman [1998] presented a set of image quality tests for printers and media, these tests include dot quality, halftone quality, line quality, text quality, color quality, smear/overspray and spatial resolution. These metrics are a part of the ImageXpert software.

Evaluation of these metrics are found in [Brill et al., 1999; Kipman, 1998]

2.9.25 Wilson et al., Δg

Wilson et al. [1997] proposed a metric for gray-scale comparison. The metric is built on a modification of the Hausdorff metric, which measures how far 2 compact non-empty subsets of a metric space are from each other. The distance between the original and reproduction is the distance between their respective subgraphs. The metric was tested on different kind of distortion, among them JPEG compression. The results indicate that the Δg for some kind of distortions are better than the RMS.

2.9.26 Guarneri et al., $PQSI$

Guarneri et al. [2005] proposed a quality metric (PQSI) for color interpolated images. The metric is divided into two error models, one for aliasing error and one for zipper error (on-off pattern caused by the interpolation process). The aliasing error is obtained by combining error masks from the red, green and blue channels. For the zipper error a procedure selecting the mask pixels belonging to an on-off pattern, and their sum resulting in the final error. The metric has been evaluated on a data set interpolated with 5 different interpolation algorithms. The results indicate a correlation with the mean opinion score.

2.9.27 Imai et al.

Imai et al. [2001] proposed a color difference metric based on Mahalanobis distance [Mahalanobis, 1936] by using covariance matrices for differences of lightness, chroma and hue angle. The CIE94 color difference can be derived from the simplified Mahalanobis perceptual difference.

2.9.28 Bouzerdoum et al., NNET

Bouzerdoum et al. [2004] proposed a method (NNET) for image quality assessment by using multilayer perception (MLP), based on neural network. The MLP is designed to extract a set of features from the original image to predict image fidelity. The features, the two mean values, the two standard deviations, covariance and MSE, are extracted from blocks in the image, and are fed as inputs to the network. Evaluation of the measure was done on the LIVE database with 354 pairs of reference and test images. The results indicate that NNET outperforms SSIM both in correlation coefficient, RMSE, MAE and standard error.

2.9.29 Carnec et al., Quality Assesser

Carnec et al. [2003] proposed the Quality Assesser. This is a reduced-reference metric, but is referred here because the input to the system is the original and reproduction. The metric is divided into two parts, perceptual representation and structural features extraction. The perceptual representation includes a low level processing and a perceptual sub-band decomposition. In the structural features extraction of a reduced representation is found based on fixation points. This is done for both the original and reproduction. The similarity measure is based on the processed original and reproduction where different similarity measures were tested. The best measure was based on structural information, other measures based solely on contrast or a combination of contrast and structure did not perform as well. The overall results indicate a strong correlation between MOS and calculated quality.

2.9.30 Teo and Heeger

Teo and Heeger [1994] described a perceptual distortion model that is consistent with spatial pattern psychophysics, it explains both contrast and orientation masking. The metric uses the steerable pyramid transform which decomposes the image into several spatial frequency and orientation bands. The detection of visible distortion is done locally with a simple squared-error norm.

2.9.31 Heeger and Teo, PDM

A model for perceptual image fidelity (PDM) was proposed by Heeger and Teo [1995]. This model is an extension of the 1994 model for perceptual distortion. The model accounts for contrast sensitivity, luminance masking and contrast masking. The model consists of 3 parts, a retinal component (responsible for contrast sensitivity and mean luminance masking), a cortical component (responsible for contrast masking) and the last component is a detection mechanism. The detection mechanism takes the distorted image and the original image, both normalized, and calculates the fidelity. The authors show an example of how PDM can predict image fidelity, where 3 JPEG compressed versions of an image are computed with approximately the same MSE.

2.9.32 Yao et al., VQM_ESC

Yao et al. [2005] proposed a visual quality metric considering error and contrast. The predicted quality is modeled by measuring error spread and isotropic local contrast. The quality index is the ratio of error spread normalized with isotropic local contrast. The results indicate a good correlation with MOS. This measure is also extended to video quality.

2.9.33 An et al., MHI

An et al. [2005] proposed an objective image quality measure, MHI (mean homogeneity index), based on homogeneity for grayscale images. The proposed measure takes the image structure into account and uses a measure based on second derivative masks to determine the local image homogeneity. The more uniform the region surrounding the pixel, the larger the homogeneity. The output is a homogeneity map, which can be averaged to obtain a measure for overall image quality. The results indicate better performance by MHI than SSIM and PSNR.

2.9.34 Xu et al.

Xu et al. [2005] proposed an metric for HDR images based on the logarithmic response of the human visual system. The metric uses the log[RGB] color space root-mean-square-error. The authors convert the pixles into a logarithmic RGB color space and then they compute the traditional RMS error.

2.9.35 Orfanidou et al., ”Busyness”

Orfanidou et al. [2008] proposed a ”busyness” metric for image quality. Ofranidou et el. defines ”busyness” as the presence or absence of details in the scene. The measure is based on a simple segmentation technique, using a Sobel filter and basic morphology functions. This metric relates directly to the spatial frequencies in the image.

2.9.36 Gorley and Holliman, SBLC

Gorley and Holliman [2008] proposed an image quality metric for stereoscopic images. The Stereo Band Limited Contrast (SBLC) accounts for HVS sensitivity to contrast in luminance changes in regions with high spatial frequency. The metric incorporate Michelson’s contrast and different algorithms to extract edges, corners and regions of high spatial frequency.

2.9.37 Other image quality metrics

An overview of several image quality metrics can be found in [Beaton, 1983; Dosselmann and Yang, 2005; Eskicioglu et al., 1995; Silva et al., 2007]. Some of the metrics described are:

- Average Difference
- Structural Content
- Normalized Cross-Correlation
- Correlation Quality
- Maximum Difference
- Image Fidelity
- Laplacian Mean Square Error
- Normalized Absolute Error
- Normalized Mean Square Error
- Czenakowski Distance
- Minkowsky Metric
- Lowest Frequency Component

- Average distance beween nearest neighboring minority-pixel pairs
- Frequency Weighted Mean Square Error
- HVS Weighted Mean Square Error
- Perceptual Equivalent Passband
- Equivalent Width
- Squared Spatial Frequency
- Modulation Transfer Function Area
- Gray Shade Frequency Product
- Integrated Contrast Sensitivity
- Perceived Modulation Ratio
- Information Content

Several of these metrics were evaluated by Eskicioglu et al. [1995] for gray scale compression

For a detail overview on research on perceptual video quality metrics see Winkler [1999] and Wang et al. [2003a]. A survey of no-reference and reduced-reference metrics is done by Engelke and Zepernick [2007].

3 Discussion and Conclusion

Evaluation of the metrics is an important step in the development of a full-reference image quality metric, without this step the performance of the metric cannot be determined. Unfortunately this is not always done to the extent that is needed in order to show the metric's performance to predict perceived difference or quality. More than 15 of the metrics reviewed have only been tested on one scene, either subjectively or objectively. This is not enough to reveal important aspects of the metric's performance. Even so some metrics have been extensively tested and the development and availability of the LIVE database [Sheikh and Bovik, 2006; Sheikh et al., 2006, 2007; Wang and Bovik, 2002; Wang et al., 2004] has resulted in more extensive testing of several metrics. This database contains JPEG compression (169 images), JPEG2000 compression (175 images), gaussian blur (145 images), white noise (145 images) and bit errors in JPEG2000 bit stream (145 images) [1] with corresponding difference mean opinion score (DMOS). Many of the metrics based on the UIQ [Wang and Bovik, 2002] have been tested on scenes from this database. One key element to the popularity of this database is the fact that is both free and available online. The use of this database makes it easy to compare metrics since they are tested with the same dataset. Also recently a new database was proposed, the TID2008 by Ponomarenko et al. [2008]. 654 observers judged 25 originals with 17 types of distortion on 4 levels. The goal of this database is to evaluate the performance of image quality metrics, and to compare and develop new metrics.

Many metrics are also benchmarked against the performance of only MSE or RMS. Many researchers have shown that these measures do not predict image difference or image quality very well. Newly developed full-reference metrics should therefore be compared against other state of the art metrics in order to determine their performance. UIQ [Wang and Bovik, 2002], SSIM [Wang et al., 2004], iCAM [Fairchild and Johnson, 2002], Quality Assesser [Carnec et al., 2003] and S-CIELAB [Zhang et al., 1997a] are available online, making a comparison against these easy. Some authors have tested their metrics against these, but still more researchers should test their newly developed metrics against other the commonly used metrics. The SSIM and S-CIELAB have been tested thoroughly, the availability online could be one reason for this.

By making the metrics available for other researchers it is also easier to discover advantages and disadvantages with the metrics, and therefore helping the development of new and better metrics.

We have given an extensive survey of full-reference image quality metrics with the aim to help researchers to find the most appropriate metric and improve the state of the art knowledge in this field. This survey will help researchers developing an universal full-reference image quality metric.

4 Acknowledgements

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

The author hereof has been enabled by Océ-Technologies B.V. to perform research activities which underlies this document. This document has been written in a personal capacity. Océ-Technologies B.V. disclaims any liability for the correctness of the data, considerations and conclusions contained in this document.

[1] Numbers from LIVE database Release 2 [Sheikh et al., 2007] (10/12/07)

References

A. J. Ahumada Jr. Computational image quality metrics: a review. In *SID International Symposium Digest of Technical Paper*, pages 305–308, 1993.

E. Allen, S. Triantaphillidou, and R. E. Jacobson. Image quality of jpeg and jpeg 2000 image compression schemes, part 1: Psychophysical measurements. *J. Elec. Imaging*, 13(1):126–138, 2004.

K. An, J. Sun, and W. Du. Homogeneity based image objective metric. In *Proceedings of 2005 International symposium on Intelligent Signal Processing and Communication Systems*, pages 649–652, Hong Kong, December 2005.

N. Avadhanam and V. R. Algazi. Evaluation of a human-vision-system-based image fidelity metric for image compression. In A. G. Tescher, editor, *Applications of Digital Image Processing XXII*, volume 3808 of *Proc. SPIE*, pages 569–579, San Jose, CA, USA, 1999.

M.A. Ben 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, 2002.

J. Bai, T. Nakaguchi, N. Tsumura, and Y. Miyake. Evaluation of image corrected by retinex method based on S-CIELAB and gazing information. *IEICE trans. on Fundamentals of Electronics, Communications and Computer Sciences*, Vol.E89-A(11):2955–2961, 2006.

E. Bando, J. Y. Hardeberg, and D. Connah. Can gamut mapping quality be predicted by color image difference formulae? In S. Daly B. Rogowitz, T. Pappas, editor, *Human Vision and Electronic Imaging X*, volume 5666 of *SPIE Proceedings*, pages 180 – 191, 2005.

P. G. J. Barten. Evaluation of subjective image quality with the square-root integral method. *Journal of the Optical Society of America A*, 7:2024–2031, Oct 1990.

P. G. J. Barten. *Contrast sensitivity of the human eye and its effect on image quality*, chapter 3, pages 25–64. HV Press, 1999.

R. J. Beaton. Quantitative models of image quality. In *Proceedings of the Human Factors Society - 27th Annual Meeting*, pages 41–45, 1983.

A. Beghdadi and R. Iordache. Image quality assessment using the joint spatial/spatial-frequency representation. *EURASIP J. Appl. Signal Process.*, (1):1–8, 2006. ISSN 1110-8657. doi: http://dx.doi.org/10.1155/ASP/2006/80537.

A. Beghdadi and B. Pesquet-Popescu. A new image distortion measure based on wavelet decomposition. In *Seventh International Symposium on Signal Processing and Its Applications*, volume 1, pages 485–488, Jul 2003.

M. Bolin and G. Meyer. A perceptually based adaptive sampling algorithm. In *SIGGRAPH '98 Conference Proceedings*, pages 409–418. ACM, 1998.

M. Bolin and G. Meyer. A visual difference metric for realistic image synthesis. In *Human Vision and Electronic Imaging IV*, volume 3644 of *SPIE*, pages 106–120, 1999.

N. Bonnier, F. Schmitt, H. Brettel, and S. Berche. Evaluation of spatial gamut mapping algorithms. In *14th Color Imaging Conference*, volume 14, pages 56–61. IS&T/SID, Nov 2006.

A. Bouzerdoum, A. Havstad, and A. Beghdadi. Image quality assessment using a neural network approach. In *Proceedings of the Fourth IEEE International Symposium on Signal Processing and Information Technology*, pages 330–333, 2004.

S. Bouzit and L. MacDonald. Colour difference metrics and image sharpness. In *8th Color Imaging Conference*, pages 262–267, Scottsdale, Arizona, 2000. IS&T/SID. ISBN / ISSN: 0-89208-231-3.

A. P. Bradley. A wavelet visible difference predictor. *IEEE Transactions on Image Processing*, 8:717–730, 1999.

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)*, volume 4, pages 2957–2960, 1991. doi: 10.1109/ICASSP.1991.151023.

M. H. Brill, J. Lubin, and D. Wolin. Perceptual scaling of quality metrics for hardcopy image evaluation. In *IS&T's NIP 15: International Conference on Digital Printing Technologies*, volume 15, pages 435–438, Orlando, Florida, Oct 1999. ISBN / ISSN: 0-89208-222-4.

A. C. Brooks and T. N. Pappas. Structural similarity quality metrics in a coding context: exploring the space of realistic distortions. In Bernice E. Rogowitz, Thrasyvoulos N. Pappas, and Scott J. Daly, editors, *Human Vision and Electronic Imaging XI*, volume 6057 of *SPIE Proceedings*, pages 299–310, San Jose, CA, USA, Feb 2006. SPIE. doi: 10.1117/12.660611.

C. R. Carlson and R. W. Cohen. A simple psychophysical model for predicting the visibility of displayed information. In *Proc. Soc. Inform. Display*, volume 21, pages 229–245, 1980.

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)*, pages 185–188, 2003.

N. Chaddha and T.H.Y. Meng. Psycho-visual based distortion measures for monochrome image and video compression. In *Proc. of the Asilomar Conference on Signals, Systems and Computers*, volume 2, pages 841–845, 1993. doi: 10.1109/ACSSC.1993.342451.

G. Chen, C. Yang, and S. Xie. Gradient-based structural similarity for image quality assessment. In *IEEE International Conference on Image Processing*, pages 2929–2932, Atlanta, GA, USA, October 2006a.

G.H. Chen, C.L. Yang, L.M. Po, and S.L. Xie. Edge-based structural similarity for image quality assessment. In *Proceedings of IEEE International Conference in Acoustics, Speech and Signal Processing*, volume 2, pages 933–936, 2006b.

S. Chen, A. Beghdadi, and A. Chetouani. Color image assessment using spatial extension to CIE DE2000. In *The International Conference on Consumer Electronics (ICCE)*, pages 1–2, Jan 2008. ISBN: 978-1-4244-1458-1.

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*, 5:467–476, 1995.

C. Chou and K. Liu. A fidelity metric for assessing visual quality of color images. In *Proceedings of 16th International Conference on Computer Communications and Networks, 2007. ICCCN 2007.*, pages 1154–1159, Aug 2007.

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*, 100:128–132, Apr 1984.

Commission Internationale de l'E'clairage. Colorimetry, 2nd ed.publication CIE 15.2, bureau central de la CIE, 1986.

Commission Internationale de l'Eclairage. Industrial colour-difference evaluation, publication CIE 116-95, bureau central de la CIE, 1995.

Commission Internationale de l'Eclairage. CIE 142-2001: Improvement to industrial colour-difference evaluation. Technical report, 2001. ISBN 978 3 901906 08 4.

S. Daly. The visible differences predictor: an algorithm for the assessment of image fidelity. In *Digital images and human vision*, pages 179–206. MIT Press, Cambridge, MA, USA, 1993. ISBN 0-262-23171-9.

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*, 9:636–650, 2000.

R. de Freitas Zampolo and R. Seara. A measure for perceptual image quality assessment. In *International Conference on Image Processing. ICIP 2003*, 2003.

R. de Freitas Zampolo and R. Seara. Perceptual image quality assessment based on bayesian networks. In *International Conference on Image processing (ICIP2004)*, 2004.

R. de Freitas Zampolo and R. Seara. A comparison of image quality metrics performrmances and practical conditions. In *IEEE international conference on Image processing*, volume 3, pages 1192–1195. IEEE, September 2005.

G. Deffner, M. Yuasa, and D. Arndt. Evaluation of display image quality: experts vs. non-experts. In *Symp Soc Inf Disp Dig*, volume 25, pages 475–478, 1994.

W. Dong, Q. Yu, C. N. Zhang, and H. Li. Image quality assessment using rough fuzzy integrals. In *Proceedings of the 27th International Conference on Distributed Computing Systems Workshops (ICDCSW'07)*, pages 1–5, Washington, DC, USA, 2007. ISBN 0-7695-2881-3. doi: http://dx.doi.org/10.1109/ICDCSW.2007.114.

R. Dosselmann and X. D. Yang. Existing and emerging image quality metrics. In *IEEE CCECE/CCGEI*, pages 1906 – 1913, 2005.

F. Dugay. Perceptual evaluation of colour gamut mapping algorithms. Master thesis, Gjøvik University College and Grenoble Institute of Technology, 2007.

F. Dugay, I. Farup, and J. Y. Hardeberg. Perceptual evaluation of color gamut mapping algorithms. *Color Research & Application*, 33(6):470–476, Dec 2008.

M. P. Eckert and A. P. Bradley. Perceptual quality metrics applied to still image compression. *Signal Process.*, 70(3):177–200, 1998. doi: http://dx.doi.org/10.1016/S0165-1684(98)00124-8.

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 Second International Workshop on Video Processing and Quality Metrics*, Scottsdale, USA, 2006.

P. G. Engeldrum. *Psychometric Scaling, a toolkit for imaging systems development*. Imcotek Press Winchester USA, 2000.

U. Engelke and H-J. Zepernick. Perceptual-based quality metrics for image and video services: A survey. In *3rd EuroNGI Conference on Next Generation Internet Networks*, 2007.

A. Eskicioglu. Quality measurement for monochrome compressed images in the past 25 years. In *Proceedings of the International Conference on Acoustics, Speech and Signal Processing*, pages 1907–1910, 2000.

A.M. Eskicioglu, P.S. Fisher, and S. Chen. Image quality measures and their performance. In *IEEE Transactions on Communications*, volume 43, pages 2959–2965, Dec 1995.

M. D. Fairchild and G. M. Johnson. Meet iCAM: A next-generation color appearance model. In *IS& T/SID Tenth Color Imaging Conference*, pages 33–38, 2002.

M. D. Fairchild and G. M. Johnson. The iCAM framework for image appearance, image differences, and image quality. *Journal of Electronic Imaging*, 13:126–138, 2004.

J.P. Farrugia and B. Peroche. A perceptual image metric in computer graphics. In *International conference on color in graphics and image processing (CGIP2000)*, pages 13–17, 2000.

J.P. Farrugia, S. Albin, and B. Peroche. A perceptual adaptive image metric for computer graphics. In *The 12th International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision (WSCG2004)*, 2004.

X-F. Feng, J. Speigle, and A. Morimoto. Halftone quality evaluation using color visual models. In *PICS 2002: IS&T's PICS Conference, An International Technical Conference on Digital Image Capture and Associated System, Reproduction and Image Quality Technologies, April 2002, Portland, Oregon*, pages 5–10, 2002.

S. R. Fernandez, G. M. Johnson, and M. D. Fairchild. Statistical summaries of iCAM image-difference maps. In *PICS*, pages 108–113, Rochester, NY, May 2003. ISBN / ISSN: 0-89208-245-3.

J. A. Ferwerda and F. Pellacini. Functional difference predictors (FDPs): measuring meaningful image differences. In *Conference Record of the Thirty-Seventh Asilomar Conference on Signals, Systems and Computers*, volume 2, pages 1388 – 1392, November 2003.

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

X. Gao, T. Wang, and J. Li. A content-based image quality metric. *Lecture notes in computer science*, 3642:231–240, 2005.

D. Gayle, H. Mahlab, Y.l Ucar, and A. M. Esckicioglu. A full-reference color image quality measure in the DWT domain. In *13th European Signal Processing Conference, EUSIPCO*, September 2005.

M. J. Gervais, L. O. Harvey Jr., and J.O. Roberts. Identifications confusions among letters of the alphabet. *Journal of experimental psychology: human perception and performance*, 10(4):655–666, 1984.

P. Gorley and N. Holliman. Stereoscopic image quality metrics and compression. In *Stereoscopic Displays and Applications XIX*, 2008.

E.M. Granger. A vision based RGB color space. In *TAGA*, pages 318–332, 2001.

E.M. Granger. A comparison of color difference data and formulas. In *TAGA 60th Annual Technical Conference*, San Francisco, CA, USA, March 2008.

I. Guarneri, M. Guarnera, A. Bosco, and G. Santoro. A perceptual quality metric for color-interpolated images. In Rene Rasmussen and Yoichi Miyake, editors, *Image Quality and System Performance II*, volume 5668 of *SPIE Proceedings*, pages 61–69, San Jose, CA, USA, Jan 2005.

S. L. Guth and H. R. Lodge. Vision: Heterochromatic additivity, foveal spectral sensitivity, and a new color model. *Journal of the Optical Society of America A (1917-1983)*, 63:450–462, April 1973.

J.Y. Hardeberg, E. Bando, and M. Pedersen. Evaluating colour image difference metrics for gamut-mapped images. *Coloration Technology*, 124(4):243–253, Aug 2008.

D. J. Heeger and P. C. Teo. A model of perceptual image fidelity. In *ICIP '95: Proceedings of the International Conference on Image Processing*, volume 2, page 2343, Washington, DC, USA, 1995. IEEE Computer Society. ISBN 0-8186-7310-9.

D. W. Hertel. Exploring s-CIELAB as a scanner metric for print uniformity. In Rene Rasmussen and Yoichi Miyake, editors, *Image Quality and System Performance II*, volume 5668 of *SPIE Proceedings*, pages 51–60, 2005.

I. E. Heynderickx and S. Bech. Image quality assessment by expert and non-expert viewers. In B. E. Rogowitz and T. N. Pappas, editors, *Human Vision and Electronic Imaging VII*, volume 4662 of *Proc. SPIE*, pages 129–137, June 2002.

G. . Hong and M. R. Luo. New algorithm for calculating perceived colour difference of images. *Imaging Science Journal*, 54(2):86–91, 2006.

G. Hong and M.R. Luo. Perceptually based colour difference for complex images. In R. Chung and A. Rodrigues, editors, *9th Congress of the International Colour Association*, volume 4421 of *Proceedings of SPIE*, pages 618–621, 2002.

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

R. Iordache and A. Beghdadi. A wigner-ville distribution-based image dissimilarity measure. In *Proceedings of the 6th International Symposium on Signal Processing and Its Applications (ISSPA '01)*, volume 2, pages 430–433, Kuala Lumpur, Malaysia, August 2001.

G. Ivkovic and R. Sankar. An algorithm for image quality assessment. In *IEEE International Conference on Acoustics, Speech, and Signal Processing, (ICASSP '04)*, pages 713–716, May 2004.

E. Jin and S. Field. A groundtruth database for testing objective metrics for image difference. In *PICS 2003: The PICS Conference, An International Technical Conference on The Science and Systems of Digital Photography, including the Fifth International Symposium on Multispectral Color Science, Rochester, NY, USA*, pages 114–119, 2003.

E. W. Jin, X-F. Feng, and J. Newell. The development of a color visual difference model (CVDM). In *PICS*, pages 154–158, Portland, Oregon, May 1998. IS&T.

G. M. Johnson and M. D. Fairchild. Darwinism of color image difference models. In *The 9th Color Imaging Conference: Color Science and Engineering: Systems, Technologies, Applications*, pages 108–112, 2001.

G. M. Johnson and M. D. Fairchild. A top down description of S-CIELAB and CIEDE2000. *Color Research and Application*, 28(6):425–435, December 2003.

M. Jourlin and J.C. Pinoli. A model for logarithmic image processing. *Journal of Microscopy*, 149(1):21–35, Januar 1988.

S. A. Karunasekera and N. G. Kingsbury. A distortion measure for image artifacts based on human visual sensitivity. In *Proc. of the IEEE International conference on Acoustics, Speech and Signal Processing*, volume 5, pages 117–120, 1994.

S. A. Karunasekera and N. G. Kingsbury. A distortion measure for blocking artifacts in images based on human visual sensitivity. *IEEE TIP*, 4:713–724, 1995.

J-S. Kim, M-S. Cho, S. Westland, and M. R. Luo. Image quality assessment for photographic images. In *AIC Colour 05 - 10th Congress of the International Colour Association*, 2005.

J-S. Kim, M-S. Cho, and B-K. Koo. *Entertainment Computing - ICEC 2006*, chapter Experimental Approach for Human Perception Based Image Quality Assessment, pages 59–68. Springer Berlin / Heidelberg, 2006.

R. Kimmel, M. Elad, D. Shaked, R. Keshet, and I. Sobel. A variational framework for retinex. *Int. J. Comput. Vision*, 52(1):7–23, April 2003. ISSN 0920-5691. doi: 10.1023/A:1022314423998. URL `http://portal.acm.org/citation.cfm?id=633022`.

R. Kimmel, D. Shaked, M. Elad, and I. Sobel. Space-dependent color gamut mapping: a variational approach. *Image Processing, IEEE Transactions on*, 14 (6):796–803, 2005. URL `http://ieeexplore.ieee.org/xpls/abs\_all.jsp?arnumber=1430768`.

Y. Kipman. Image quality metrics for printers and media. In *PICS: IS&T's 1998 Image Processing, Image Quality, Image Capture, Systems Conference*, Portland, Oregon, USA, May 1998. ISBN 0-89208-211-9.

Y.-K. Lai, C.-C. J. Kuo, and J. Li. New image compression artifact measure using wavelets. In J. Biemond and E. J. Delp, editors, *Visual Communications and Image Processing*, volume 3024 of *Proc. SPIE*, pages 897–908, January 1997.

Y.-K. Lai, J. Guo, and C.-C. J. Kuo. Perceptual fidelity measure of digital color images. In B. E. Rogowitz and T. N. Pappas, editors, *Human Vision and Electronic Imaging III*, volume 3299 of *Proc. SPIE*, pages 221–231, July 1998.

E. P. Lam and K. C. Loo. An image similarity measure using homogeneity regions and structure. In Susan P. Farnand and Frans Gaykema, editors, *Image Quality and System Performance V*, volume 6808, 2008.

C. Lambrecht and J. Farrell. Perceptual quality metric for digitally coded color images. In *The European Signal Processing Conference*, Trieste, Italy, 1996. URL `citeseer.ist.psu.edu/article/vandenbrandenlambrec96perceptual.html`.

E. C. Larson and D. M. Chandler. Unveiling relationships between regions of interest and image fidelity metrics. In William A. Pearlman, John W. Woods, and Ligang Lu, editors, *Visual Communications and Image Processing*, volume 6822 of *SPIE Proceedings*. SPIE, Jan 2008.

Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner. Gradient-based learning applied to document recognition. volume 86, pages 2278–2324, November 1998.

J. Lee and T. Horiuchi. Image quality assessment for color halftone images based on color structural similarity. *IEICE Trans. Fundamentals*, E91A:1392–1399, 2008.

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*, 51(5):391–401, September/October 2007.

M-S. Lee, L-Y. Liu, and F-S. Lin. Image similarity comparison using dual-tree wavelet transform. In *Proceedings of PSIVT, Advances in Image and Video Technology, First Pacific Rim Symposium*, pages 189–197, Dec 2006.

B. Li, G. Meyer, and R. Klassen. A comparison of two image quality models. In *Human Vision and Electronic Imaging III*, volume 3299 of *SPIE proceedings*, pages 98–109, 1998.

Q. Lin. Halftone image quality analysis based on a human vision model. In J. P. Allebach and B. E. Rogowitz, editors, *Proc. SPIE, Human Vision, Visual Processing, and Digital Display IV*, volume 1913, pages 378–389, September 1993.

C. Liu, G. M. Johnson, M. D. Fairchild, and G. Braun. Perception and modeling of halftone image quality using a high-resolution LCD. In *Thirteenth Color Imaging Conference*, pages 165–170, November 2005.

J. Lubin. *Vision Models for Target Detection and Recognition*, chapter A visual discrimination model for imaging systems design and evaluation, pages 245 –

283. World Scientific Publishing Co, Inc., River Edge, NJ, USA, 1995. Singapore: World Scientific.

J. Lubin. A human vision system model for objective picture quality measurements. In *International Broadcasting Convention*, pages 498–503, 1997a.

J. Lubin. Sarnoff JND vision model: Algorithm description and testing. Technical report, Sarnoff Corporation, 1997b. ftp://ftp.its.bldrdoc.gov/dist/ituvidq/old2/jrg003.rtf.

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*, 103:86–94, 1987.

M.R. Luo, G. Cui, and B. Rigg. The development of the CIE 2000 colour-difference formula: CIEDE2000. *Color Research and Application*, 26(5):340–350, 2001.

P.C. Mahalanobis. On the generalized distance in statistics. In *Proceedings National Institute of Science*, volume 12, pages 49–55, India, 1936.

S. Mallat and W.L. Hwang. Singularity detection and processing with wavelets. *IEEE Transactions on Information Theory*, 38:617–643, Mar 1992.

S. Mallat and S. Zhong. Characterization of signals from multiscale egdes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14:710–732, Jul 1992.

J.L. Mannos and D.J. Sakrison. The effects of a visual fidelity criterion on encoding of images. *IEEE transactions on information theory*, 20(4):525–536, 1974.

R. Mantiuk, K. Myszkowski, and H.-P. Seidel. Visible difference predicator for high dynamic range images. In *IEEE International Conference on Systems, Man and Cybernetics*, volume 3, pages 2763–2769, October 2004.

R. Mantiuk, S. Daly, K. Myszkowski, and H.P. Seidel. Predicting visible differences in high dynamic range images: model and its calibration. In Bernice E. Rogowitz, Thrasyvoulos N. Pappas, and Scott J. Daly, editors, *Human Vision and Electronic Imaging X*, volume 5666 of *SPIE proceedings*, pages 204–214, San Jose, CA, USA, Jan 2005. SPIE.

A. Mayache, T. Eude, and H. Cherifi. A comparison of image quality models and metrics based on human visual sensitivity. In *IEEE int. Conf. Image Proc.*, volume 3, pages 409–413. IEEE, oct 1998.

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*, pages 680–688, Salt Lake City, UT, October 2004. IS&T.

F. Mindru and J. Jung. Structure and color based comparison of halftone images. In *International Congress of Imaging Science (ICIS06)*, 2006.

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*, pages 14–16, 1994.

T. Mitsa and K.L. Varkur. Evaluation of contrast sensitivity functions for the formulation of quality measures incorporated in halftoning algorithms. In *Proc. IEEE Int. Conf. Acostics, speech, signal processing*, volume 3, pages 313–316, 1992.

M. Miyahara, K. Kotani, and Ralph Algazi. Objective picture quality scale (PQS) for image coding. *IEEE TRANSACTIONS ON COMMUNICATIONS*, 46(9): 1215–1226, Sep 1996.

J. Morovic and P. Sun. Visual differences in colour reproduction and their colorimetric correlates. In *10th Color Imaging Conference*, pages 292–297, Scottsdale, Arizona, 2002. IS&T/SID.

K. T. Mullen. The contrast sensitivity of human colour vision to red-green and blue-yellow chromatic gratings. *The Journal of Physiology*, 359:381–400, 1985.

J. Munkberg, P. Clarberg, J. Hasselgren, and T. Akenine-Möller. High dynamic range texture compression for graphics hardware. *ACM Transactions on Graphics*, 25(3):698–706, July 2006.

S. Nakauchi, S. Hatanaka, and S. Usui. Color gamut mapping based on a perceptual image difference measure. *Color Research and Application*, 24:280 –291, 1999.

R. Näsänen. Visibility of halftone dot textures. *T-SMC*, 14:920–924, 1984.

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.

L. Neumann, K. Matkovic, and W. Purgathofer. Perception based color image difference. In *Computer Graphics Forum*, volume 17, pages 233 – 242, Sep 1998.

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*, 14:2111–2127, September 1997.

N. Nill. A visual model weighted cosine transform for image compression and quality assessment. *IEEE Transactions on Communications (COM)*, 33:551–557, 1985.

N.B. Nill and B.H. Bouzas. Objective image quality measure derived from digital image power spectra. *Optical Engineering*, 31(4):813–825, April 1992.

F. Nilsson. Objective quality measures for halftoned images. *Journal of Optical Society of America A*, 16:2151–2162, 1999.

F. Nilsson and B. Kruse. Objective quality measures of halftones images. In *IS&T's NIP 13: International Conference on Digital Printing Conference*, pages 353–357, Mar 1997.

M. Orfanidou, S. Triantaphillidou, and E. Allen. Predicting image quality using a modular image difference model. In Susan P. Farnand and Frans Gaykema, editors, *Image Quality and System Performance V*, volume 6808 of *SPIE Proceedings*, San Jose, USA, Jan 2008. SPIE/IS&T.

S. Ouni, M. Chambah, M. Herbin, and E. Zagrouba. Are existing procedures enough? image and video quality assessment: Review of subjective and ojective metrics. In Susan P. Farnand and Frans Gaykema, editors, *Image Quality and System Performance V*, volume 6808 of *SPIE Proceedings*, San Jose, CA, USA, Jan 2008. SPIE.

T. N. Pappas and D. L. Neuhoff. Least-squares model-based halftoning. *IEEE Transactions on Image Processing*, 8(8):1102–1116, Aug 1999.

M. Pedersen. Importance of region-of-interest on image difference metrics. Master's thesis, Gjøvik University College, 2007.

M. Pedersen and J. Y. Hardeberg. Rank order and image difference metrics. In *CGIV 2008 Fourth European Conference on Color in Graphics, Imaging and Vision*, pages 120–125, Terrassa, Spain, Jun 2008. IS&T.

M. Pedersen and J. Y. Hardeberg. A new spatial hue angle metric for perceptual image difference. In *2009 Computational Color Imaging Workshop*, Saint Etienne, France, Mar 2009a.

M. Pedersen and J.Y. Hardeberg. Shame: A new spatial hue angle metric for perceptual image difference. In *Vision Sciences Society 9th Annual Meeting*, Naples, FL, USA, May 2009b.

M. Pedersen, J. Y. Hardeberg, and P. Nussbaum. Using gaze information to improve image difference metrics. In B. Rogowitz and T. Pappas, editors, *Human Vision and Electronic Imaging VIII (HVEI-08)*, volume 6806 of *SPIE proceedings*, San Jose, USA, Jan 2008. SPIE.

S. Pefferkorn and J-L. Blin. Perceptual quality metric of color quantization errors on still images. In Bernice E. Rogowitz and Thrasyvoulos N. Pappas, editors, *Human Vision and Electronic Imaging III*, volume 3299 of *SPIE Proceedings*, pages 210–220, 1998.

E. Peli. Contrast in complex images. *Journal of the Optical Society of America A*, 7:2032–2040, 1990.

A.B. Poirson and B.A. Wandell. Apperance of colored patterns: pattern-color separability. *Journal of the Optical Society of America A*, 10(12):2458–2470, 1993.

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*, pages 403–408, Cairns, Queensland, Australia, Oct 2008. `http://www.ponomarenko.info/tid2008.htm`.

H. Rushmeier, G. Ward, C. Piatko, P. Sanders, and B. Rust. Comparing real and synthetic images: Some ideas about metrics. In *Eurographics Rendering Workshop*, 1995.

R. J. Safranek and J. D. Johnston. A perceptually tuned sub-band image coder with image dependent quantization and post-quantization data compression. In *ICASSP*, volume 3, pages 1945–1948, 1989.

A. Samet, M.A.B. Ayed, N. Masmoudi, and L. Khriji. New perceptual image quality assessment metric. *Asian Journal of Information Technology*, 4(11): 996–1000, 2005.

C. Sano, T. Song, and M. R. Luo. Colour differences for complex images. In *IS&T/SID's Eleventh Color Imaging Conference: Color Science and Engineering Systems, Technologies, Applications*, pages 121–126, Scottsdale, Arizona, Nov 2003. ISBN / ISSN: 0-89208-248-8.

T. Scheermesser and O. Bryngdahl. Texture metric of halftone images. *J. Opt. Soc. Am. A*, 13:18–24, 1996.

T. Scheermesser and O. Bryngdahl. Spatially dependent texture analysis and control in digital halftoning. *J. Opt. Soc. Am. A*, 14:827–835, 1997.

T. Seim and A. Valberg. Towards a uniform color space: A better formula to describe the munsell and osa color scales. *Color research and applications*, 11 (1):11–24, 1986.

G. Sharma. *Digital Color Imaging Handbook*. CRC Press, Inc., Boca Raton, FL, USA, 2002. ISBN: 084930900X.

H. R. Sheikh. *Image Quality Assessment Using Natural Scene Statistics*. PhD thesis, The University of Texas at Austin, 2004.

H. R. Sheikh and A. C. Bovik. Image information and visual quality. *IEEE Transactions on Image Processing*, 15(2):430–444, 2006.

H. R. Sheikh, A. C. Bovik, and G. de Veciana. An information fidelity criterion for image quality assessment using natural scene statistics. *IEEE Trans. Image Processing*, 14(12):2117–2128, December 2004.

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*, 15(11):3440–3451, 2006.

H.R. Sheikh, Z. Wang, L. Cormack, and A.C. Bovik. LIVE image quality assessment database release 2. Online: http://live.ece.utexas.edu/research/quality, 2007.

A. Shnayderman, A. Gusev, and A. M. Eskicioglu. A Multidimensional image quality measure using singular value decomposition. In Yoichi Miyake and D. Rene. Rasmussen, editors, *Image Quality and System Performance*, volume 5294 of *Proceedings of SPIE*, pages 82–92, San Jose, CA, USA, December 2004. doi: 10.1117/12.530554.

A. Shnayderman, A. Gusev, and A. M. Eskicioglu. An svd-based grayscale image quality measure for local and global assessment. *IEEE Transactions On Image Processing*, 15(2):422–429, 2006.

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*, volume 6579 of *Presented at the Society of Photo-Optical Instrumentation Engineers (SPIE) Conference*, San Jose, CA, USA, Apr 2007. SPIE.

D. A. Silverstein and S. A. Klein. DCT image fidelity metric and its application to a text-based scheme for image display. In J. P. Allebach and B. E. Rogowitz, editors, *Proc. SPIE, Human Vision, Visual Processing, and Digital Display IV*, volume 1913, pages 229–239, San Jose, CA, USA, sep 1993. SPIE.

Tao Song and Ronnier Luo. Testing color-difference formulae on complex images using a crt monitor. In *IS&T/SID Eighth Color Imaging Conference*, pages 44–48, Scottsdale, Arizona, 2000. IS&T/SID.

J. Sullivan, L. Ray, and R. Miller. Design of minimum visual modulation halftone patterns. *IEEE Trans. Systems, Man, and Cybernetics*, 21:33–38, 1991.

J. Sullivan, R. Miller, and G. Pios. Image halftoning using a visual model in error diffusion. *J. Opt. Soc. Am. A*, 10:1714–1724, 1993.

P-L. Sun and J. Morovic. What differences do observers see in colour image reproduction experiments. In *CGIV 2002: The First European Conference on Colour Graphics, Imaging and Vision*, pages 181–186, Poitiers, France, April 2002. IS&T.

C. C. Taylor, J. P. Allebach, and Z. Pizlo. The image fidelity assessor. In *IS&Ts 1998 PICS Conference*, Portland, Oregon, USA, 1998. IS&T.

P. C. Teo and D. J. Heeger. Perceptual image distortion. *IEEE International Conference Image Processing*, 2:982 – 986, 1994.

A. Toet and M.P. Lucassen. A new universal colour image fidelity metric. *Displays*, 24:197–204, 2003.

O. Veryovka, A. Fournier, and J. W. Buchanan. Multiscale edge analysis of halftoned images. In B. E. Rogowitz and T. N. Pappas, editors, *Human Vision and Electronic Imaging III*, volume 3299 of *Proc. SPIE*, pages 461–472, San Jose, CA, USA, Jan 1998. SPIE.

X. Wan, D. Xie, and J. Xu. Quality evaluation of the halftone by halftoning algorithm-based methods and adaptive method. In Luke C. Cui and Yoichi Miyake, editors, *Image Quality and System Performance IV*, volume 6494 of *SPIE Proceedings*, San Jose, California, Jan 2007. SPIE.

Z. Wang and A.C. Bovik. A universal image quality index. *IEEE Signal Processing Letters*, 9:81–84, 2002.

Z. Wang and M. R. Luo. Experimental filters for estimating image differences. In *CGIV 2008 - Fourth European Conference on Color in Graphics, Imaging and Vision*, 2008.

Z. Wang and E. Simoncelli. Translation insensitive image similarity in complex wavelet domain. In *IEEE international conference on acoustics, speech and signal processing*, volume 2, pages 573–576, March 2005.

Z. Wang, H. R. Sheikh, and A. C. Bovik. *The Handbook of Video Databases: Design and Applications*, chapter Objective video quality assessment, pages 1041–1078. CRC Press, 2003a.

Z. Wang, E. P. Simoncelli, and A. C. Bovik. Multi-scale structural similarity for image quality assessment. In *Proceedings of the 37th IEEE Asilomar Conference on Signals, Systems and Computers*, Nov 2003b.

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*, 13(4):600–612, 2004.

A. B. Watson. DCT quantization matrices visually optimized for individual images. In Jan P. Allebach and Bernice E. Rogowitz, editors, *Human Vision, Visual Processing, and Digital Display IV*, volume 1913 of *SPIE Proceedings*, pages 202–216, San Jose, CA, USA, February 1993a. SPIE.

A.B. Watson. DCTune: a technique for visual optimization of dct quantization matrices for individual images. In *Soc. Inf. Display Digest Tech. Papers XXIV*, number 5, pages 946–949, Santa Ana, CA, USA, 1993b. SID.

E.H. Weber. *A Source Book in the History of Psychology*, chapter On Weber's law, pages 64–66. Hardvard University Press, 1965.

S. J. P. Westen, R. L. Lagendijk, and J. Biemond. Perceptual image quality based on a multiple channel HVS model. In *Proceedings ICASSP-95 (IEEE International Conference on Acoustics, Speech and Signal Processing)*, volume 4, pages 2351–2354, Detroit, Michigan, USA, 1995. IEEE.

D. L. Wilson, A. J. Baddeley, and R. A. Owens. A new metric for grey-scale image comparison. *Intern. J. Computer Vision*, 24:5–17, 1997.

S. Winkler. Issues in vision modeling for perceptual video quality assessment. *Signal Process.*, 78:231–252, 1999.

P. W. Wong. A mixture distortion criterion for halftones. Technical report, HP Labs, March 1997. URL `http://www.hpl.hp.com/techreports/97/HPL-97-49.html`.

W. Wu, Z. Pizlo, and J. P. Allebach. Color image fidelity assessor. In *IS&T's 2001 PICS Conference Proceedings*, pages 148–152, Monreal, Quebec, Canada, 2001. IS&T.

R. Xu, S. N. Pattanaik, and C. E. Hughes. High-dynamic-range still-image encoding in jpeg2000. *IEEE Computer Graphics and Applications*, 25(6):57–64, 2005.

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 *IEEE International Symposium on Circuits and Systems (ISCAS)*, volume 4, pages 3793 – 3796, Kobe, Japan, May 2005. IEEE.

S. Yao, W. Lin, Z.K Lu, E.P. Ong, M.H. Locke, and S.Q. Wu. Image quality measure using curvature similarity. *Image Processing, 2007. ICIP 2007. IEEE International Conference on*, 3:437–440, 2007. ISSN 1522-4880. doi: 10.1109/ICIP.2007.4379340.

H. Yee. A perceptual metric for production testing. *Journal of Graphics Tool*, 9 (4):33–40, 2004.

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*, volume 3299 of *SPIE proceedings*, pages 160–172, San Jose, CA, USA, Jan 1998. SPIE.

Q. Yu and K. J. Parker. Quality issues in blue noise halftoning. In *Proc. SPIE, Color Imaging: Device-Independent Color, Color Hardcopy, and Graphic Arts III, Giordano B. Beretta; Reiner Eschbach; Eds.*, volume 3300, pages 376–385, San Jose, CA, USA, Jan 1998. SPIE.

Q. Yu, K. J. Parker, R. Buckley, and V. Klassen. A new metric for color halftone visibility. In *IS&Ts PICS Conference*, pages 226–230, Portland, OR, USA, May 1998. IS&T - The Society for Imaging Science and Technology.

X. Zhang and B. A. Wandell. Color image fidelity metrics evaluated using image distortion maps. *Signal Processing - Special issue on image and video quality metrics*, 70:201 – 214, 1998.

X. Zhang and B.A. Wandell. A spatial extension of CIELAB for digital color image reproduction. In *Soc. Inform. Display 96 Digest*, pages 731–734, San Diego, 1996.

X. Zhang, J.E. Farrell, and B.A. Wandell. Application of a spatial extension to CIELAB. In *Very high resolution and quality imaging II*, volume 3025 of *SPIE proceedings*, pages 154–157, San Jose, CA, USA, Feb 1997a.

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 *COMP-CON97 Digest of Papers*, pages 44–48, Washington, DC, USA, 1997b. IEEE Computer Society.

APPENDIX

A Image quality metric overview

In Table 1 we show an overview of the metrics in this survey. The metrics are sorted in chronological order, then the name of the metric if given any by the authors, then the author names in column three. For the type column the metrics are grouped by type. ID = image difference, IQ = image quality, IF = image fidelity, CD = color difference, HT = halftoning, DP = difference predictor, VQ = video quality , IS = image similarity. For the column HVS, the metric must have a HVS model or CSF filter that simulates the HVS. The metrics like SSIM, who indirectly simulate the HVS is set to ”no”. MS indicate whether the metrics are multiscale. S/NS indicate whether the metric is spatial (FFS, FVS, VFS, VVS) or non-spatial (NS). We have divided the spatial metrics into 3 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, VVS = variable size of filter and variable calculation. Fixed size indicate that the filter, block or similar is fixed for the whole image, variable size indicate that the filter or block changes according to the image content. Fixed computation indicate the same calculation within the filter or block, variable calculation indicate that the calculation is dependent on the image content. C/G indicate whether the metric are for color or grayscale images. The test column indicate what kind for evaluation that has been performed. This is either objective or subjective, for the metrics where the authors has done the subjective test this is marked with (A). Scenes indicate 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 of the number of scenes, while the second number in ( ) indicate the total number of images (originals $\times$ number of modifications). For observers the total number of observers is stated, and inside the ( ) number of experts are stated if this information is given by the author. - indicate that this information is not available or not stated by the authors. The last column refers to the section where the metric is reviewed in the article.

Table 1: Metrics: ID = image difference, IQ = image quality, IF = image fidelity, CD = color difference, HT = halftoning, DP = difference predictor, VQ = video quality , IS = image similarity.

Year	Metric	Author(s)	Type	HVS	MS	S/NS	C/G	Test	Scenes	Modification	Observers	Comment	Section
1976	ΔE^*_{ab}	CIE	CD	No	No	NS	Color	-	-	-	-		2.2.1
1984	CMC	Clarke et al.	CD	No	No	NS	Color	-	-	-	-		2.2.2
1984		Näsänen	HT	Yes	No	FFS	Gray	sub.	4 (128)	Dot profiles	3	1 observers in the main experiment, 2 in control experiment.	2.9.8
1986	SVF	Seim and Valberg	CD	No	No	NS	Color	obj.	-		-	-	2.2.7
1987	BFD	Luo and Rigg	CD	No	No	NS	Color	-	-	-	-		2.2.3
1989		Safranek and Johnston	IQ	No	No	FFS	Gray	obj.	30	-			2.9.21
1990	SQRI	Barten	IQ	No	No	FFS	Gray	sub.	5 (35)	Resolution	20	sub. data from Westerink and Roufs.	2.9.16
1991	LIPMSE	Brailean et al.	IF	Yes	No	FFS	Gray	sub.	1 (1)	Gaussian blur	1 (A)	Metric used to restore a blurred image.	2.1.8
1992		Mitsa and Varkur	HT	Yes	No	VFS	Gray	sub.	11 (44)	Halftoning	12	3 metrics tested, same procedure but different CSFs.	2.1.4
1993	VDP	Daly	DP	Yes	No	FFS	Gray	-	-	-	-		2.5.1
1993		Watson	IQ	No	No	FFS	Gray	obj.	-	2 (8)	-		2.7.1
1993		Silverstein and Klein	IQ	No	No	FFS	Gray	obj.	-	-	-		2.7.2
1993		Lin	HT	Yes	No	FFS	Gray	obj.	1 (5)	Halftoning	-	CSF filter from Mannos et al. and Nill et al. tested.	2.1.5

Continued on Next Page…

Table 1 – Continued

Year	Metric	Author(s)	Type	HVS	MS	S/NS	C/G	Test	Scenes	Modification	Observers	Comment	Section
1993		Chaddha and Meng	IQ	No	No	FFS	Gray	sub.	6 (48)	Compression artifacts	60	Also extended to video. Only the results from one scene presented.	2.9.7
1994		Mitsa and Alford	HT	Yes	No	VFS	Gray	sub.	11 (44)	Halftoning			2.1.10
1994	E_θ	Karunasekera and Kingsbury	IQ	Yes	No	FFS	Gray	sub.	1 (8)	Lapped Orthogonal Transform	8	Observer expertise not stated.	2.9.19
1994		Teo and Heeger	IQ	Yes	Yes	VFS	Color	s/o	1 (2)	-	(A)		2.9.30
1995	ΔE_{94}	CIE [1995]	CD	No	No	NS	Color	-	-	-	-	Tested in various papers on various type of scenes.	2.2.4
1995	PDM	Heeger and Teo	IF	Yes	Yes	VFS	Gray	sub.	1 (3)	JPEG	(A)	Extension of Teo and Heeger [1994]	2.9.31
1995	PSPNR	Chou and Li	IQ	Yes	No	FFS	Gray	obj.	-	-	-		2.1.16
1995		Rushmeier et al.	IS	Yes	No	FFS	Gray	obj.	Various	Synthetic images	-	Comparing real and synthetic images	2.1.9
1995	PEM	Westen et al.	IQ	Yes	No	VFS	Gray	sub.	6(105)	PCM, DPCM, DCT and SBC coding at different bit rates	7 (5)		2.9.22
1995	VDM	Lubin	IQ	Yes	Yes	VFS	Gray	sub /obj	-	-	-	Various testing	2.6.1
1996	S-CIELAB	Zhang and Wandell	ID, HT	Yes	No	FFS	Color	obj.	-	JPEG-DCT, halftoning and patterns	-		2.3.1
1996	PQS	Miyahara et al.	IQ	Yes	No	FFS	Gray	sub.	5(25)	global and local distortion	9 (9)		2.9.20

Continued on Next Page…

Table 1 – Continued

Year	Metric	Author(s)	Type	HVS	MS	S/NS	C/G	Test	Scenes	Modification	Observers	Comment	Section
1996	CMPSNR	Lambrecht and Farrell	IQ	Yes	No	VFS	Color	sub.	1 (400)	JPEG	2		2.1.17
1996		Scheermesser and Bryngdahl	HT	No	No	VVS	Gray	obj.	2 (8)	Halftoning	-	Also tested on 2 test pattern.	2.9.14
1996	PQS	Miyahara et al.	IQ	Yes	No	FFS	Gray	sub.	5 (25)	-	9 (9)		
1997	Sarnoff JND Vision Model	Lubin	IQ	Yes	Yes	FFS	Color	sub.	5 (15)	MPEG-2 with different bit-rates	20	Also tested on JPEG data.	2.6.2
1997		Neumann et al.	ID	No	No	VFS	Color	obj.	-	-	-	-	2.3.12
1997		Wong	HT	No	No	FFS	Gray	-	-	-	-		
1997		Nijenhuis and Blommaert	IQ	No	No	NS	Gray	sub.	2 (25)	Interpolation	6		2.9.17
1997		Lai et al.	IF	Yes	Yes	VFS	Color	sub.	1(1)	JPEG2000	(A)		2.8.1
1997		Scheermesser and Bryngdahl	HT	No	No	VVS	Gray	obj.	2 (-)	Halftoning	-		2.9.15
1997	Δg	Wilson et al.	IS	No	No	VFS	Gray	sub.	1 (5)	JPEG, different distortion types	(A)		2.9.25
1998	IFA	Taylor et al.	IF	Yes	Yes	FFS	Gray	obj.	-	-	-	-	2.3.10
1998	CVDM	Jin et al.	ID	Yes	No	FFS	Color	obj.	-	-	-		2.3.8
1998		Lai et al.	IF	Yes	No	FFS	Color	sub.	1(1)	JPEG2000	(A)		2.8.1
1998		Yu et al.	HT	Yes	No	FFS	Color	sub.	6 (36)	Halftoning	8 (8)		2.9.4
1998	KLK	Yu and Parker	HT	No	No	FFS	Color	sub.	5 (20)	Halftoning	10		2.9.5
1998		Veryovka et al.	HT	Yes	Yes	VFS	Gray	obj.	1/1 (3/3)	Halftoning	-	-	2.8.2
1998	E	Yeh et al.	VQ	Yes	No	FFS	Gray	sub.	1 (9)	Block artifacts	8	Sequence of 64 frames.	2.9.11

Continued on Next Page…

Table 1 – Continued

Year	Metric	Author(s)	Type	HVS	MS	S/NS	C/G	Test	Scenes	Modification	Observers	Comment	Section
1998	CCETT visual metric	Pefferkorn and Blin	IQ	Yes	No	FFS	Color	sub.	6 (30)	MPEG-2	16		2.3.14
1998		Bolin and Meyer	DP	Yes	No	FFS	Color	sub.	-	-	(A)		2.6.3
1998		Franti	IQ	No	No	FFS	Gray	sub.	3 (42)	Compression	15-39	Possiblilty for color images	2.4.18
1999	WVDP	Bradley	DP	Yes	No	FFS	Gray	obj.	1 (3)	Noise	-		2.5.3
1999	PDM	Avadhanam and Al-gazi	IF	Yes	No	FFS	Color	sub.	5/5(50/75)	Compression	5(2)/5(2)		2.9.18
1999	LSMB	Pappas and Neuhoff	HT	Yes	No	FFS	Gray	-	-	-	-	Used for halftoning optimization	2.9.13
1999	QM_a	Nilsson	HT	Yes	No	VVS	Gray	obj.	1(3)	Halftoning	-		2.9.23
1999	PD	Nakauchi et al.	ID	Yes	No	FFS	Color	sub.	8 (48)	Gamut mapping	10	Used to optimize gamut mapping	2.3.4
2000	DM and NQM	Damera-Venkata et al.	IQ	No	No	FVS	Gray	sub.	-	Various	(A)		2.9.1
2000		Farrugia and Per-oche	ID	Yes	No	FFS	Color	obj.	4(8)		(A)		2.3.19
2001	ΔE_{00}	CIE [2001]	CD	No	No	NS	Color	-	-	-	-		2.2.5
2001	CIFA	Wu et al.	ID	Yes	Yes	FFS	Color	sub.	1(1)	Hue	(A)		2.3.11
2001		Johnson and Fairchild	IQ, HT	Yes	No	FFS	Color	sub.	1(72)	Sharpness	-	S-CIELAB with different CSF filters.	2.3.2
2001		Imai et al.	CD	No	No	NS	Color	sub.	6(12)	-	-		2.9.27
2001	SNR_W	Iordache and Begh-dadi [2001]	IS	No	No	FFS	Gray	sub.	1(3)	Salt-and-pepper noise, blurring, JPEG	5 (0)		2.1.11

Continued on Next Page…

Table 1 – Continued

Year	Metric	Author(s)	Type	HVS	MS	S/NS	C/G	Test	Scenes	Modification	Observers	Comment	Section
2002	iCAM	Fairchild and Johnson	ID	Yes	No	FFS	Color	-	-	-	-		2.3.6
2002		Hong and Luo	ID	No	No	NS	Color	obj.	2	Local color change	-		2.3.15
2002	UIQ	Wang and Bovik	IQ	No	No	FFS	Gray	sub.	1 (8)	Different distortion types	22		2.4.1
2002		Feng et al.	IQ	Yes	No	FFS	Color	sub.	2 (14)	Halftoning		Extension of CVDM	2.3.9
2002	ΔI_{cm}	Morovic and Sun	ID	No	No	FFS	Color	sub.	7 (32)	Gamut mapping	-		2.3.7
2002	WMSE	Ayed et al.	IQ	No	No	FFS	Gray	obj.	2 (14)	Noise	-	-	2.1.6
2003	Q_{color}	Toet and Lucassen	IF	No	No	FFS	Color	sub.	2 (21)	Quantization	4-16		2.4.2
2003	FDP	Ferwerda and Pellacini	DP	Yes	No	NS	Gray	sub.	8 (24)	Computer generated images	18		2.5.2
2003	MSSIM	Wang et al.	IQ	No	Yes	FFS	Gray	sub.	29 (344)	JPEG and JPEG2k	-	Scenes from LIVE	2.4.5
2003	$M-SVD$	Shnayderman et al.	IQ	No	No	FFS	Gray	sub.	5 (30)	JPEG, JPEG2k, G.Noise, G.Blur , Sharpening, DC shift	10 (5/5)	Color extension possible	2.9.12
2003	SNR_{WAV}	Beghdadi and Pesquet-Popescu	IQ	No	No	FFS	Gray	sub.	1 (3)	Gaussion noise, JPEG and grid pattern	>25		2.1.12
2003	*DQM* and *CQM*	de Freitas Zampolo and Seara	IQ	No	No	FFS	Gray	sub.	4 (45)	Frequency distortion	7		2.9.2
2003	Quality Assesser	Carnec et al.	IQ	Yes	No	VFS	Color	sub.	- (90)	JPEG and JPEG2k	-	Also tested on LIVE	2.9.29
2004	*B-CQM*	de Freitas Zampolo and Seara	IQ	No	No	NS	Gray	sub.	1 (81)	Frequency distortion and noise injection			2.9.3

Continued on Next Page…

Table 1 – Continued

Year	Metric	Author(s)	Type	HVS	MS	S/NS	C/G	Test	Scenes	Modification	Observers	Comment	Section
2004	SSIM	Wang et al.	IQ	No	No	FFS	Gray	sub.	29 (344)	JPEG and JPEG2k	-	Scenes from LIVE	2.4.4
2004		Ivkovic and Sankar	IQ	Yes	No	FFS	Gray	obj.	5 (20)	Contrast streching, white noise, blur, JPEG2k	-		2.9.6
2004	NNET	Bouzerdoum et al.	IQ	No	No	FFS	Gray	sub.	-	JPEG and JPEG2k	-	Scenes from LIVE	2.9.28
2004	I^*	McCormick-Goodhart et al.	IQ	No	No	FFS	Color	sub.	-	2 printsystems	-		2.9.10
2004	HDR-VDP	Mantiuk et al.	DP	No	Yes	VFS	Grey	sub.	-	Quantization, noise	(A)		2.5.4
2004	pdiff	Yee	ID	Yes	Yes	FFS	Color	-	-	Film production	-		2.3.13
2004	IFC	Sheikh et al.	IF	No	No	FFS	Gray	sub.	29 (344)	JPEG, JPEG200, Noise, Blur	20 - 25	Scenes from LIVE	2.8.3
2005	CBM	Gao et al.	IQ	No	No	FFS	Gray	sub.	29 (344)	JPEG and JPEG2k	-	Scenes from LIVE	2.4.9
2005	CWSSIM	Wang and Simoncelli	IS, IQ	No	No	FFS	Gray	obj.	1 (12)	Distortion as JPEG, noise etc.	-		2.4.11
2005	*PQSI*	Guarneri et al.	IQ	No	No	-	Color	sub.	-	5 Interpolation algorithms	-		2.9.26
2005	M-DWT	Gayle et al.	IQ	No	No	NS	Color	sub.	5 (30)	JPEG, JPEG2k, blur, noise, sharp and DC-shift	14	Stated as a color metric, but only operates on luminance	2.8.5
2005	VQM_ESC	Yao et al.	IQ	No	No	FFS	Gray	sub.	- (344)	JPEG and JPEG2k	-	Scenes from LIVE	2.9.32
2005	MHI	An et al.	IQ	No	No	FFS	Gray	sub.	1 (4)	JPEG, JPEG2k, gaussian noise and speckled noise	(A)		2.9.33
2005		Kimmel et al.	IS	No	Yes	FFS	Color	sub.	3 (3)	Gamut mapping	(A)	Used for gamut mapping optimization	2.3.5

Continued on Next Page. . .

Table 1 – Continued

Year	Metric	Author(s)	Type	HVS	MS	S/NS	C/G	Test	Scenes	Modification	Observers	Comment	Section
2005		Xu et al.	ID	No	No	NS	Color	sub.	-	Compression	(A)		2.9.34
2005	NwMSE	Samet et al.	IQ	Yes	No	FFS	Gray	sub.	-	JPEG2k, JPEG, and blurring	Scenes from LIVE		2.1.7
2006	VIF	Sheikh and Bovik	IF	Yes	No	FFS	Gray	sub.	29 (344)	JPEG and JPEG2k	-	Scenes from LIVE	2.9.9
2006	WCWSSIM	Brooks and Pappas	VQ	Yes	Yes	FFS	Color	sub.	3 (5)	Video compression and transmission distortion	-	Various testing of the metric.	2.4.12
2006	DTWT-SSIM	Lee et al.	IS	No	No	FFS	Gray	obj.	10 (4860)	Blurring, scaling ,rotation and shift	-	Tested on handwritten data as a similarity measure.	2.4.13
2006	ESSIM	Chen et al.	IQ	No	No	FFS	Gray	sub.	- (489)	JPEG2k, JPEG, and blurring	-		2.4.6
2006	GSSIM	Chen et al.	IQ	No	No	FFS	Gray	sub.	- (489)	JPEG2k, JPEG, and blurring	-		2.4.7
2006	UQI-HVS	Egiazarian et al.	IQ	Yes	No	FFS	Gray	sub.	2 (44)	Noise, blur, JPEG and JPEG2000	56		2.4.3
2006	PSNR-HVS	Egiazarian et al.	IQ	Yes	No	FFS	Gray	sub.	2 (44)	Noise, blur, JPEG and JPEG2000	56		2.1.14
2006		Mindru and Jung	IQ	Yes	No	FFS	Color	sub.	1 (3)	Halftoning	(A)		2.4.14
2006	$SSIM_{color}$	Bonnier et al.	ID	No	No	FFS	Color	sub.	15 (90)	Gamut mapping	22		2.4.8
2006	mPSNR	Munkberg et al.	IQ	No	No	NS	Color	sub.	16 (-)	HDR	(A)		2.1.15
2007	RCBM	Dong et al.	IQ	No	No	FFS	Gray	obj.	29 (204)	JPEG		Scenes from LIVE	2.4.10
2007	SEME	Silva et al.	IS	No	No	FFS	Gray	sub.	(233)	JPEG	-	Scenes from LIVE	2.4.17
2007	CISM	Lee et al.	HT	Yes	No	FFS	Color	obj.	1 (28)	Halftoning	-		2.4.15

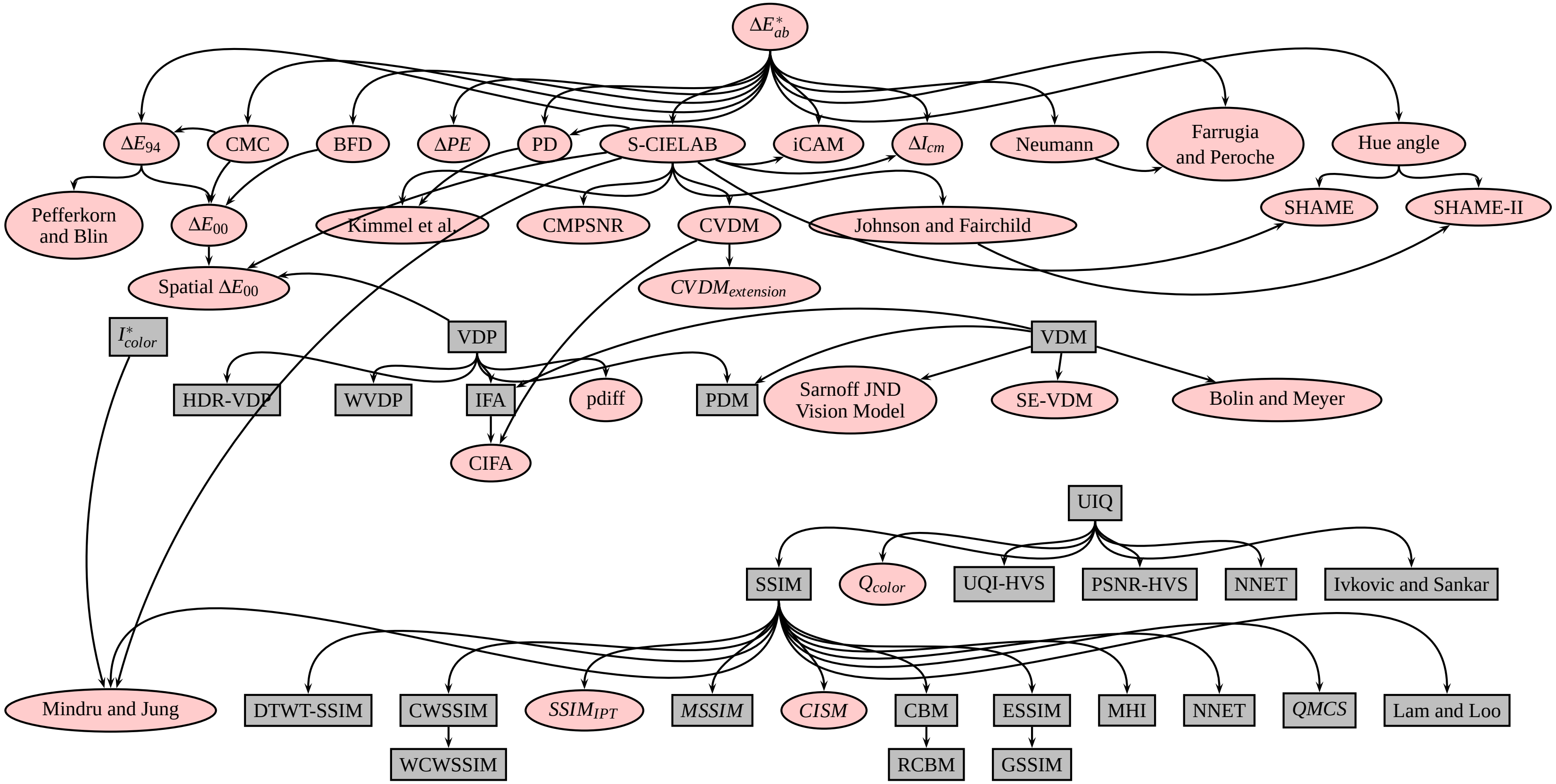
Continued on Next Page…

Table 1 – Continued

Year	Metric	Author(s)	Type	HVS	MS	S/NS	C/G	Test	Scenes	Modification	Observers	Comment	Section
2007	P-CIELAB (ΔPE)	Chou and Liu	IF	No	No	VFS	Color	sub.	3 (6)	JND profile and JPEG	(A)		2.3.18
2007	*QMCS*	Yao et al.	IQ	No	No	FFS	Gray	sub.	29 (344)	JPEG and JPEG2k	-	Scenes from LIVE	2.8.4
2007	$D\acute{E}COR-WSNR$	Wan et al.	HT	Yes	No	FFS	Gray	obj.	1 (3)	Error diffusion	-		2.1.13
2008	DP	Granger	CD	No	No	NS	Color	obj.	-				2.2.6
2008		Lam and Loo	IQ	No	No	NS	Gray	sub.	2 (6)	Noise			2.4.16
2008	SBLC	Gorley and Holliman	IQ	No	No	NS	Gray	sub.	3 (54)	JPEG compression	20		2.9.36
2008	”busyness”	Orfanidou et al.	IQ	No	No	FFS	Gray	sub.	10 (80)	JPEG and JPEG2k compression	10	Psychophysical data from Allen et al. [2004]	2.9.35
2008	Spatial ΔE_{00}	Chen et al.	IQ	Yes	No	FFS	Color	obj.	1 (1)	Blurring	-		2.3.3
2009	SHAME	Pedersen and Hardeberg; Pedersen and Hardeberg	IQ	Yes	No	FFS	Color	Sub	-	-	-	Various testing	2.3.16
2009	SHAME-II	Pedersen and Hardeberg	IQ	Yes	No	FFS	Color	Sub	-	-	-	Various testing	2.3.17

B Metric map

This map shows the connections between the different metrics available. Only the most common metrics are shown, i.e. those with connections to other metrics. Connections between them can be that they are influenced, directly descender or have adapted one or several modules from the parent metric. Other connections can also be found, but they are not necessarily shown here. A gray square indicates a metric for grayscale images, while a red oval indicates a metric for color images.


ΔE^*_{ab}
ΔE_{94}
CMC
BFD
ΔPE
PD
S-CIELAB
iCAM
ΔI_{cm}
Neumann
Farrugia and Peroche
Hue angle
Pefferkorn and Blin
ΔE_{00}
Kimmel et al.
CMPSNR
CVDM
Johnson and Fairchild
SHAME
SHAME-II
Spatial ΔE_{00}
$CVDM_{extension}$
I^*_{color}
VDP
VDM
HDR-VDP
WVDP
IFA
pdiff
PDM
Sarnoff JND Vision Model
SE-VDM
Bolin and Meyer
CIFA
UIQ
SSIM
Q_{color}
UQI-HVS
PSNR-HVS
NNET
Ivkovic and Sankar
Mindru and Jung
DTWT-SSIM
CWSSIM
$SSIM_{IPT}$
MSSIM
CISM
CBM
ESSIM
MHI
NNET
QMCS
Lam and Loo
WCWSSIM
RCBM
GSSIM