The Performance of Image Difference Metrics for Rendered HDR Images

Jørn Skjerven

Master's Thesis
Master of Science in Media Technology
30 ECTS
Department of Computer Science and Media Technology
Gjøvik University College, 2011

Avdeling for
informatikk og medieteknikk
Høgskolen i Gjøvik
Postboks 191
2802 Gjøvik

Department of Computer Science
and Media Technology
Gjøvik University College
Box 191
N-2802 Gjøvik
Norway

The Performance of Image Difference Metrics for Rendered HDR Images

Jørn Skjerven

1st July 2011

Contents

1 Abstract

HDR is a field in image processing that has received a lot of attention in the later years. Techniques for capturing, tone map back to viewable data has been proposed. Many different ideas have been pursuited, some with a background in the Human Visual System (HVS), but the same problem with determining the quality of these reproductions still exist. In low dynamic range imaging, the solution to the problem has been either to do a visual inspection and compare the reproduction against an original, but as this is a labour intensive and time consuming and highly subjective process, and the need for automated measures which can predict quality has resulted in different image difference metrics.

As for comparison of HDR and LDR, this is no trivial task. Currently, no method of automated comparison has been deemed a viable solution due to the difference in in dynamic range.

In this master thesis, we present a novel framework extending on recent research which enables us to compare HDR and LDR content, and from this using standard image difference metrics to evaluate the quality of these. These measures are tested against data from a perceptual experiment to verify the stability and quality of the framework. Initial results indicate that the proposed framework enables us to evaluate the quality of such reproductions on the tested scenes, but that some problems are still unsolved.

Keywords: High dynamic range, perceptual uniform encoding, tone mapping operators, image difference metrics, perceptual experiment.

2 Acknowledgements

This master thesis is the ending of my three year study at the Master in Media Technology in the Faculty of Computer Science and Media Technology at Gjøvik University College.

I would like to thank and show my deepest appreciation for my excellent supervisor Ivar Farup for his kind support, motivation and help during the last half year when this thesis is written. Through the many visits to his office and small talk the work grew forward and ended up in this report.

I would also like to thank professor Jon Yngve Hardeberg for his excellent class *Advanced Color Imaging* during the first half of the second semester where my interest for color and color science really came to life.

I would also like to express my deepest gratitude to the staff of The Norwegian Color Research Laboratory that supported me and given good feedback during my Master Thesis, especially Marius Pedersen, Peter Nussbaum, Arne Magnus Bakke, Gabriele Simone, Dibakar Pant and Raju Shrestha. Without them, I would not have gotten where I am today.

I would also like to thank my family and girlfriend for her kind support during long days of work and late nights with work.

Last but not least, I offer my regards to all those who have taken part of my experiment or supported me in any respect during the work with this Master Thesis

Jørn Skjerven
July, 2011

3 Introduction

Since the beginning of man, pictures of different scenes has been depicted on different media. One thing that always has been an important, is how we perceive the motive. One of the essential parts of this is the dynamic range and how this is mapped. On most media it is not possible to reproduce a perfect copy, so different techniques has been developed to produce a more realistic rendering. In paintings, Leonardo da Vinci is accredited the technique Chiaroscuro, which emphasizes the difference between light and dark to create an illusion of high dynamic range. This was further adapted into photography and film, with Ansel Adams and his zone system to transfer natural light into specific values on negatives. Today the same problems exist with digital cameras. Systems cannot capture the same range that the eye can, and a lot of adaptation processes that are not fully understood influences how we perceive motives.

3.1 Areas of Research

The thesis touches a wide area of research. The main work lies in the field of High Dynamic Range(HDR), the capturing, rendering and displaying of such data on Low Dynamic Range(LDR) environments, and how these to can be compared using known measures of quality in the form of image difference metrics.

3.2 Problem Description

Today the cameras get more and more elaborate, but still there are cases where the equipment cannot give a dataset which produces an accurate reproduction of a scene. By introducing HDR, we can extend the dynamic range that is available in the raw data, and in post processing, we can choose the different elements that are to be included. HDR imaging can also be used to emphasize other aspects of images and give hyperreal reproduction which seems more real than reality itself. This makes it very hard to evaluate which reproductions are good ones, and earlier research has mainly relied on subjective measures to evaluate quality of the reproductions against the rendering intent.

3.3 Justification

In this thesis the hypothesis that by using our framework, quality metrics developed for LDR images can be used to compare HDR against LDR reproductions are proposed. A direct comparison is not possible due to the difference in dynamic range. Some research in the field has been done, and by using the state of the art techniques we will extend this to a wider set of metrics and image attributes.

3.4 Research Question

From the problem described above, we will look into three main topics: The quality of different tone mapping operators, the quality of image difference metrics when comparing HDR vs LDR reproductions and usage of original scenes as a reference in perceptual experiments. From these topics, we will try to answer the following questions:

1. How does the new tone mapper STRESS by Kolås compare to a set of state of the art tone mappers?
2. What image attributes are important in a HDR image, and how are they weighted against each other ?
3. Is it possible to compare a HDR original against a LDR reproduction using a set of image difference metrics, and how well can they predict the quality of such a reproduction?
4. How well are the different image difference metrics settings able to determine the image quality of a HDR vs LDR compared to each other?
5. Is using a original scene a viable option as a reference when comparing a LDR reproduction against a HDR captured of the scene?

3.5 Research Methodology

To try to answer our research question, we are going to look into the whole HDR pipeline, how HDR images are captured and processed back to viewable data on a standard LDR monitor. Earlier research have shown that directly comparison of HDR and LDR content using conventional methods does not yield probable results. We will therefore present a processing framework how they can be compared based on research, and calculate a set of different scores that represent the different properties of image quality by the usage of image quality metrics.

To check the findings, we will build two different scenes with different viewing conditions which we will perform a perceptual experiment on. The same set of attributes are checked for. In addition we will use our framework to produce comparable image data which results in a set of quality scores.

To compare the results from the perceptual experiment with the quality scores, a correlation of the results will be done, and from this data a conclusion can be drawn.

3.6 Thesis Organization

The thesis is divided into four main parts:

- State of the art
- Experimental setup
- Result
- Conclusion

In the first part, we present the State of the art in the different field of HDR image capture and tone mapping. We also present a set of image quality measures. In addition we present the components which our framework is built upon.

The second part is used to describe how the different parts of the experiment is carried out, and the different setting/parameters used to generate images and quality scores.

The results part will consist of the gathered results, where we present the results from the experiment. From these we can use statistical methods, among them correlation to check for dependance between the different datasets.

For the last chapter we will try to conclude if the findings will fulfill the research question, and if the results are statistically viable. A discussion of the results and a set of further research questions is performed.

4 State of the Art

To capture images of different scenes has been done since the beginning of man, starting with crude paintings in caves to todays digital photography. Different techniques have been done to try to capture the sensation of the scene during the age, and especially painters such as Leonardo da Vinci and Rembrandt whose pictures give a very good representation of the full dynamic range that is present in the motive with the limited range that the paints and canvas has.

Transferred into photography, tone mapping was done in still another ways, and systems were developed to increase the chance to get a good exposure of a difficult scene. In todays digital world, much of this has been incorporated into the logic circuits of available cameras, and in most cases the default settings are capable of give a very good reproduction of the motive.

As cameras continue do develop, but measurements has stuck since the start of photography. One such value is ΔEV, or exposure value, which is a number that represent the equivalent shutter speeds and aperture combinations for a given scene. The best DSLR cameras today are capable to capturing a dynamic range of ΔEV [9,12], but for some scenes like the ones built in this experiment, this is still not enough. To overcome this limitation, the usage of HDR imaging can make us able to capture the entire dynamic range and make reproductions which corresponds better to the original than a single capture can give.

This then needs to be mapped down to a range which is displayable by a monitor by using different tone mapping operators(TMOs), which preserves the different parts of the image which is considered important.

The last step that has gotten a lot of attention, is the usage of automated methods to measure the quality of images. These image difference metrics (IDM) are mainly been developed for a standard LDR images, and the reproductions of HDR scenes have mainly been given a subjective representation of the quality. Some research have been developed, but no distinct measures that have been proven to give simple measures on comparing HDR and LDR content has been proposed.

4.1 Image Assembly

There are very few devices that can capture HDR images directly, and those that exist are very expensive and not generally available. In the past, most HDR material were manually rendered using 3D computer graphics.

Another way that has been developed to generate HDR images, or radiance maps, is by using a standard camera, and taking multiple exposures of the same scene with each image at a different exposure time so that every pixel in the scene will be properly exposed in one of the shots. However, in some of the shots other pixels will be over- or under exposed, and therefore it is possible and desirable to ignore very bright and dark pixels.

Assuming that the capturing device is perfectly linear, each exposure can be transferred into the same domain by dividing each pixel by the image's exposure time. The corresponding pixels can then be averaged across exposures, excluding the pixels mentioned above.

4.1.1 Camera Response Function

As cameras are not perfectly linear, it becomes important to account for the camera's response. Two methods to take this into account has been proposed. Debevec and Malik [1] demonstrated a simple and robust way to derive the camera response curve from a set of exposures, extending earlier work by Mann and Picard [2]. The idea is that by capturing different exposures of a static scene, they are effectively sampling the camera response function at each pixel. Debevec and Malik uses linear optimizations to find a smooth curve that minimizes the mean square error over the derived response function.

Another technique was presented by Mitsunaga and Nayar [3] where they approach the problem of deriving the response function with the use of polynomial approximation. The main advantage is that their method will resolve the exact response ratios in addition to the camera response function. This is an important addition because in low-grade consumer equipment aperture and shutter speeds may not be included in the output from the camera.

Mitsunaga and Nayar define the following N-dimensional polynomial for their camera response function:

$$f(M) = \sum_{n=0}^{N} c_n M^n \tag{4.1}$$

where any response can be modeled by the calibration that is determining as the order of N and the coefficients c_n

The final response function is defined by the $N+1$ coefficients of this polynomial, $c_0, \ldots, c_n$. To determine these coefficients, they minimize the following error function for a given candidate exposure ratio, $R_{q,q+1}$ (the scale ratio between exposure q and $q+1$):

$$\epsilon = \sum_{q=1}^{Q-1} \sum_{p=1}^{P} \left[\sum_{n=0}^{N} c_n M_{p,q}^n - R_{q,q+1} \sum_{n=0}^{N} c_n M_{p,q+1}^n \right]^2 \tag{4.2}$$

The minimum is found by determining where the partial derivates with the respect to the polynomial coefficients are all zero, which is the same as solving the following system of $N+1$ linear equations:

$$\frac{\partial \epsilon}{\partial c_n} = 0 \tag{4.3}$$

Following the method by Devlin and Malik, they only solve the response up to some arbitrary scaling. By defining $f(1) = 1$, the dimensionality of the system is reduced by one coefficient by substituting

$$c_N = 1 - \sum_{n=0}^{N-1} c_n \tag{4.4}$$

which can be written as a $N \times N$ system. The reader is pointed to Reinhard et al. [4, chapter 5] for a full explanation of the transformation.

The original Mitsunaga and Nayar formulation only considered adjacent exposures. In practice, the system is more stable if all exposure combinations are considered. The error function becomes

a triple sum by including a sum over $q' \neq q$ instead of just comparing q to $q+1$. This results gets repeated in the sums of the combined system of equations, where $d_{q,p,n}$ is replaced by

$$d_{p,q,q',n} = M^n_{p,q} - R_{q,q'} M^n_{p,q'} \tag{4.5}$$

The actual exposure ratio between the images, Mitsunaga and Nayar apply a interactive technique where the system of equations is solved repeatedly, and between each solution we update the exposure ratios using

$$R^{(k)}_{q,q+1} = \sum_{p=1}^{P} \frac{\sum_{n=0}^{N} c^{(k)}_n M^n_{p,q}}{\sum_{n=0}^{N} c^{(k)}_n M^n_{p,q+1}} \tag{4.6}$$

Iterations will complete when the polynomial is no longer changing significantly:

$$| f^{-1}_{(k)}(M) - f^{-1}_{(k-1)} | < \epsilon, \forall M \tag{4.7}$$

The only problem left is to determine the polynomial degree N. Mitsunaga and Nayar recommend solving for every polynomial up to some maximum exponent, and accepting the solution with the smallest error, ϵ. Test have shown that the solution process proceeds quickly, so a solution is found quite fast.

Mitsunga and Nayar recommend to ensure that the same degree is selected for each color channel, so a combined ϵ is preferred in the final test.

The different methods for camera response curves requires a set of image samples from the exposure sequence. This may either be a intelligent gathered set of samples or it is possible to use every pixel from every image, but this will both increases computation time and reduce stability due to misaligned and noisy data. To get a good sampling, the procedure in Reinhard et al. [4, Chapter 5.7.3] results in a set which should minimize both exposure misalignment and noise from the image sensor.

4.1.2 Ghost Removal

A problem with multiple exposures is movement of objects in the scene. These may appear as "ghosts" in the combined result due to their multiple locations. A lot of different algorithms and methods try to address how we can remove these unwanted elements, but we will only present some of them here.

Kang et al. [5] is one of the later who addresses the problem of alignment by warping pixels according to local content. The problem with this method, is that even if removal of ghosts can be done correctly for objects that change postures as well as positions, it will still leave the problem of filling the holes that were obstructed in some views, but not in others. The method can also produce some artifacts due to low sampling.

Khan et al. [6] later presented an interactive solution to the ghosting problem, where a weighting function is introduced

$$P(x|F) = M^{-1} \sum_{m=1}^{M} K_H(x - y_m) \tag{4.8}$$

where x is a vector that belongs to class F, H is a symmetric, positive definite $d \times d$ bandwidth matrix, M is the numbers of vectors in the class and y_m is the m^{th} vector in the class. Furthermore

$$K_H(x) = |H|^{-\frac{1}{2}} K(H^{-\frac{1}{2}} x) \tag{4.9}$$

where K is a d-variate kernel function and H is a symmetric, positive definite $d \times d$ matrix. To represent K, a Gaussian density is used. Each pixel is represented by a vector in a feature space, three dimensions represent color and two represent the location of the pixel in the image, $x_{ijr} \in \mathbb{R}^5, i = 1, 2 \ldots I, j = 1, 2, \ldots J, r = 1, 2, \ldots R$

For this vector, the background is represented as $p \times q \times s$, which assumes that all vectors are part of the background.

Weights are assigned to each vector based on a simple hat-function, and three weights for each pixel are derived. The final weight, ω_{pqs} is a average of these, and using this weight the probability that a vector x_{ijr} is an actual part of the background is calculated. Thus

$$P(x_{ijr}|F) = \frac{\sum_{p,q,s \in N(x_{ijr})} \omega_{pqs} K_H(x_{ijr} - y_{pqs})}{\sum_{p,q,s \in N(x_{ijr})} \omega_{pqs}} \tag{4.10}$$

Vectors that capture moving objects will get a lower probability than a vector that is a part of the of the background. From this a HDR image can be generated using the chosen weights and this will result in a diminishing of the ghosts compared to a image using only the initial weights.

The process of kernel density estimation is repeated, with the initial weight of the vectors y_{pqs} as the weights from the previous iteration as follows:

$$\omega_{pqs,t+1} = \omega(Z_s(p,q)) * P(x_{pqs}|F) \tag{4.11}$$

where $\omega(Z_s(p,q)$ is the initial weight for the pixel located at (p,q) with exposure s, $P(x_{pqs|F})$ is the weight of the pixel as determined by kernel density estimation and $\omega_{pqs,t+1}$ is the weight that will be used as the kernel density estimation in iteration $t+1$.

By multiplying the newly computed weights with the initial weights, the distribution around a pixel which consists of over- and under-exposed pixels will not get high weights as their initial weights are low.

This results in a image that removes moving objects from a set of captured images and gives a plausible image.

4.1.3 Image Storage

For storage of HDR data, there has been proposed a wide range of different formats and encodings. The Radiance RGBE developed by G.W. Larson [7] was the first encoding, and with the HDR format as a container, it has found a widespread use in the graphics community. In the later years, there has been proposed many different encodings and formats. In this thesis, we will only look at one of these formats, OpenEXR.

OpenEXR [8], or "EXtended Range format", is one of the more popular during the later years and was developed by Industrial Light and Magic. This format stores the data in four different channels R,G,B and A which is used for alpha(opacity) values. The format stores 48 bits pr pixel with a

dynamic range of $1 : 10^{10.7}$ which is considered enough for all practical usages. In addition the format covers the entire visible gamut, something which RGBE does not.

Another interesting point with the OpenEXR format is the support for both HALF, FLOAT and UINT values in each channels, where the HALF data format is directly compatible with NVIDIA Cg and later CUDA programming language. This makes it very easy to implement into most programming languages and processing can be sped up using graphic cards for image processing tasks that can be processed in parallel.

The formula for converting from an encoded HALF value is

$$H = \begin{cases} (-1)^S 2^{(E-15)} \left(1 + \dfrac{M}{1024}\right) & 1 \leq E \leq 30 \\ (-1)^S 2^{-14} \dfrac{M}{1024} & \text{E} = 30 \end{cases} \tag{4.12}$$

where S is the sign bit, E is the expontent (0-31). If the exponent is 31, the value is either infinity if $M = 0$, or not a number (NaN). Zero is represented by zero bits.

For writing and reading the OpenEXR data, a API is included in the IlmImf library available in a C++ implementation. Several lossless compression options is included, but most widely used is the PIZ wavelet compression, which gives a compression ratio of about 3:5 (60 %).

This library was used in a pre-project to this master thesis to develop reader and writer functions for Matlab to import and export data to and from OpenEXR files.

4.2 Tone Mapping Operators

Displays that shows images should show a faithful representation of the scenes depicted in the image. This is not a new problem, and artist and photographers have been struggling with this since the birth of painting and photography. The main problem, is that light intensity levels and contrast in the environment cannot be reproduced on the display medium, be this film, paint canvas or digital panels.

Thumblin and Rushmeyer [9] formally described this problem, and introduced the usage of visual models to solve it. The model is based on the idea that luminance are calculate from the original scene, and by usage of a forward model we can calculate a set of intermediate values. By reversing these, we get at tone mapped image, and the ultimate goal is to have perceptual match between the output between the input and the output.

This quickly become one of the main way of thinking for producing the best perceptual match between the original an a tone-mapped reproduction. Most of the work done has been applied to compressing the luminance values to make the dynamic range of an image to fit the range of a given media.

The visual pathway has a signal-to-noise ratio of less than 2 orders of magnitude for each channel [4, chapter 7], but we are still able to operate in a wide order of illumination, and gives the simultaneously perceive the detailed contrast in both light and dark parts of a HDR scene. This makes it very difficult to map a image data into the narrow range that is available for different output units.

Simple functions such as scaling and compression was early shown not to give good reproductions, and this inspired researchers to look at how the Human Visual System (HVS) works. The different aspects of the HVS is often mimicked to some extent by most TMOs, and a wide variety of

Table 1: Table of all presented tone Mappers

Name	Type	Publication
Miller Brightness-Ratio-Preserving Operator	Global	[10]
Tumblin-Rushmeyer Brightness-preserving Operator	Global	[9]
Ward Contrast-based Scale Factor	Global	[11]
Logarithmic and Exponential Mappings	Global	[4]
Reinhard and Devling Photoreceptor Model	Local	[12]
Chiu Spatially Variant Operator	Local	[13]
Reinhard Photograchic Tone Reproduction	Local	[14]
Durand Fast Bilateral Filtering	Local	[15]
STRESS	Local	[16]
Rahman Retinex	Local	[17]
Fattal Gradient Domain	Local	[18]
Fairchild iCAM	Local	[19]
Pattanaik Multiscale Observer Model	Local	[20]
Mantiuk Perceptual Framework for Contrast Processing	Local	[21]

operators have been developed.

There has been a presented a lot of different tone mapping operators, and this is not a full or complete list of them. A small selection of the operators either used in our experiment or is related either as previous works or are of historical interest.

It is common practice to classify tone mapping operators into two main groups:

- Global
- Local

A global operator uses the entire image as a neighborhood for a image, and therefore the compression curve will be the same for all pixels. As a result, global operators are frequently less computationally less expensive than local operators.

The other type is the local operator. This operator is designed to compute local adaptation of each pixel based on the pixel value itself as well as the neighborhood of surrounding pixels that are of interest. This local adaptation then drives the adaptation curve for the given pixel, and due to the fact that the neighborhood of the pixel helps determine how the pixel should be compressed, a bright pixel with dark pixels surrounding it will be treated differently than a bright pixel with a bright neighborhood.

A more thorough explanation of the different algorithms is given in the next section where we present a selection of TMO algorithms and a list of them are found in Table 1.

4.2.1 Global Tone Mapping Operators

Miller Brightness-Ratio-Preserving Operator

This is the first tone-reproduction operator in the field of digital image processing. It was documented in 1984 by Miller and colleagues [10] which tried to introduce the computer graphics to the lighting engineering community. In this paper a discussion of the usefulness of an algorithm in lighting design, and that output should be given in radiometric or photometric quantities rather than scaled pixel intensities. This results in images that are not directly displayable on LDR displays.

As a result, a global TMO that aims to preserve the sensation of brightness both before and after the dynamic range reduction was developed. They defined brightness as follows:

$$Q = kL_v^b \tag{4.13}$$

Q is approximated as a power function of luminance L_v. Miller et al. assert that an image may look the same both before and after dynamic range reduction by making a model where the brightness ratio is kept constant. We therefore can write the two visual elements Q_1 and Q_2 to be equivalent with the compressed counterpart $Q_1^{'}$ and $Q_2^{'}$ when the ratios are constant. This may be written as

$$\frac{Q_1}{Q_2} = \frac{Q_1^{'}}{Q_2^{'}} \tag{4.14}$$

It should be noted that visual equivalence between pairs of brightness values is not the same, in other words $Q_1 \neq Q_1^{'}$.

The procedure starts by converting image data to brightness values $Q_w(x,y)$. The the maximum brightness, $Q_{w,max}$ is determined and the data is normalize by dividing all values on the maximum.

We then determine the display device's maximum brightness, $Q_{d,max}$ based on the maximum luminance value using the same luminance brightness relationship. Display brightness is then determined using the following:

$$Q_d(x,y) = \frac{Q_w(x,y)}{Q_{w,max}} Q_{d,max} \tag{4.15}$$

These brightnesses are then converted to luminances by applying the inverse of the brightness function. By using Stevens work on brightness [22], Miller et al. argues that produced the most plausible results. The relationship between luminance and brightness was then calculated to the following:

$$Q = (-1,5log_{10}(L_v) + 6,1)L_v^{0,338L_v^{0,034}} \tag{4.16}$$

As Miller et al.'s work aimed for lighting design, their operator is suitable for compressing luminance ranges that are typically found indoors. This assert that actual room luminance is between 100 and 1000 cd/m^2. For practical purposes, each image is normalized between 0 and 1000 cd/m^2, so it is not unreasonable to assume that the operator is not suitable for images with a higher dynamic range. This makes this operator mainly of interest for historical reasons.

Tumblin-Rushmeyer Brightness-preserving Operator

As Miller et al. was the first ones to introduce computer graphics to the field of lighting design, Tublin and Rushmeyer [9] were the ones that introduced the problem of tone reproduction to the field of computer graphics in 1993.

Tumblin-Rushmeyer also based their work on Stevens' psychophysical data, based on the idea that the HVS already had solved the problem of dynamic range reduction. Their work resulted in the Tumblin-Rushmeyer Operator which exists in two different forms, the original and a revised version. The reason for giving out a revised version, was that the original was calibrated using sieverts, which is not widely used in the imaging community.

As it is based on the same data as the Miller's Operator, the brightness function is slightly altered:

$$Q(x,y) = C_0 \left(\frac{L(x,y)}{L_a} \right)^{\gamma} \tag{4.17}$$

Q stands here for the brightness(or perceived brightness) measured in brils. L is the luminance and L_a is the adaptation luminance, both measured in cd/m^2. The constant $C_0 = 0.3698$ is introduced to allow the formulae to be stated using SI units, and γ is the measure for contrast sensitivity and is in itself a measure for adaptation luminamce (L_a). The function may be evaluated both for a HDR image as well as for the intended display device, which two sets of brightness values. These two then are functions of the input luminances, world luminance (measured or derived from HDR image, with subscript w) and display luminance (with subscript d). In contradiction with Miller et al. they do not match image and display brightness ratios, these are simply equated as follows:

$$\begin{aligned} Q_w(x,y) &= C_0 \left(\frac{L_w(x,y)}{L_{wa}} \right)^{\gamma(L_{wa})} \\ Q_d(x,y) &= C_0 \left(\frac{L_d(x,y)}{L_{da}} \right)^{\gamma(L_{da})} \\ Q_w(x,y) &= Q_d(x,y) \end{aligned} \tag{4.18}$$

The gamma function $\gamma(L)$ models Stevens' human contrast sensitivity for the image and the display as follows:

$$\gamma(L) = \begin{cases} 2.655 & \text{for L} > 100\ cd/cm^2 \\ 1.855 + 0.4 log_{10}(L + 2.3 * 10^{-5}) & \text{else} \end{cases} \tag{4.19}$$

As the world or image luminance, the equation for $L_d(x,y)$ is solved for the display luminance:

$$L_d(x,y) = L_{da} \left(\frac{L_w(x,y)}{L_{wa}} \right)^{\frac{\gamma(L_{wa})}{\gamma(L_{da})}} \tag{4.20}$$

The adaptation luminance L_{da} for the display and L_{wa} for the image. Display adaptation luminance is often between 30 and 100 cd/m^2, but for HDR display devices this may be higher.

For dim scenes where L_{da} map to mid-range display luminances, will result in a gray uniform appearance in the display. The way of remedy this is to introduce a scale factor $m(L_{wa})$ which is

dependent on the world adaptation level, L_{da}:

$$m(L_{wa}) = (\sqrt{C_{max}})^{\gamma_{wd}-1} \tag{4.21}$$

which results in

$$\gamma_{wd} = \frac{\gamma_w}{1.855 + 0.4log(L_{da})} \tag{4.22}$$

In the equation above, C_{max} is the maximum displayable contrast which is typically between 30-100 for LDR displays, but may be quite a bit higher on new LCD screens. Screens with up to 400 cd/m^2 is readily available.

The full operator can then be written as follows:

$$L_d(x,y) = m(L_{wa})L_{da}(\frac{L_w(x,y)}{L_{wa}})^{\frac{\gamma(L_{wa})}{\gamma(L_{da})}} \tag{4.23}$$

It is worth noticing that this operator is calibrated in the SI unit cd/m^2, and the input image must also be specified in the same units. The normal approach with normalizing, applying gamma correction and then multiplying with 255 to make into standard RGB values will not work because the operator already includes a display gamma correction step.

To do the mapping from other units, only a pre-scaling by a factor determined by trial and error is available, which makes it very hard to implement in general.

Ward Contrast-based Scale Factor

This operator aims to preserve the contrasts in a reproduction. Gregory Ward published a tone mapping operator in 1994 [11] that matches just noticeable differences,(JNDs), one might see in a image with ones an observer viewing a LDR display device may distinguish. In this way, differences are preserved without giving the limited number of display steps to differences we cannot see.

The operator maps image (or world) luminances L_d to display luminances L_w linearly:

$$L_d(x,y) = mL_w(x,y) \tag{4.24}$$

Based on the studies by Blackwell [23], the equation for the scale factor m can be solved as follows:

$$m = \frac{1}{L_{d,max}} \left(\frac{1.219 + (\frac{L_{d,max}}{2})^{0,4}}{1.219 + L_{wa}^{0,4}} \right)^{2.5} \tag{4.25}$$

$L_{d,max}$ is here specified as the displays maximal luminance, and the display adaptation level is estimated to half of this value, and should be specified by the user. The world adaptation level may be estimated as the log average of the image:

$$L_{wa} = \exp\left(\frac{1}{N} \sum_{x,y} \log(10^{-8} + L_{x,y}) \right) \tag{4.26}$$

This equation sum up all log luminance values. Black pixels will result in a singularity, so a small offset is added to all pixels. This offset could be omitted and only non-zero pixels could be included, but the difference would be so small that both ways are treated as equal.

Worth notice is also the fact that this operator also needs the input in SI units, so the same scaling as the Tumblin-Rushmeyer operator must be done. This pre-scaling of the image will result in a brighter result.

Logarithmic and Exponential Mappings

These are maybe the easiest of the nonlinear mappings that exist. The main purpose of these is to provide a baseline result which all other operators may be compared against.

The logarithm is a compressive function for values larger than 1, and therefore a mapping of luminances must be done. One way of doing this is by using the following function:

$$L_d(x,y) = \frac{\log_{10}(1+L_w(x,y))}{\log_{10}(1+L_{max})} \tag{4.27}$$

There is also a second way that maps world luminances to display luminances by means of the exponential function:

$$L_d(x,y) = 1 - \exp(-\frac{L_w(x,y)}{L_{av}}) \tag{4.28}$$

Here the function is bound between 0 for black pixels and 1 for infinitely bright pixels which will never occur. The division by the average luminance L_{av} causes pixels to be mapped to $1-1/\ln \approx 0.63$. Because this average is slightly larger than 0.5, the arithmetic average if often used instead of log average luminance.

The images produced from logarithmic mapping is considered to be somewhat dull, and exponential mapping produces images that overall are much brighter. Different types of lock-down values like L_{max} and L_{av} will have a great impact on the images as these values map the brightest pixels.

4.2.2 Local Tone Mapping Operators

Chiu Spatially Variant Operator

The first to observe that a spatially varying operator may be useful was Chiu et al. [13] They found that artists often use spatially varying techniques to fool the eye into believing that a much larger dynamic range is present in a scene than the range that actually exist, such as Chiaroscuro.

The basic formula for their operator multiplies each pixel's luminance by a scaling factor, $s(x,y)$ which is dependent on the pixel itself and its neighbors:

$$L_d(x,y) = s(x,y)L_w(x,y) \tag{4.29}$$

Chiu et al. presents a solution using a low-pass filter on the input image to represent a local average, and that most low-pass filters would produce similar results. Reinhard et al. [24] does this by using a Gaussian filter as follows:

$$L_d(x,y) = \frac{1}{kL_w^{blur}(x,y)}L_w(x,y) \tag{4.30}$$

Here k is a constant that controls the weight given to the blurred image relative to the unblurred input. The problem with this operator is that small kernels produce heavy haloing effect. The work with this TMO is intended to be exploratory in the field of local operators, and to highlight the effects and main problems with these.

Reinhard and Devling Photoreceptor Model

The model presented by Reinhard and Devlin [12] is directly inspired by photoreceptor physiology. It is based on the idea that some of the parts of the HVS responds logarithmically on some of the operation range. The HVS responds more like a sigmoidal functions, and these have been mapped by Naka and Rushton [25]. For the purpose of tone reproduction, Hood et al. [26] gave the following representation:

$$V_{x,y} = \frac{I(x,y)}{I(x,y) + \sigma(I_a(x,y))} \tag{4.31}$$

I is here the photoreceptor input, V the photoreceptor response and σ the semisaturation constant. This constant can be calculated from the adaptation value I_a as follows:

$$\sigma(I_a(x,y)) = (f\, I_a(x,y))^m \tag{4.32}$$

Here f is a scale factor to steer overall luminance, and can initially be estimated to 1. m is used to steer the overall impression of contrast, and is dependent on the minimum, maximum and average luminance, and can be derived from the image by the following formula:

$$\begin{aligned} m &= 0.3 + 0.7k^{1.4} \\ k &= \frac{L_{max} - L_{av}}{L_{max} - L_{min}} \end{aligned} \tag{4.33}$$

k here can be interpreted here as the key of the image(average dark-/lightness), I_a is the adaptation level, and may be computed in traditional way as average luminances of the images(log units), but more elaborate computations of I_a may be in place to achieve better light and chromatic adaptation.

Strong color casts may also be removed by interpolation between the luminance values $L(x,y)$ and the RGB values for each pixel. The formula

$$I_a(x,y) = cI_{\mathrm{r|g|b}}(x,y) + (1-c)L(x,y) \tag{4.34}$$

produces the same effect as the von Kries color correction if we set $c = 1$, and no color correction if $c = 0$. This is referred to as "chromatic adaptation".

The next step is the adaptation level which can be thought of as determined by the current light level a receptor is exposed to, as well as the levels the receptor was exposed to in the past. Because of saccadic eye movements and other connectivity in the retina, it can be assumed that the current adaptation level is a function of the pixel value itself and all other pixels in the image.

This may be solved using a simple interpolation between pixel values and global averages:

$$I_a(x,y) = aI_{\mathrm{r|g|b}}(x,y) + 1 - a)I^{av}_{r|g|b} \tag{4.35}$$

Here a is the interpolation weight and controls image appearance which is is connected to light adaptation.

Light and chromatic adaptation may be combined in a bilinear interpolation as follows:

$$
\begin{aligned}
I_a^{local}(x,y) &= cI_{\mathrm{r|g|b}}(x,y) + (1-c)L(x,y) \\
I_a^{global}(x,y) &= cI_{\mathrm{r|g|b}}(av) + (1-c)L^{av} \\
I_a(x,y) &= aI_a^{local}(x,y) + (1-a)I_a^{global}
\end{aligned} \tag{4.36}
$$

Using the default parameters, it produces plausible results for a large class of images. Because all the parameters has an intuitive effect on the result, adjustment of any of them is fast and straightforward.

Durand Fast Bilateral Filtering

The tone operator proposed by Durand and Dorsey [15] is based on a decomposition of the input image into a base layer consisting of large-scale features and a detail layer with smaller details. Only the base layer is contrast reduced, and therefore the details are preserved. This can result in halo-effects, and there Durand and Dorsey introduce a fast and robust edge-preserving filter to counteract this.

This edge-preserving filter blurs small variations of a signal(details and noise in a image), but in this case this information is important and must be preserved. This is done by using a bilateral filter based on the work on the work of Tomasi and Manduchi [27]. This is rather a computationally expensive function, so Durand and Dorsey proposes that by splitting the density differences into a number of segments and afterward recombining them yields an approximate solution which in practice is indistinguishable from using the accurate spatial processing.

The computation is given by

$$
\begin{aligned}
D_j^{smooth}(x,y) &= \frac{1}{k_j(x,y)} \sum_u \sum_v b_j(x,y,u,v) D(x-u, y-v) \\
k_j(x,y) &= \sum_u \sum_v b_j(x,y,u,v) \\
b_j(x,y,u,v) &= f\left(\sqrt{(x-u)^2 + (y-v)^2}\right) g\left(D(x-u, y-v) - D_j\right)
\end{aligned} \tag{4.37}
$$

Here $k_j(x,y)$ is the weight factor for normalizing the result, $b_j(x,y)$ is the bilateral filter with components f and g. f is the luminance domain filter and g is the gaussian. D_j is a quantized set of possible values for pixels(x,y) and the final output is a linear combination of the two smoothed values $D_j^s mooth$ and $D_{j+1}^s mooth$. These are chosen such that D_j and D_{j+1} are the closest two values to the input density D.

The above equation may be executed in the Fourier domain for each segment j in contrast to real bilateral filtering which cannot be expressed by convolution, and from this gain a substantial speedup and still retain a very good approximation of the output from a true bilateral filter.

Durand and Dorsey argues and gives an implementation of how the tone mapping algorithm can be further optimized by introducing subsampling. As all operations, except the final interpolation aims at low-pass filtering, which can be incorporated with little loss in quality. The usage of a nearest-neighbor down-sampling scheme is proposed as this does not alter the histogram of the input image. Tests show that a subsampling factor of 4-5 results in reduced execution time, and a

downsampling factor up to 25 did not result in any noticeable artifacts.

The actual contrast reduction is as described only done on the base layer. Durand and Dorsey follow Tumblin et al. [28] and does the compression using a scale factor in the log domain. The parameter *base contrast* is as default set to 5 is shown to work well, but may be varied.

Colors calculated in the log domain and are recomposed after the contrast reduction, because the pixel differences then directly corresponds to contrast and it is a more uniform treatment of the whole range.

In the filtering phase, several influence fluctuation were considered. The Huber minimax estimator is the best when it comes to decrease the strengths of halos, but does not eliminate them. The Lorentz is also good, but only the Gaussian and the Turkey's biweight are able to correctly decompose the image. Durand et al. experimented with different scales of the spatial kernel, but this seems to have little effect on the result and is important because this allows us to keep the kernel constant. This was set to 2 % of the image size, and by default the Gaussian biweight is used.

The operator proposed by Durand is very fast, and even very large images with 10 Mpixel is processed in very short time. The details in an image is a little bit blurred, and other operators may be better, but the overall photorealistic appearance is better preserved than previous operators.

STRESS

Kolås et al [16] developed a framework called STRESS, Spatio-Temporal Retinex-inspired Envelope with Stochastic Sampling. The framework has a wide range of usages, but in this setting, only tone mapping is looked into.

The basic idea for STRESS is based on the fact that the HVS adjusts itself so that bright details in a picture may appear darker than a part that if measured is actually darker. This effect is calculated using two envelope functions, referred to as the max envelope E^{max} and the min envelope E^{min}.

These envelopes are computed in a iterative manner, where M pixel intensity values $p_j, i \in 1, ..., M$, are sampled at random with a probability $1/d$, where d is the Euclidian distance from the sampled pixel to the pixel evaluated. The pixels are sampled from the center of the image with a radius R.

From the samples found above, the max and min samples are found as follows:

$$s_i^{max} = \max_{j \in \{0,...,M\}} p_j \tag{4.38}$$

$$s_i^{min} = \min_{j \in \{0,...,M\}} p_j \tag{4.39}$$

The sample point p_0 is always a part of the sample points, and therefore $s_i^{min} \leq p_0 \leq s_i^{max}$ and the range r_i of the samples and the relative values v_i of the center pixel will be as

$$r_i = s_i^{max} - s_i^{min} \tag{4.40}$$

$$v_i = \begin{cases} 1/2 & \text{if } r_i = 0 \\ (p_0 - s_i^{min})/r_i & \text{else} \end{cases} \tag{4.41}$$

To get a better estimate of the values, these are averaged over the N iterations

$$\bar{r} = \frac{1}{N}\sum_{i=1}^{N} r_i \tag{4.42}$$

$$\bar{v} = \frac{1}{N}\sum_{i=1}^{N} v_i \tag{4.43}$$

Kolas et al. argues that averaging r_i and v_i directly ensures that the the algorithm is edge-preserving and does not introduce halo artifacts.

The actual envelopes are then computed from from the estimated average range and pixel value as follows:

$$E^{min} = p_0 - \bar{v}\bar{r} \tag{4.44}$$

$$E^{max} = p_0 + (1-\bar{v})\bar{r} = E^{min} + \bar{r} \tag{4.45}$$

It is worth noticing that the envelopes is dependent on the whole image.

The algorithm has the three input variables described above, M as the number of sample points, R as the radius of the sample points, and N as the number of iterations.

The numbers of sample points, M can range from just one in addition to the sample point. Tests show that as the number of sample points grow, the algorithm grows more global to the point where it is reduced to global linear contrast stretching.

As for the number of iterations, N this will affect the chromatic noise in the result. The higher the number of iterations, the lower amount of noise. With N in the range of the radius R has shown to give very limited amounts of noise, and can be used as a criterion for maximization.

The last input is the radius R. This is the maximum distance from that samples can be collected, and controls the locality for the spatial maxima and minima for the adjustment. Kolås et al. argues that this is not a critical parameter, and the value in real world application should be chosen large enough to cover the entire image.

As described, only tone mapping of HDR images to a LDR copy is presented, but the algorithm can also be used to do local contrast enhancements of grayscale images, do local correction to color images and do full color to grayscale conversion. The algorithm has also been implemented for usage in movies and has shown good results as a temporal color correction method for degraded movies.

It can directly be applied to the image without any modification as a tone-mapping function as long as the HDR data is mapped in a gamma corrected or a perceptually uniform space. The algorithm will rearrange spatial and local values in the image while preserving the edges and compressing the gradients.

The standard output of the images does not only do a tone rendering, but also local color adjustment, contrast stretching and luminosity normalization, while not introducing artifacts such as halos. It is noted in Kolås et al. that images can look a little like high-pass filtered images if they are compressed to much.

Rahman Retinex

Rahman and Jobson [17] developed their interpretation of of the retinex theory for a wide variety of applications. Their operator does not differ much from Chiu's approach with blurring the image, but the operator comes in two different scales: single-scale and multiscale. The single-scale is a version almost a copy of Chiu's operator, except that all transformations are done in the log domain. Also the placement of the logarithms are a bit special as they are convolved with the Gaussian filter kernel, and have been empirically determined to produce visually improved results. The operator also works independently on each channel, so the convolution needs to be performed three times pr picture.

For the multiscale retinex the process of creating Gaussians is repeated several times with different kernel sizes. This effectively creates a stack of n images with increased blurring. The operator then gives a weighted sum for each image, determined by the user. It has been shown that by using a power function a straightforward control of the weights is achieved.

For a stack of N image levels, the normalized weights are computed by the following formula:

$$w_n = \frac{(N-n-1)^f}{\sum_{n=0}^{N}(N-m-1)^f} \tag{4.46}$$

Here f is defined as a user parameter which determines the relative weight of each of the scales. For equal weighing, we should set this to 0. Values larger than this emphasizes smaller scales and negative numbers brings forward larger scales. The form for the multiscale retinex is as follows:

$$I_d(x,y) = \exp\left(\sum_{n=0}^{N} w_n \left(log(I_w(x,y)) - k\log\left(I_{w,n}^{blur}(x,y)\right)\right)\right) \tag{4.47}$$

f we have already described, but also k is in many ways similar the user parameters which describe Chiu's operator that specifies the relative weight of the blurred image. There are however three main differences. The operator works in the log domain which results in that large image values lie closer together. Secondly it operates on the three color channels independently and third the algorithm operates on multiple scales, which also is incorporated in other TMOs.

Fattal Gradient Domain

Fattal et al. [18] presented in 2002 a new way to do tone mapping, by manipulating the gradient field of the luminamces.

The idea behind their method is that drastic changes in the luminance on a HDR image must be seen in the increase of in large magnitude luminance gradients. Smaller details will then be connected to gradients of smaller magnitude, and by identifying the large gradients and reduce their magnitude without altering their direction will compress the luminance. The alteration is progressive, so larger gradients are penalized more than smaller changes. This results in that drastic luminance changes are heavily compressed, and fine details are preserved.

For image data which can be viewed as a 2D function, Fattal et al. define the gradient ∇I by orthogonally project the gradient onto a finite set of orthonormal basis functions. This is done by searching the space of all 2D potential functions for a function I whose gradient gives the least

distance when checked with the least-square method. This can be formally written as the the integral

$$\int\int F(\nabla I, G)dxdy \tag{4.48}$$

Here G is the gradient multiplied by the attenuation function and

$$F(\nabla I, G) = ||\nabla I - G||^2 = \left(\frac{\partial I}{\partial x}G_x\right)^2 + \left(\frac{\partial I}{\partial y}G_y\right)^2 \tag{4.49}$$

This function must according to the variational principle, satisfy the Euler-Lagrange equation

$$\frac{\partial F}{\partial I} - \frac{d}{dx}\frac{\partial F}{\partial I_x} - \frac{d}{dy}\frac{\partial F}{\partial I_y} = 0 \tag{4.50}$$

By substitution of F and divide by 2, we get the well-known Possion equation:

$$\nabla^2 I = \nabla \cdot G \tag{4.51}$$

The next step is the progressive reduction function that alters the magnitudes of the luminances for each pixel, denoted as $\phi(x, y)$

Gradients cannot simply be reduced at the resolution where it is detected as this would lead to halos around strong edges. Fattal et al. proposes that all found gradients are propagated to the full resolution of the image and that all compression is performed there. This is done using a Gaussian pyramid with at least 32 steps. For each step, the gradient ∇H is calculated, and by using central differences this can be written as follows:

$$\nabla H_k = (\frac{H_k(x+1, y) - H_k(x-1, y)}{2^{k+1}}, \frac{H_k(x, y+1) - H_k(x, y-1)}{2^{k+1}}) \tag{4.52}$$

For each of these levels k, a scaling factor $\phi_k(x, y)$ is determined based on the gradient there:

$$\phi_k(x, y) = \frac{\alpha}{||\nabla H_k(x, y)||}\left(\frac{||\nabla H_k(x, y)||}{\alpha}\right)^{\beta} \tag{4.53}$$

The α parameter here determines which gradient magnitudes that remain unchanged, while β controls which gradients should be attenuated. $\beta < 1$ assumes that gradients larger than α should be reduced while gradients smaller than α are slightly magnified.

As for color, the same approach as Tumblin and Turk [28] and Schlick [29] is used:

$$C_out = \left(\frac{C_{ind}}{L_{in}}\right)^s L_{out} \tag{4.54}$$

where C represents each of the R,G and B channels, L_{in} and L_{out} denotes the luminances before and after the compression and the exponent s controls the saturation. Fattal et al. recommends that this is kept between 0.4 and 0.6.

This method is the first that looked into image compression into other domains, and inspired a lot of newer TMOs. It has also shown very good results on a lot of different usages.

Fairchild iCAM

Mark D. Fairchild presented in 2002 a model for color appearance called iCAM [19]. Most models before this, like the CIECAM97, CIECAM02 and the Hunt model, are intended for usage in simplified environments, but the iCAM model works in more complex scenes where the effects of neighboring pixels and local adaptation of these are taken into account. This model is heavily inspired by the models of Chiu and Rahman because of their spatial tone-reproduction operators also are relevant for spatially variant color appearance models.

Based on CIECAM02, iCAM omits the sigmoidal compression step, but adds spatially variant processing by using two separate Gaussian-blurred images. The input is expected in XYZ device independent coordinates. The model consist of several stages, The first is a chromatic adaptation transform using sharpened cone responses. This pushes the colors in the image towards D_{65} white point. The whitepoint information may be applied to all channels for chromatic adaptation, or only for the Y-channel for achromatic adaptation. The next step is the von Kries transformation, where the images is divided into filtered versions of the image for each RGB-channel. In this step the scaling factor D plays the same role as the k in Chiu's and Rahmans operator and have the same effect, but it also determines the chromatic adaptation.

The transformation is given by

$$\begin{aligned} R_c(x,y) &= R'(x,y)\left(Y_W \frac{D}{W_{R1}(x,y)} + 1 - D\right) \\ G_c(x,y) &= G'(x,y)\left(Y_W \frac{D}{W_{G1}(x,y)} + 1 - D\right) \\ B_c(x,y) &= B'(x,y)\left(Y_W \frac{D}{W_{B1}(x,y)} + 1 - D\right) \end{aligned} \tag{4.55}$$

where the whitepoint $Y_W = 95.05, 100.0, 108.88$,

Large values of D results in that the whitepoint of each pixel is pushed against the D_{65} whitepoint, so compression, halo and chromatic adaptation is all related to this.

To do further compression, a exponential function is executed in the LMS color space. The function is modified on a per-pixel basis by the surrounding map of pixels derived from the luminance channel of the input image as follows

$$\begin{aligned} L'(x,y) &= |L(x,y)|^{0.43F_L(x,y)} \\ M'(x,y) &= |M(x,y)|^{0.43F_M(x,y)} \\ S'(x,y) &= |S(x,y)|^{0.43F_S(x,y)} \end{aligned} \tag{4.56}$$

This surrounding map $S(x,y)$ is a low-pass filtered version of the luminances with a Gaussian filter kernel size of $1/3$ of the size of the image. This results in a function which may be seen as the spatially variant extension of CIECAM02's factor for partial adaptation. The function F_L is given by

$$F_L = \frac{1}{1.7}\left(0.2\left(\frac{1}{5S(x,y)+1}\right)^4 (5S(x,y)) + 0.1\left(1-\left(\frac{1}{5S(x,y)}\right)^4\right)^2 \sqrt{35}S(x,y)\right) \tag{4.57}$$

The model is also reversed. The first step is to invert the exponential function. This does not require a spatially variant white point as the global white point was pushed towards D_{65}. The result is

that $Y_e = 100.00, 100.00, 100.00$. The next step will then be to reverse the von Kries transformation, and because full adaptation is assumed, D is set to 1, which simplifies the scaling to multiplying the channel by the dividend of Y_e and Y_W.

These steps are executed in their appropriate color spaces, the final steps is to clip the top 99% of all the pixels, normalize and do the gamma correction.

This model is best suited for images with a medium dynamic range, as the trade-off between compression and the presence of halos is less a critical for these classes of images than for extreme HDR images.

Pattanaik Multiscale Observer Model

This model is one of the more complete color appearance models and consist of several steps, which results in a set of color appearance correlates. A tone-reproduction operator may be derived from these correlates by executing the reverse model and insert the the characteristics of the image display [20].

There are several versions of this model, some with steps that cancel each other out, so this is a representation of the model as described in [4].

The models first step is to account for light scatter. This is done for the ocular media, and is followed by a spectral sampling, which is used to model the photoreceptor output which results in four images representing the rods, L, M and S cones. A image at level s is represented with the following triplet:

$$\left(L_s^{blur}(x,y), M_s^{blur}(x,y), S_s^{blur}(x,y)\right) \tag{4.58}$$

These are converted into six-level Difference of Gaussians (DoG) stacks which is a representation of the bandpass behavior we see in the HVS:

$$\begin{aligned} L_s^{DoG}(x,y) &= \left(L_s^{blur}(x,y) - L_{s+1}^{blur}(x,y)\right) G\left(L_{s+1}^{blur}(x,y)\right) \\ M_s^{DoG}(x,y) &= \left(M_s^{blur}(x,y) - M_{s+1}^{blur}(x,y)\right) G\left(M_{s+1}^{blur}(x,y)\right) \\ S_s^{DoG}(x,y) &= \left(S_s^{blur}(x,y) - S_{s+1}^{blur}(x,y)\right) G\left(S_{s+1}^{blur}(x,y)\right) \end{aligned} \tag{4.59}$$

The low-pass image at level $s = 7$ is retained, and forms the basis for reconstruction. All pixels in this step are adapted to a linear combination of the mean value $\bar{L}_7^{blur}, \bar{M}_7^{blur}, \bar{S}_7^{blur}$ of the low-pass image:

$$\begin{aligned} L_7^{blur}(x,y) &= L_7^{blur}(x,y) G\left((1-A)\bar{L}_7^{blur} + AL_7^{blur}\right) \\ M_7^{blur}(x,y) &= M_7^{blur}(x,y) G\left((1-A)\bar{M}_7^{blur} + AM_7^{blur}\right) \\ S_7^{blur}(x,y) &= S_7^{blur}(x,y) G\left((1-A)\bar{S}_7^{blur} + AS_7^{blur}\right) \end{aligned} \tag{4.60}$$

where the parameter A controls how much the dynamic range should be reduced. This can be a value between 0 and 1.

The inverse model is then applied, and values such as mean luminance $L_{d,mean}$ of the display device is set. The gain factor is computed and the low-pass image is adapted to the display luminance as follows:

$$\begin{aligned} L_7^{blur}(x,y) &= \frac{L_7^{blur}(x,y)}{G(L_{d,mean})} \\ M_7^{blur}(x,y) &= \frac{M_7^{blur}(x,y)}{G(L_{d,mean})} \\ S_7^{blur}(x,y) &= \frac{S_7^{blur}(x,y)}{G(L_{d,mean})} \end{aligned} \tag{4.61}$$

The stack of DoGs is then added to the adapted low-pass image one scale at a time, $S = 6 \cdots 0$:

$$\begin{aligned} L_7^{blur}(x,y) &= \max\left(L_7^{blur}(x,y) + \frac{L_s^{DoG}(x,y)}{G(L_7^{blur}(x,y))}, 0\right) \\ M_7^{blur}(x,y) &= \max\left(M_7^{blur}(x,y) + \frac{M_s^{DoG}(x,y)}{G(L_7^{blur}(x,y))}, 0\right) \\ S_7^{blur}(x,y) &= \max\left(S_7^{blur}(x,y) + \frac{S_s^{DoG}(x,y)}{G(L_7^{blur}(x,y))}, 0\right) \end{aligned} \tag{4.62}$$

The final step is to convert back to XYZ and to RGB where gamma correction is done according to the display gamma.

The model results in some haloing artifacts due to the fact that visual angle to pixels is needed. It is discussed that by including a weighting scheme so that higher frequencies get higher weights can be beneficial. A scale factor k used for the images with an index S of the band-pass pyramid can be defined:

$$k = (6 - S)g \tag{4.63}$$

where the parameter g is defined by the user between 1 to 5. Higher values increases contrast in the tone-mapped image, but may produce halos if pushed to high.

Pattanaik's Multiscale observer model is the first to introduce the concept of center-surround processing, and has later been used in other tone mappers such as Ashikhmin [30] and Reinhard et al. photographic tone operator described in Section 4.2.2.

Reinhard Photographic Tone Reproduction

The Reinhard Photographic tone reproduction tone operator [14] takes its ideas from analogue photography where a lot of different methods are used to produce realistic "renderings" of captured scenes. The same problems that exists in digital photography exists in the analogue world with the limitations of films, processing and photographic papers.

Ansel Adams tried to formalize some of these methods by introducing the "Zone System", which was first described in the 1940s. This is a system of "practial sensiometry", where the photographer uses measured information to increase the chance for a final good print.

The Zone System is defined by 11 print zones, from pure black to pure white named with the roman numerals I - XI. The first step is to determine the middle brightness which is subjectively measured, often set to zone V. The next step is to find the dynamic range of the image. This can be expressed as the difference between the highest and lowest distinguishable scene zones. The last step is to determine the key of the zone. Here "key" means if the scene is subjectively light, (high-key) normal or dark (low-key). All measurements are done for the luminances in the different parts of the scene.

The tone operator is divided into two different parts, the initial luminance mappings and the "dodge-and-burning" step. For the initial luminances the log-average luminance is used as an approximation of the key in the scene. This can be computed by the following formula:

$$\bar{L}_w = \frac{1}{N} \exp(\sum_{x,y} log(\delta + L_w(x,y))) \tag{4.64}$$

Here $L_w(x,y)$ is the "world luminance" for pixel (x,y), N is the total number of pixels in the image and δ is a small value to avoid the singularity that appears for black pixels.

As for the key of the image, this is for normal-key mapped to the middle-grey of the image. Reinhard et al. proposes that this may be modeled using the following equation:

$$L(x,y) = \frac{\alpha}{\bar{L}_w} L_w(x,y) \tag{4.65}$$

The value $L(x,y)$ is a scaled luminance and the α equals the scaling factor. This value may be adjusted by the user in the range 0 - 1. The parameter α is often referred to as the "key value".

This equation shows a problem as most scenes mostly have a predominantly normal dynamic range, but have a few high luminance regions near highlights or in the bright sky for outside scenes. Reinhard et al. proposes an controlled burning of these areas by introducing the following:

$$L_d(x,y) = \frac{L(x,y)\left(1 + \frac{L(x,y)}{L_{white}^2}\right)}{1 + L(x,y)} \tag{4.66}$$

L_{white} is here the smallest luminance that will be mapped to pure white. If this is set to L_{max} or higher, no burn-out will occur.

The last step is the dodge-and-burn. This step is necessary for images that have a very high dynamic range because fine detail is lost. The idea behind dodge-and-burn is to give different parts of the picture different exposure time. In digital photography this may be thought of as defining a key value for every pixel, which is equal to setting a local α in Equation (4.65).

In the process of dodging-and-burning, entire regions are bound by large contrasts. To determine these areas, a measurement of local contrast computed at multiple scales is often used. A variety of these functions have been proposed, and Reinhard et al. chose to use use a center-surround function derived from Blommaerts model for brightness perception. This function has the form

$$R_i(x,y,s) = \frac{2}{\pi(\alpha_i s)^2} \exp\left(-\frac{x^2+y^2}{(\alpha_i s)^2}\right) \tag{4.67}$$

These functions can be called profiles, and will operate at different scales. The resulting response V_i will be a function of the image location, scale s and luminance distribution L as follows:

$$V_i(x,y,s) = L(x,y) \otimes R_i(x,y,s) \tag{4.68}$$

The smallest Gaussian profile will only be slightly larger than one pixel, and therefore the accuracy of the above function is very important. By performing the integration in terms of the error function, Reinhard et al. shows that they gain high enough accuracy without the usage of supersampling.

The center-surround can therefore be defined as follows:

$$V(x,y,s) = \frac{V_1(x,y,s) - V_2(x,y,s)}{\frac{2^\Phi \alpha}{s^2} + V_1(x,y,s)} \tag{4.69}$$

Here V_1 is the center, V_2 is the surround and Φ is a parameter which controls the sharpening. To ease computation for the model, the center size is set to the same as the surround of the current scale. Reinhard et al. presents that a ratio of 1.6 between the center-surround gives the best result.

The purpose of this function is to to measure the locality for each pixel, for which a scale s_m can be derived from. This can be found by choosing the largest neighborhood with fairly the same luminances. This can be formally written:

$$|V(x,y,s_m)| < \epsilon \tag{4.70}$$

When the first threshold, ϵ, is found, $V_1(x,y,s_m)$ may serve as a local average for the pixel, and can be converted into a local operator:

$$L_d(x,y) = \frac{L(x,y)}{1 + V_1(x,y,s_m(x,y))} \tag{4.71}$$

Reinhard et al. observe with this operator that relatively bright regions satisfies $L < V_1$, and here the display luminance will be decreased, and thus the contrast increased. This is very similar to the photographic process of "dodging", and where $L > V_1$ the regions will "burned" and get lighter.

If s_m is to small and V_1 is close to the luminance L, the local operator is reduced to the global operator, L_d. If the value is to big, this will cause dark rings around bright areas to appear.

The work of Reinhard et al. is a good approximation of the dodge and burning technique used in analogue photography, and the results are very similar to the effects seen when the exposure of film is locally changed to darken or brighten the final print.

Mantiuk Perceptual Framework for Contrast Processing

This TMO is a framework for converting images to contrast and by the usage of a transducer function which maps appearing contrast to physical contrast.

The model consist of several steps. First the pixel luminance values are transformed to physical contrast values using the observations of Peli [31] where contrast is not viewed as only periodic nor local, and should be considered quasilocal. Peli proposed to measure contrast as a difference between

selected levels of a Gaussian pyramid, but this introduces halo artifacts at sharp edges. Mantiuk et al. use a logaritmic ratio G as a low-pass measure to overcome this problem:

$$G_{i,j}^k = \log_{10}\left(\frac{L_i^k}{L_j^y}\right) = x_i^k - y_j^k \tag{4.72}$$

Here L_i^k and L_j^k are luminance values for neighboring pixels i and j. For a single pixel there are now two or more conrast measures $G_{i,j}^k$ depending how many neighboring pixels that are considered. Test have shown that for tone-mapping purposes, usage of the two nearest neighbors are sufficient for complex images. For other applications like mapping from color-to-gray, a neighborhood of 20-30 pixels should be considered.

The next step will be to use a transducer function to transfer the contrast G to hypothetical response of the HVS. Here Mantiuk et al. follows the transducer-function found by Wilson [32], and derives the function $T(G) = R$. This is simplified for the case of threshold stimuli:

$$T(0) = 0 \qquad\qquad T(G_{threshold}) = 1$$

or

$$T^{-1}(0) = 0 \qquad\qquad T^{-1}(1) = G_{threshold}$$

where T^{-1} is the inverse transducer. The detection threshold, $G_{threshold}$ is approximated with 1% of the contrast which is commonly used in digital images. This assumes that the detection threshold for all spatial frequencies is the same, and we can approximate a function as follows fro the suprathreshold:

$$\Delta T \approx \frac{dT(g)}{dG}\Delta G(G) = 1$$

The function states that a unit increase of response R on the right side should correspond to the increase of G equal to the discrimination threshold ΔG for the contrast G.

This can be solved using its differential part, but it is more convenient to solve this numerically by solving the equal differential equation:

$$\frac{dT^{-1}(R)}{dR} = \Delta G(T^{-1}(R))$$

Here $T^{-1}(R) = G$ is the inverse response function, G is a nonnegative logarithmic ratio and R is the response from the HVS.

By solving this numerically, we get a transducer function which is very similar to a logarithmic or exponential function, but none are an exact fit. If an accurate model is not necessary, the transducer can be approximated by the following function:

$$T(G) = 54.09288 \cdot G^{041850} \tag{4.73}$$

or for the inverse transducer

$$T^{-1}R = 7.2232 \cdot 10^{-5} \cdot R^{2.3895} \tag{4.74}$$

The major difference between this transducer and other operators that take an similar approach, like Ashikhmin [30] is that this transducer operates in the contrast domain rather than in the luminance domain.

The above step has was to convert images to perceptually linearized contrast responses, and then restoring these to images through the inverse. The framework then can be used for several things, but as we are only interested in tone mapping, we will only describe the contrast mapping step where the actual scaling is done.

Mantiuk et al. calls this contrast compression instead of tone mapping. Other operators try to overcome the insufficient dynamic range of a display, or the proper reproduction of luminances on a screen whereas Mantiuk et al. does not try to overcome this nontrivial step, but instead do a linear scaling in the visual response space. This is expressed as follows:

$$\hat{R}^k_{i,j} = R^k_{i,j} \cdot l \tag{4.75}$$

where $R^k_{i,j}$ is the visual response and l is the scale factor.

This can be achieved because $R^k_{i,j}$ is perceptually linearized. The compression scheme done here resembles the gradient domains method proposed by Fattal et al. [18], but the method is based on perceptual characteristics of the eye instead of Fattal et als. ad-hoc approach.

Mantiuk et al. tests their approach for a broad range of images, and the only apparent problem they found, was that in low light condition, image noise was magnified. No halo or ringing effects were observed.

The logarithmic ratio defined in Equation (4.72) can be used to transform luminance to contrast by minimization of the distance between a set of contrast. This can be formally written as the minimization of the function:

$$f\left(x^1_1, x^1_2, \ldots, x^1_N\right) = \sum_{k=1}^{K} \sum_{n=1}^{N} \sum_{j \in \Phi i} p^k_{i,j} \left(G^k_{i,j} - \hat{G}^k_{i,j}\right)^2 \tag{4.76}$$

with the regard to pixel values x^1_i on the finest level of the pyramid, Φ_i is the set of neighbors of the pixel, N is the total number of pixels, K is the number of levels in the Gaussian pyramid, $p^k_{i,j}$ is a constant weight factor to control for mismatch in desired contrast $\hat{G}^k_{i,j}$ and actual contrast $\hat{G}^k_{i,j}$.

The objective function reaches its minimum when all its derivatives $\frac{\partial f}{\partial x_i}$ equal 0:

$$\frac{\partial f}{\partial x_i} = \sum_{k=1}^{K} \sum_{n=1}^{N} \sum_{j \in \Phi i} 2p^k_{i,j}(x^k_i - x^k_j - \hat{G}^k_i, j) = 0 \tag{4.77}$$

for $i = 1, \ldots, N$. The equation may be rewritten using a matrix notations as follows:

$$A \cdot X = B \tag{4.78}$$

where X is the column vector of $x_1, \ldots, x_N$ which holds the pixel values for the resulting image, A is a $N \times N$ square matrix and B is an N-row vector.

Mantiuk et al. proposes to solve these using a biconjugate gradient method [33], which gives a result in acceptable time. The method involves a iterative procedure where the image is stored in a vector X and is refined for each iteration. This method is chosen because this only requires an efficient computation of the product $\Psi = A \cdot X$.

The contrast is then computed between a pixel and its four neighbors with the same level of a Gaussian pyramid, and if only the first level of the pyramid is considered, the problem is reduced to the solution of the Poisson's equation as in Fattal et al.

The last step in Mantiuk et al. extensive framework is the reconstruction of color. Here they follow the same method proposed by most tone mapping algorithms, which involves rescaling the RGB-channels proportionally to the luminance and desaturation of the colors to compensate for the higher local contrast:

$$C_{out} = \frac{X - l_{min} + s(C_{in} - L_{in})}{l_{max} - l_{min}} \tag{4.79}$$

Here C_{in} and C_{out} are the input and output pixels for the different channels, L_{in} is the input luminance and X is the result of the optimization. All values are calculated in the logarithmic domain. This function is similar to the functions used by Tumblin and Rushmeyer [28], but the method proposed my Mantiuk et al. is given in the logarithmic domain an includes a linear scaling. The resulting color values, C_{out} can then be directly mapped to pixels values of a gamma-corrected display.

4.3 Perceptual Experiment

In the field of perceptual experiments there is a wide variety of different experiments that can be used to compare images. One of the simplest to perform, is the pairwise comparison where a test subject will compare two different images and choose the one preferred or is evaluated better in some quantitative measure. The first to introduce this was L.L. Thurstone with his law of comparative judgement [34].

The full definition of Thurstone law is defined as

$$S_i - S_j = x_{ij}\sqrt{\sigma_i^2 + \sigma_j^2 - 2r_{ij}\sigma_i\sigma_j} \tag{4.80}$$

where S_i is the psychological scale vale of stimuli i, x_{ij} is the sigma corresponding to the proportions of occasions which the stimuli i is judged to exceed the magnitude of stimuli j, σ_i is the discrimininal dispersion of a stimuli R_i and r_{ij} is the correlation between the dicriminal deviation of stimuli i and j.

The law consist of five different cases, each consisting of different assumptions which simplifies the law. The most important case for us is case V, where it is shown that the discriminal dispersions are spescified to be uniform and uncorrelated. This is the simplest form of the law, and can be represented as follows:

$$x_{ij} = \frac{S_i - S_j}{\sigma} \tag{4.81}$$

where

$$\sigma = \sqrt{\sigma_i^2 + \sigma_j^2} \tag{4.82}$$

4.4 Perceptual Uniform Encoding

Aydın et al. presented in [35] a method to extend quality metrics to HDR images by the usage of perceptual uniform (PU) encoding. The method is based on the fact that details on bright displays are much more visible than the same details viewed on a dark display. This is confirmed through a psychophysical experiment.

The results plotted in a cvi (contrast versus intensity) and luminance diagram shows that there is a big difference in threshold contrast. This contradicts the Weber-Fechner law which states that this ratio stays constant, and it can be concluded that the Weber-Fechner law is a very inaccurate model of luminance masking for todays bright displays.

To try to find a better function that describe this behavior, Aydın et al. estimates the contrast detection thresholds as a function of pixel luminance. A lot of factors affect this in complex images, but to try to simplify the problem, all these factors are reduced except adapting luminance L_a. All the other factors is assumed to be taken care of in the quality metric. Following this, the cvi function is defined as follows:

$$\text{cvi}(L, L_a) = \left(\max_x [\text{CSF}(L_a, x) MA(|L - L_a|)] \right)^{-1} \tag{4.83}$$

where CSF is the contrast sensitivity function from Daly [36] and x corresponds to all the parameters except adapting luminance L_a and background luminance L. The MA function is a description of maladaptation, which can be described to the fact that the HVS is not capable of perfectly adapting to each tiny luminance patch in a scene. The functions used to account for this, is based on so-called *probe-on-flash* experiments and these data are furthermore combined with the cvi functions as done by Irawan et al. [37].

To make the extension spatially independent, Aydın et al. makes two assumptions that simplify the luminance adaptation process, a minimal luminance level, L_{a-min} and that the HVS is perfectly adapted for all luminance levels above L_{a-min}. The last point makes the adapting luminance the equal of the luminance in the pixel ($L_a = L$).

This results in the final estimates of the detection thresholds:

$$t(L) = \text{cvi}(L, \max(L, L_{a_m in})) \tag{4.84}$$

Using these thresholds, the author shows that the distortion visibility is uniform along all encoded values when the differentials of the encoded values are proportional to the detection thresholds. This is obtained using a recursive formula:

$$f_i = f_{i-1}(1 + t(f_{i-1})) \text{ where } f : L^{'} \rightarrow L, i \in [2 \cdots N] \tag{4.85}$$

where f_1 is the minimum luminance to be encoded, N is selected so that it is larger than the maximum luminance to be encoded.

The forward mapping function for a given luma value is given as $PU : L \rightarrow L^{'}$, is achieved by using the value f as a lookup table and find the nearest (or interpolated) index i.

Furthermore, the method is made backward compatible to the sRGB non-linearity within the range of 0.1 - 80 cd/m^2:

$$\sum_{L=0.1}^{80} ((sPU(L, L_{a_m in} + m) - l^{-1}(L)))^2 \tag{4.86}$$

where the summation is performed for 256 logarithmically distributed luminance values. s can be looked upon as the absolute sensitivity factor, which varies between observers, m adjusts the absolute encoding response to fit to sRGB. The term PU is the inverse of Equation (4.85), and l^{-1} is the reverse of the sRGB non-linearity.

Aydın et al. argues that a look-up table should be used instead of power or logarithmic functions to describe the PU encoding because of better accuracy and that it is faster to compute.

4.5 Image Quality Metrics

Image quality metrics is defined as a algorithm or method that tries to show the amount of distortions or artifacts on a perfect image. In the field of photography, the original is the scene which is captured, and imaging capture systems can introduce noise, and therefore it is really hard to measure the quality using other than using subjective measurements.

In image reproduction, different components such as processing or reproduction technique may introduce artifacts or distortions on the original signal, and here a wide variety of methods and algorithms have been and still is developed.

Image quality can be divided into a wide range of factors. Čadík et al. presented the following attributes in Figure 1 to be most important for image quality in tone mapping settings in [38]:

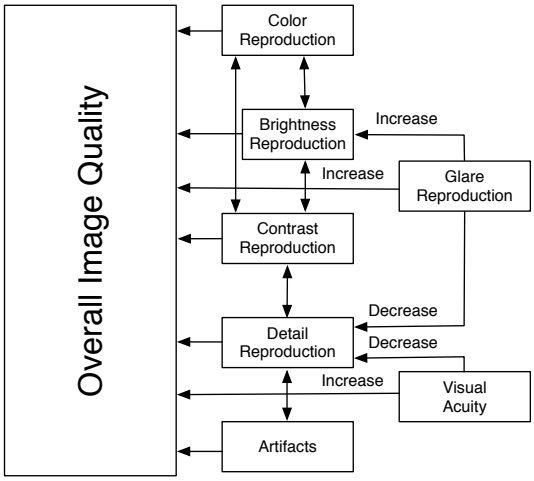


Figure 1: Image Quality attributes in tone mapping.

As shown here, image quality is dependent on the different factors, and therefore it is very hard to give a single measure which will give a correct score in all possible situations.

Under we therefore present some state of the art metrics that tries to evaluate these attributes.

4.5.1 Measurement Metrics

In this section we present a set of measurement metrics that only does simple measurement on differences in image data. No psychophysical information is taken into account.

Mean Square Error

Mean Square Error(MSE) was first introduced by Carl Friederich Gauss [39], and has become one of the most widely used loss functions in statistics, especially in analyzing the performance of linear regression and variation i datasets explained by randomness.

Formally MSE for a unbiased sample in a population of size n can be expressed as

$$\mathrm{MSE}(\bar{X}) = E((\bar{X} - \mu)^2) = \left(\frac{\sigma}{\sqrt{n}}\right) = \frac{\sigma^2}{n} \tag{4.87}$$

In the field of computer graphics, one of the usages of this measure is to measure the difference between an original and a distorted image. Here the distance for each pixel is compared, which can be expressed as

$$\mathrm{MSE} = \frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N}\left(x_{i,j} - x'_{i,j}\right)^2 \tag{4.88}$$

where M and N are the size pictures, and x and x' are the corresponding pixels in each image.

This metric is only looking at the pixel values, and does not take any perceptual information into account. Therefore it should be only used as a reference for other metrics

Peak Signal-to-Noise Ratio

Peak Signal-to-Noise Ratio (PSNR) is closely linked to MSE, and describes the ratio between the maximum possible energy of a signal and the power of the corrupting noise affecting the quality of the signal. As many signals often have a wide dynamic range, PSNR is often expressed in dB, which is a logarithmic scale.

In computer graphics this is often used both in images and video and can be considered a valid quality measure [40], and is defined as

$$\begin{aligned}\mathrm{PSNR} &= 10 \cdot \log_{10} \frac{(2^n - 1)^2}{\mathrm{MSE}} \\ &= 10 \cdot \log_{10} \frac{255^2}{\mathrm{MSE}}\end{aligned} \tag{4.89}$$

where n is the number of bits in the input image. For color images, alternative color spaces such as YCbCr or HSL is often used.

$\Delta\mathbf{E}^*_{ab}$

ΔE is a measure of the distance between to two different colors. ΔE^*_{ab} is also just named Delta E or $\Delta E94$ and defined by CIE in [41].

The theory is based on the fact that a ΔE of under 1 are not detectable by the HVS, but due to perceptual problems with the different color spaces, the measure has undergone some changes over the years. This has been studied to a great extent because the HVS is more sensitive to certain colors, and the JND for different colors in the spectrum is shown to vary.

This version is often used as a reference against other evaluations, but has in the later years been replaced by CIEDE2000.

ΔE^*_{94} is defined in the CIE $L^*a^*b^*$ color space as the difference between as sample color $L_2a_2b_2$ and a reference color $L_1a_1b_1$ as follows:

$$\Delta E^*_{94} = \sqrt{\left(\frac{\Delta L^*}{K_L}\right)^2 + \left(\frac{\Delta C}{K_C S_C}\right)^2 + \left(\frac{\Delta H}{K_H S_H}\right)^2} \tag{4.90}$$

where

$$\begin{aligned}
\Delta L &= L_1 - L_2 \\
\Delta C &= C_1 - C_2 \\
\Delta H &= \sqrt{\Delta a^2 + \Delta b^2 + \Delta C^2} \\
C_1 &= \sqrt{a_1^2 + b_1^2} \\
C_2 &= \sqrt{a_2^2 + b_2^2} \\
\Delta a &= a_1 - a_2 \\
\Delta b &= b_1 - b_2 \\
S_L &= 1 \\
S_C &= 1 + K_1 C_1 \\
S_H &= 1 + K_2 C_1 \\
K_L &= 1 \\
K_C &= 1 \\
K_1 &= 0.045 \\
K_2 &= 0.015
\end{aligned}$$

This is the version used for graphic arts. If the metric should be used on paper prints, other values should be inserted.

4.5.2 Perceptual Metrics

In this section, a small set of image difference metrics that take into account HVS data, and tries to measure different parts of these.

UIQ

Wang and Borvik [42] presented a Universal Image Quality Index,(UIQ) which is based on a mathematical approach much like MSE and PSNR. This means that no actual viewing conditions or HVS models are incorporated, but it distinguishes different types of distortions. In this thesis this metric is included as it forms a basis for SSIM-index.

The quality index is defined as

$$Q = \frac{4\sigma_{xy}\bar{x}\bar{y}}{(\sigma_x^2 + \sigma_y^2)[(\bar{x})^2 + (\bar{y})^2]} \tag{4.91}$$

where

$$\begin{aligned}
\bar{x} &= \frac{1}{N}\sum_{i=0}^{N} x_i \\
\bar{y} &= \frac{1}{N}\sum_{i=0}^{N} y_i \\
\sigma_x^2 &= \frac{1}{N-1}\sum_{i=0}^{N}(x_i - \bar{x})^2 \\
\sigma_y^2 &= \frac{1}{N-1}\sum_{i=0}^{N}(y_i - \bar{y})^2 \\
\sigma_{xy} &= \frac{1}{N-1}\sum_{i=0}^{N}(x_i - \bar{x})(y_i - \bar{y})
\end{aligned} \tag{4.92}$$

The range of Q ranges from $[-1, 1]$, where 1 is a perfect match and only occur when $y_i = x_1$ for all pixels.

Wang et al. describes that the UIQ models any distortion as a combination of three factors:

- Loss of correlation
- Luminance distortion
- Contrast distortion

These three can be summed into Q as follows:

$$Q = \frac{\sigma_{xy}}{\sigma_x\sigma_y} \cdot \frac{2\bar{x}\bar{y}}{(\bar{x})^2 + (\bar{y})^2} \cdot \frac{2\sigma_x\sigma_y}{\sigma_x^2 + \sigma_y^2} \tag{4.93}$$

The first component in the above equation is the loss of correlation, and so on.

When applied to images, the implementation uses a square sliding window with a given size over the image. The quality of each windows is summed up as follows:

$$Q = \frac{1}{M}\sum_{j=1}^{M} Q_j \tag{4.94}$$

where M is the number of steps, Q_j is the local quality index.

The proposed size if the window $B = 8$, and Wang et al. shows that this value works very well.

SSIM index

SSIM index, or structured similarity was proposed by Wang et al. [43] and tries to quantify the visible difference between a distorted image and a reference. The metric is based on UIQ presented in Section 4.5.2, and exploits the fact that natural images are highly structured, and that the pixels in such images exhibit strong dependencies, especially when they are spatially proximate. The HVS is also assumed to be highly adapted to extract structural information from the viewing field, and a measure of changes in the structure can be used a a quality measure.

Wang et al. uses the luminance on the surface of an object, and they define the structural information in the image as the attributes that represents the structure, independent of the average luminance and contrast. This assumption makes the metric independent of the illuminant.

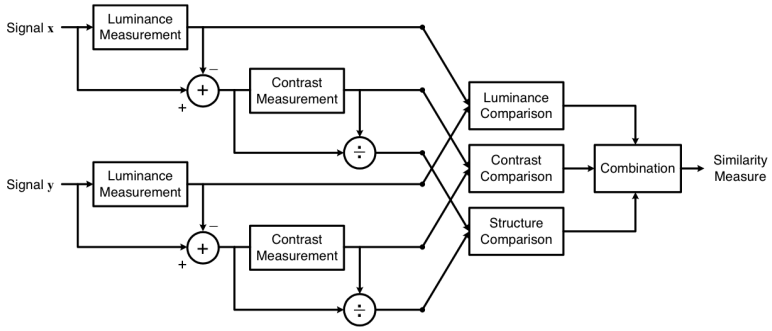


Figure 2: SSIM Flowchart

The system diagram shown in Figure 2 shows the proposed assessment. Signal x and y, where one is the reference (perfect quality), and one is the distorted reproduction. First the luminance of the signals are compared. Assuming discrete signals, the mean intensity can be estimated as follows:

$$\mu_x = \frac{1}{N}\sum_{i=1}^{N} x_i \tag{4.95}$$

The luminance comparison $l(x,y)$ is then a function of μ_x and μ_y

The next step is to extract the mean intensity of the signal, which eliminates the luminance component. On the resulting signal, the standard deviation is used as an estimate of the signal contrast. In discrete form this can be given by

$$\sigma = \left(\frac{1}{N-1}\sum_{i=1}^{N}(x_i - \mu_x^2)\right)^{\frac{1}{2}} \tag{4.96}$$

The contrast $c(x,y)$ is then a comparison of the σ components for the two signals.

Third, the signal is normalized by its own standard deviation, so that both the signals have unit standard deviation. The last step, the structure comparison $s(x,y)$ is then done on the resulting

signal which is now denoted as follows:

$$s(x) = \frac{x - \mu_x}{\sigma_x} \quad \text{and} \quad s(y) = \frac{y - \mu_y}{\sigma_y} \tag{4.97}$$

The last step is then to combine the different components to yield a overall similarity measure

$$S(x,y) = f(l(x,y), c(x,y), s(x,y)) \tag{4.98}$$

The luminance comparison is defined as

$$l(x,y) = \frac{2\mu_x\mu_y + C_1}{\mu_x^2\mu_y^2 + C_1} \tag{4.99}$$

Here C_1 is a constant that is included to avoid instability when the denominator is close to zero. Wang et al. gives the value

$$C_1 = (K_1 L)^2$$

where L is the dynamic range of the image, and $K \ll 1$ is a small constant. If we let R represent the change of relative background luminance which can be written as $\mu y = (1-R)\mu_x$. By substituting this into Equation (4.99) gives

$$l(x,y) = \frac{2(1+R)}{1 + (1+R)^2 + \frac{C_1}{\mu_x^2}} \tag{4.100}$$

which with values of C_1 very small, gives a function that is qualitatively consistent with Weber's Law.

The contrast measure is defined almost the same way:

$$c(x,y) = \frac{2\sigma_x\sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2} \tag{4.101}$$

$C_2 = (K_2 L)^2$ with $K \ll 1$. The important feature here, is that $\Delta\sigma = \sigma_y - \sigma_x$ is less sensitive to high base contrast σ_x than low base contrast. This is the same as seen in the HVS with the contrast-masking feature.

The structure comparison is defined as

$$s(x,y) = \frac{\sigma_{xy} + C_3}{\sigma_x\sigma_y + C_3} \tag{4.102}$$

C_3 is defined as $\frac{C_2}{2}$.

The resulting similarity metric is then combined

$$SSIM(x,y) = [l(x,y)]^\alpha \cdot [c(x,y)]^\beta \cdot [s(x,y)]^\gamma \tag{4.103}$$

α, β and γ are parameters used to adjust the relative importance of each of the components.

This then adds up to following form for the practical use of the SSIM index:

$$SSIM(x,y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_2^x + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \tag{4.104}$$

The heritage from UIQ, presented in Section 4.5.2 is very prominent, and if we set $C_1 = C_2 = 0$ it actually performs the same way although the results may be unstable for certain values.

The local measure for the variables is obtained by using a 8×8 window that slides across the image. Wang et al. proposes the use of a Gaussian weighting function to overcome the problem of blocking effects to the index map. This weighting functions is implemented using the two parameters $K_1 = 0.01; K_2 = 0.03$ Wang et al. claims that the algorithm is fairly insensitive to variations in these values. This local measure can be displayed directly as a quality map. Also, a single numeric value accounting for global error can be computed as the average of all the local measurements.

SSIM has gained a significant popularity in the later years because of its easy-to-understand formulation and low implementation complexity. In the context of HDR imaging, a lot of different experiments has been done, and it is recommended that one transfer the luminance values to a more perceptually uniform space. Also adjustment of the constants $[C_1, C_3]$ and the dynamic range L has to be performed.

HDR-VDP

HDR-VDP is based on S. Dalys [36] original VDP, or Visible Difference Metric. This metric is considered to be one of the leading channel decomposition-based fidelity metric, where both a reference and a distorted image is decomposed and a difference score and map is calculated.

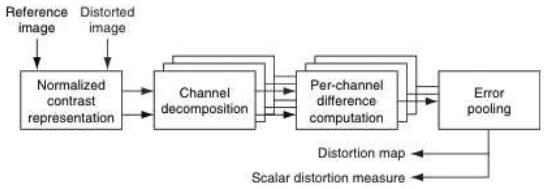


Figure 3: Flowchart of ultichannel perceptual-based fidelity metrics
(Image a courtesy of Tunç Ozan Aydın)

The flow of such a metric is described in Figure 3

Mantiuk et al. [44] proposed an extension to VDP, making it able to correctly determine visible differences between HDR images. The metric takes two images as an input, an reference and a distorted, and produces an output map of pixels differences that is probable that can be perceived.

The first step of the metric is the optical transfer function (OTF), which simulate eye optics. The pixel values are transferred into a retinal response space. This is often referred to as "luminance

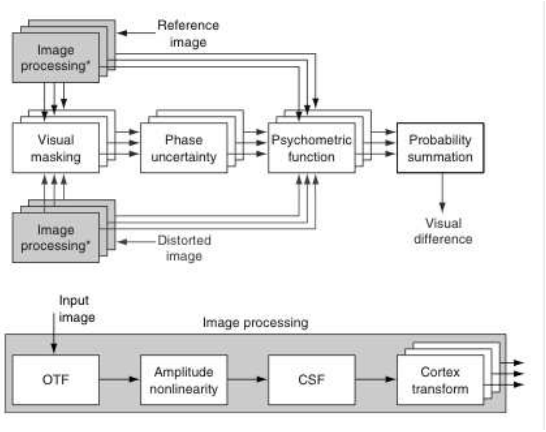


Figure 4: Workflow of VDP. Multiple arrows indicate parallel processing of spatial frequency and orientation channels. (Image courtesy of Rafał Mantiuk)

masking". Mantiuk et al. uses the same OTF function as proposed by Deely et al. [45]:

$$OTF = \exp\left[-\left(\frac{\rho}{20.9 - 2.1d}\right)^{1.3-0.07d}\right] \tag{4.105}$$

where d is the pupil diameter, and ρ is the spatial frequency in cycles per degree(cpd) The pupil diameter is dependent on the global adaptation luminance. The full description of this function is described in detail in Reinhard et al. [4, Chapter 10.7.1]

Amplitude nonlinearity modeling is about modeling the fact that the HVS is not linearly sensitive to contrast. This is modeled using the contrast-versus-intensity function, which is defined as follows to be compatible with the contrast sensitivity function(CSF) used in HDR VDP at a given adaptation level, L_a:

$$\text{CVI}(L_a) = \frac{1}{\max_{\bigvee \rho}[\text{CSF}(\rho, L_a, \ldots)]} \tag{4.106}$$

The CVI function is used to derive luma mappings ($l_{HDR}(L)$) which specifies the retinal response. The CSF function is described in more detail further down.

For a full explanation of all the details, the reader is referred to Reinhard et al. [4, Chapter 10.7.2]

The next step is the CSF function, which is one of the most important characteristics of the HVS in image perception. This is also evident, as many visual appearance models uses the CSF as a key component.

The CSF is investigated many experiments, and even in simple detection experiments, a lot of factors can influence the measured sensitivity. The most important factors that have been determined are the following: Gratings, spatial and temporal frequency, color, background luminance, grating orientation, spatial extent, eccentricity with respect to fovea and observation distance.

For HDR imaging, is is important that the CSF is valid for the full range of luminances that can be seen by the human observer, and Daly has taken this into consideration and defined the CSF as follows:

$$\mathrm{CSF}(\rho,\phi,L_a,i^2,d,\chi) = P \cdot min\left[S_1\left(\frac{\rho}{r_a \cdot r_x \cdot r_\phi}\right) \cdot S_1(\rho)\right] \tag{4.107}$$

where

$$\begin{aligned} r_a &= 0.856 \cdot d^{0.14} \\ r_x &= \frac{1}{1+0.24\chi} \\ r_\phi &= 0.11cos(4\phi)+0.89 \\ S_1(\rho) &= [(3.23(\rho^2 i^2)^{-0.3})5+1]^{-\frac{1}{5}} \cdot A_l \epsilon\rho e^{B_i \epsilon\rho}\sqrt{1+0.06e^{B_i\epsilon\rho}}A_l \qquad = 0.801(1+0.7L_a^{-1})^{-0.2} \\ B_l &= 0.3(1+100L_a^{-1})^{0.15} \end{aligned} \tag{4.108}$$

The parameters are as follows: ρ, spatial frequency in cycles per visual degree; ϕ, orientation; L_a, light adaptation in cd/m^2; i^2, the stimulus size in deg^2; d, distance in meters, χ, eccentricity; ϵ, constant $\epsilon = 0.9$; P, absolute peak sensitivity;

It is important to notify that the functions A_l and B_l has been corrected against the original publication [36].

Before this can be used in the HDR-VDP model, the CSF needs to be correlated to the OTF. This is done by dividing the CSF with the OTF, which results in the need for taking glare into account. This is very important in the extended dynamic range of HDR, and is done by scaling the values to JND units. The transfer to this space also normalizes the values. The normalized JND scaling and the optical light scattering can be excluded from the original CSF with the following formulation:

$$\mathrm{nCSF}(\rho,\phi,L_a,i^2,d,\chi) = \frac{\mathrm{CSF}(\rho,\phi,L_a,i^2,d,\chi)}{\mathrm{OTF}(\rho)} \cdot \mathrm{CVI}(L_a) \tag{4.109}$$

The last step is to normalize the nCSF and use this to modulate the response values obtained from the amplitude nonlinearity step.

The final step in the image processing pipeline, is the cortex transform. Here HDR-VDP follows its predecessor and uses a multichannel technique that models the combined radial frequency and the orientational selectivity of cortical neurons. This is done by decomposing the input image into a pyramid-style representation which is decomposed into six frequency channels by using band-pass filters. These again are filtered into six orientational-selective bands, but we omit the lowest base-bands, and a total of 31 bands are obtained. The actual filtering is done in the Fourier domain, but all other steps require that channels are transferred back to the spatial domain.

Inside the cortex transform, we perform three different filters on each channel. The first is the visual masking, and its computational model is relatively simple. It is referred to as the "threshold elevation model" [36] and is defined as follows:

$$t_e^{k,l}[i,j] = \max\left[1.0, |c_n^{k,l}[i,j]|^s\right] \tag{4.110}$$

The contrast elevation factor, $t_e^{k,l}$ is found for the spatial channel k, orientation channel l and pixel i,j. The value $c_n^{k,l}$ is the detection threshold-scaled contrast, which is a result of the cortex transform. The last parameter s is a factor that controls the contrast elevation, ranging from 0.6 to 1.0. To account for the visual masking in the threshold elevation, we need to normalize the signal by the contrast elevation factor. This results in a more uniform spread of the phase shift uncertainty, and is the second step of the model. It is defined as:

$$e_n^{k,l}[i,j] = \frac{\left|c_{\underline{\underline{r}}}^{k,l}[i,j] - c_{\underline{\underline{d}}}^{k,l}[i,j]\right|}{t_e^{k,l}[i,j]} \tag{4.111}$$

where $e_n^{k,l}$ is the distortion signal, $c_{\underline{\underline{r}}}^{k,l}$ and $c_{\underline{\underline{d}}}^{k,l}$ are the JND-normalized contrast signals for the reference and the distorted image.

The last two steps in the HDR-VDP is used to express the distortion visibility in terms of probabilities within each channel, and then these probabilities are summed up across all channels. This is done in the same way as the original VDP. The psychometric function that describes the increase of detection probability is defined as follows:

$$p^{k,l}[i,j] = 1.0 - \exp^{-\left|\alpha \cdot e_n^{k,l}[i,j]\right|^h} \tag{4.112}$$

where $p^{k,l}$ defines the detection probability, h is the slope of the function, and $\alpha = [-\ln(1-0.75)]^{\frac{1}{h}}$ ensures that the probability equals 0.75 for the contrast signal $e_n^{k,l}[i,j]$ which equals 1 JND.

The summation of these probabilities are then performed using a product series equation defined as:

$$P_t[i,j] = 1 - \prod_{k=1,\ldots,K;l=1,\ldots,L} \left(1 - P_m^{k,l}[i,j]\right) \tag{4.113}$$

where $P_t[i,j]$ is the total probability of distortion detection summed across the all the bands for given pixel i,j.

The output is displayed as a in-context difference map that also can be imposed on the reference as a grating map.

S-CIELAB

S-CIELAB is like ΔE a perceptual color fidelity metric that measures the accurate reproduction of a color. The main idea behind the metric, is that ΔE_{94} does not take into account spatial patterns and color discrimination. Many studies, among them Zhang and Wandell [46] resulted in a spatial extension to CIELAB that corresponds better to perceived colors.

This transform is done by choosing three spatial filters estimated from psychophysical measurements of color appearance. The input image is converted from XYZ to opponent-colors with the following transformation

$$\begin{aligned} O_1 &= 0.279X + 0.72Y + 0.107Z \\ O_2 &= -0.449X + 0.29Y - 0.077X \\ O_3 &= 0.086X - 0.59Y + 0.501Z \end{aligned} \tag{4.114}$$

Then the data in each plane is filtered by a two-dimensional separable spatial kernel in the form

$$f = k \sum_i \omega_i E_i \tag{4.115}$$

where

$$E_i = k_i \exp\left[\frac{-(x^2+y^2)}{\sigma_i^2}\right] \tag{4.116}$$

k_i is a scale factor chosen so that E_i sums to 1, and therefore k is chosen so that foreach color plane the two-dimensional kernel f sums to one.

The parameters ω, σ for the three color planes are a set of weights and spreads(degrees of visual angle).

Total Variation of Difference

Pedersen et al. [47] presents in this yet unpublished article a way to use total variation as a novel image quality metric. This metric is inspired by Total Variation (TV) by Rudin et al. [48] and contrast filtering (CSF) by Peli [49].

The image data is first converted to CIEXYZ color space, and the contrast c for each pixel using Peli's function is calculated to account for variations and frequency bands. This is compared against the contrast sensitivity threshold T for the corresponding channel for each band as L_i:

$$L_i(x,y) = \begin{cases} c(x,y) & if\, c(x,y) > T \\ 0 & else \end{cases} \tag{4.117}$$

The final image L_f is then the sum of all the bands:

$$L_f(x,y) = \sum_{i=1}^{n} L_i(x,y) \tag{4.118}$$

Pedersen et al. argues that since the CIEXYZ color space is not orthogonal, it is necessary to separate the X and Z channel into a color and luminance part filtered with their respective contrast sensitivity function. The luminance bandpass is obtained from the color channel as follows:

$$X_{BL} = \frac{X_L}{Y_L} Y_B \tag{4.119}$$

where X_{BL} is the luminance bandpass information, X_L is the lowpass information in the color channel and Y_L is the lowpass information in the luminance channel.

The color information is found the same way:

$$X_{BC} = X_B - X_{BL} \tag{4.120}$$

where X_{BC} is the color information in the color channel, and X_B is the bandpass information in the same color channel, and the resulting data for both the reference and the reproductions is then converted to the OSA-UCS color space.

The new Total Variation of Difference (TVD) for the original L_O and reproduction L_R is defined as

$$TVD = \sqrt{\sum_j \left(\int |\nabla L_{O_j} - \nabla L_{R_j}| dA \right)^2} + \lambda \int \sqrt{\sum_j (L_{O_j} - L_{R_j})^2} dA \tag{4.121}$$

where $\sqrt{\sum_i (\int |\nabla L_{O_j} - \nabla L_{R_j}| da)^2}$ is the TV term while $\lambda \int \sqrt{\sum_j (L_{O_j} - L_{R_j})^2} dA$ is the color difference term (CD).

This unfortunately results in quite a lot of values, so Pedersen et al. tries to overcome this by using pooling strategies. The TV term is replaced by the standard outer norm(L2) over the color channels with different L1 values, and the CD term the standard outer term is replaced with other statistical functions. Initial results from tests on datasets done by Pedersen et al. indicates that best results compared to perceptual data is obtained when only the TV term, $\lambda = 0$ is considered.

SSIM-ipt

Bonnier et al. presented in [50] a new extension of the classical SSIM algorithm which extends it to color images. Bonnier proposes that this SSIM can be used as an gamut mapping algorithm by first converting both the original and the reproduction of the image to the IPT color space developed by Ebner et al [51]. The IPT color space is constructed to improve the hue uniformity that is missing from CIELAB and CIELUV. The colorspace introduces three channels, I which represents intensity, whereas P and T encode color opponency. The P is named for *protan*, and encodes an approximation of the red-green opponency whereas T, named for *tritan* encodes approximate yello-blue opponency.

The conversion to the IPT color space is achieved by first going to a variant of the LMS space close to the Hunt-Pointer-Estevez cone primaries. These are discussed in Reinhard et al. [52, chapter 10.4.1]. The color space assumes a normalization into CIE D65. It is further discussed that real cone signals cannot generate negative components, this may be done for real colors when converting to IPT. To correct for this, it is important to maintain the proper sign before applying the nonlinear compression.

Bonnier et al. proposes that the usage of a separate SSIM of each channel in the image which is combined using a geometrical mean gives a good correlation in the tested images, but the author recommends further study to draw any conclusions.

Dynamic Range Independent Image Quality Metric

Aydın et al. [53] proposed a metric (DRI) that is independent to the dynamic range of the images that are compared. The metric can be considered a hybrid between contrast detection and SSIM metrics as it includes HVS model data from similar to that found in HDR-VDP, but instead of reporting the contrast changes, analyses these further for structural changes.

The structural changes that Aydın et al. argues are important for are the following:

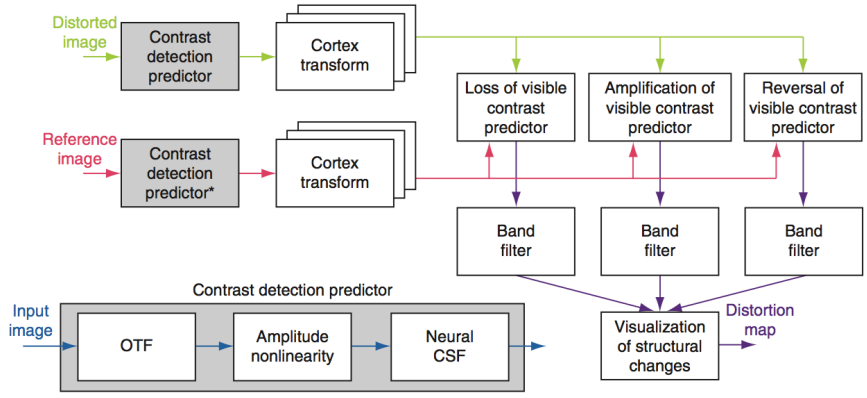


Figure 5: DRI metric workflow. (Figure courtesy of Tunç Ozan Aydın and Rafał Mantiuk)

- Loss of visible contrast
- Amplification of invisible contrast
- Reversal of visible contrast

These are defined as follows

$$\begin{aligned} &\text{loss of visible contrast:} && P^{k,l}_{loss} = P^{k,l}_{r/v} \cdot P^{k,l}_{t/i}, \\ &\text{amplification of invisible contrast:} && P^{k,l}_{ampl} = P^{k,l}_{r/i} \cdot P^{k,l}_{t/v}, \\ &\text{reversal of visible contrast:} && P^{k,l}_{rev} = P^{k,l}_{r/v} \cdot P^{k,l}_{t/v} \cdot R^{k,l} \end{aligned} \tag{4.122}$$

where k and l are the spatial bands and orientation indices, the subscript $r/\cdot$ denotes reference, and $t/\cdot$ denotes test image, the subscript $\cdot/v$ visible and $\cdot/i$ invisible contrast. The parameter R equals 1 if the polarity of contrast in the reference and the test image differ:

$$R^{k,l} = \left[C^{k,l}_r \cdot C^{k,l}_t < 0 \right] \tag{4.123}$$

As shown in Figure 5, each image is separately run through the contrast detection parameter also seen in HDR-VDP, but the cortex transform differs from the standard way. The map from the contrast predictor can contain spatial frequencies that do not belong to a particular subband, which results in noise. To avoid this problem, Aydın et al. proposes to run a Fourier transform on the map with a cortex filter for each component:

$$\hat{B}^{k,l}_{loss} = \mathscr{F}^{-1} \left\{ \mathscr{F}\{P^{k,l}_{loss}\} \cdot B^{k,l} \right\} \tag{4.124}$$

where $B^{k,l}$ is defined as the product of the *dom* and *fan* filter. For an explanation of these, please refer to [53, 17]

As it can be assumed that detection of each distortion in each band is an independent process, the probability that a distortion is detected in any band is given by

$$P_{loss} = 1 - \prod_{k=1}^{N} \prod_{l=1}^{M} \left(1 - \hat{P}_{loss}^{k,l}\right) \tag{4.125}$$

The probability maps for P_{ampl} and P_{rev} is computed the same way.

The distortions found in these maps are then visualized on a in-context map where luminance of the distorted image is copied into all RGB channel. On top of that *green* for loss of visible contrast, *blue* for amplification of invisible contrast and *red* for reversal of visible contrast. If there are overlapping distortions, only the distortion with the highest detection probability will be displayed.

Aydın et al. claim that this metric is able to generate meaningful results for images even if the dynamic range of the images are different. A comparison of a LDR with a HDR reference yields to an overall loss of visible contrast spread throughout the entire image, which is a result of contrast compression. This is also observed in perceptual experiments.

5 Experiment

In this chapter we will present the process of how we will try to answer the research questions given in the introduction.

The first step will be to set up a set of HDR scenes. These scenes will be constructed with different scenes with different complexity and illuminated by different light sources. The setup of the two scenes are described in Section 5.1.

The next step would be to capture HDR images of the given scenes. As HDR cameras are not generally available, a different method is chosen. The process for this is described in Section 5.2.

For viewing HDR images directly a HDR monitor is required. As no such device is accessible, a tone mapping of these are required to be able to view them on a standard monitor. As a lot of different algorithms are available for this, we chose only a set of these. A reproduction of each HDR is produced for all the tone mappers. This process is described in Section 5.3.

For the same reason as stated above, the only way to do a perceptual experiment on the proposed setup, is to do the comparison of the reproductions against the original scene and by this assume that the HDR can be considered a perfect reproduction of the scen. The comparison was done using a paired comparison with two of the reproductions on a calibrated monitor against the given scene. This step is described in Section 5.4.

The comparison of an HDR original against an LDR reproduction is not trivial. For this we developed a framework using perceptual coding to get two images that are perceived the same way. By using this framework, we can transfer the image data between different colorspaces and most image difference metrics can be used without much alterations. This part is described in Section 5.5.

The last step will be to run a set of image difference metrics on the image data which we propose is made comparable through our framework. Here we compare the different HDR images as the original against the tone mapped reproductions of the same scene for evaluated metrics. The different metrics and the settings used to acquire a set of quality scores is described in Section 5.6.

The full procedure described above is summed up in Figure 6.

5.1 Scene Setup

As for the setup of the scene, we used the bunker room at Gjøvik University College (GUC), where the scenes could stand undisturbed during the whole project. The room itself is painted in dark gray, and gives a relative neutral background. The light in the room can be dimmed to a very low setting. This is evident in our measurement of the darker parts of our scene. Both of the scene setups were built using a box with the measurement 45×60×45 cm. These measurements were chosen as this gives us an aspect ratio of the scene of 4:3, which is also the preferred size of images that should be evaluated [54]. The inside the box is clad with cloth to minimize any glare. Furthermore, the box is then divided into two compartments by a center wall which will give us a bright and dark part. Furthermore, we try to block the light so that it only shines on the scene, and not to illuminate the background. In addition we set up a wall blocking any direct lighting from the light source to both

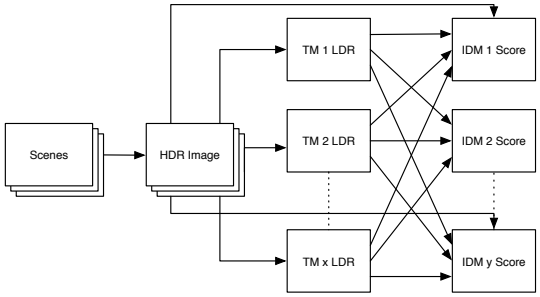


Figure 6: Full overview of experiment workflow

the camera and the the position of the viewer. A sketch of the setup is given in Figure 7. Also a image of the setup during the experiment is available in Figure 10.

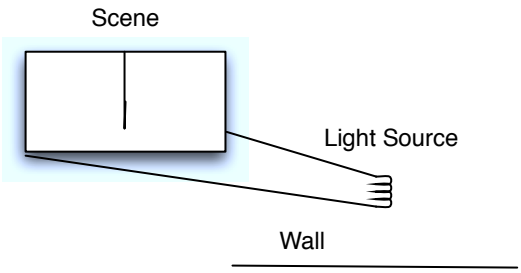


Figure 7: Sketch of scene setup

5.1.1 Simple Scene

In our simple scene, we inserted a set of painted wooden blocks with uniformly colored surfaces. This was chosen to give a set of simple surfaces that could be compared for both the dark and the light part. Colors on the blocks are the same, but their placement and structure vary. In addition we added some balls of crumpled aluminum foil to increase the dynamic range of the scene.

The calculated dynamic range of the final scene, was calculated to $\Delta EV = 13.8224$. This results in a dynamic range of 14486.17 which corresponds to an optical density (OD) of 4.1609. An explanation of these terms and how they were obtained, is given in Chapter 5.2. All other measurements under are taken using a Minolta CS-1000 spectroradiometer.

To illuminate this scene, we used a lamp with two 100 watt Philips incandescent bulbs. These gives off a warm white light which we measured on a white patch as follows:

$$xy = [0.4625, 0.4141] \tag{5.1}$$

The brightest spots on in the scene was determined to be in the reflection of the aluminum ball in the light part of the scene, and this was used to calculate the maximum luminance in the scene. This was measured to

$$Y_{max} = 668.1 \text{ cd/m}^2 \tag{5.2}$$

For the minimum luminance in the scene, we used a spot 5 cm from the center wall in the dark part of the scene as our reference point. The luminance values were here measure to

$$Y_{min} = 0.21 \text{ cd/m}^2 \tag{5.3}$$

All these measures were performed three times and averaged.

Figure 8: Best possible capture with a single exposure of simple scene

A best effort single capture which the camera is able to produce of the given scene is presented in Figure 8.

5.1.2 Advanced Scene

The advanced scene has the same physical measurements as the simple scene, and was divided into the same compartments. The only physical difference in the box is that it is mirrored around the light source as shown in Figure 7.

As for items in the scene, we used a similar set of uniform wooden blocks which was used in the simple scene. In addition two aluminum balls were added, but also some objects that have complex surfaces and textures. Some of these objects werescrews, some pens, two coke bottles and two badges with the Colorlab logo on. The setup were constructed so that all components in the dark part had a similar component in the light part of the scene. To furthermore increase the dynamic range in the scene, we placed a mirror which reflected the light source into a lens which gives us a focused glare spot. This was aimed to hit the camera and the observer. Lastly, we added a small sphere which was polished to almost perfect reflection which we used as a visual reference that the glare did not get overexposed to much in the capture process.

To illuminate the scene, we used a studio lamp with a 500 watt halogen lamp. Compared to the incandescent bulbs used on the simple scenes, the light has a much more bluish tint to it which is also evident in the images.

The light was measured on the same neutral patch as the simple scene. The measurements from this were the following:

$$xy = [0.4339, 0, 4750] \tag{5.4}$$

As for the maximal luminance in the scene, this was determined to be in the focused glare on the lens. When we tried to measure this with the CS-1000, this was over the range that could be measured, which indicated that the luminance was above 10 000 which is the upper measurable limit.

To overcome this limitation, we acquired a neutral density (ND) filter that could be fitted on top of the focusing lens of the spectroradiometer. The used filter was a Fader ND adjustable ND filter from Light Craft Workshop [55]. This was adjusted to it's minimum setting, and we measured the white patch to get how much light was let through:

$$Y_{filter} = 460.2 \tag{5.5}$$

We then removed the filter and measured the same white patch

$$Y_{noFilter} = 1481.77 \tag{5.6}$$

which gives us a blocking factor of 3.2196, which is around $\frac{1}{4}$ which is corresponds to a static ND4.

We then attached the ND filter back on, and used the same setting to measure the glare in the lens:

$$Y_{maxFilter} = 6356 \tag{5.7}$$

and multiplied with the blocking factor this results in the following

$$Y_{max} = 20463.78 \tag{5.8}$$

This high dynamic range was also observed in the capturing process, as the value $\Delta EV = 16.3558$ which corresponds to the dynamic range of 83863.8438. [1] This value corresponds to an OD value of 4.9236

A best effort single capture which the camera is able to produce of the scene is presented in Figure 9.

5.2 Capture of Images

When capturing our scenes, we used a Nikon D90 camera with a Sigma 18-250mm F3.5-6.3 DC OS HSM lens attached. As noted in Section 5.1.1, the pictures should be reproduced with a size of 9-15

[1] These values are the final values after ghost removal. See discussion in 5.2

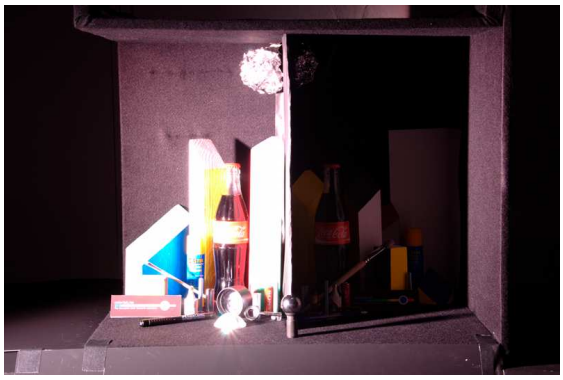


Figure 9: Best possible capture with a single exposure of advanced scene

cm × 12-20 cm, which gives us the aspect ratio of 4:3, the same size as the scene.

To accomplish this during the capture process, we placed the camera on a stable tripod 180 cm from the scene. The camera was then placed in a height of 110 cm as this was measured to be the height of the eyes of a normal person which is seated on a rigid and sturdy chair which was used during the perceptual experiment.

For each of the scenes, the whitepoint of the scene was measured using the cameras internal whitepoint meter, and a profile was created. These profiles were saved, and used for all the captures of the given scene. The colorspace for the camera was set to sRGB.

The lens was then zoomed in to a setting of 50 mm which covered most of the scene, the aperture were set to f16 which results in that we use a relative small part of the center of the lens. This gives us higher depth clarity which ensures that the entire scene as well as any part of the background are as clear as possible. An ISO setting of 200 were used to produce the lowest possible amount of noise in the image but still be able to keep the exposure times low.

Image data were stored in Nikon NEF format. This gives us 12 bit raw data from the sensor, but also all settings used is stored inside the file and is used when decoding the image. All settings on the camera were turned to manual, and any artificial enhancement such as DLighting, were turned off. The capture resolution was set to the maximum possible on the camera, 4288 × 2848 which corresponds to around 12 Mpixels.

The actual capture process were done by selecting a 2 second delay of the shutter after pushing the release button on the camera, as no remote control were available. This makes sure that any movement of the camera were as small as possible for each capture.

Only the exposure time were altered during each capture of the scenes. It is discussed how many shots that should be used, but we tried to use settings so that we had about 1 picture per ΔEV which is recommended by the stitching software used. This resulted in a total of 15 shots for the advanced scene, and 10 for the simple scene. The number of samples on the simple scene was chosen as larger

numbers of samples resulted in renderings with much less dynamic range.

The timings and calculated ΔEV are presented in Table 2 for the simple scene, and Table 3 for the advanced scene.

Table 2: Simple Scene; exposure times and exposure values

Exposure time in seconds	EV
30	+1.55 EV
10	-0.03 EV
5	-1.03 EV
2.5	-2.36 EV
1	-3.36 EV
0.25	-4.36 EV
0.2	-5.68 EV
0.05	-7.68 EV
0.025	-8.39 EV
0.01	-10.00 EV

Table 3: Advanced Scene; Exposure time and exposure values

Exposure time in seconds	EV
20	+4.29 EV
10	+3.29 EV
5	+2.29 EV
2.5	+1.29 EV
1	-0.03 EV
0.5	-1.03 EV
0.25	-2.03 EV
0.125	-3.03 EV
0.05	-4.36 EV
0.025	-5.36 EV
0.0125	-6.36 EV
0.00625	-7.36 EV
0.003125	-8.36 EV
0.002	-9.00 EV
0.001	-10.00 EV

These values were calculated using Luminance [56] and Picturenaut [57].

The actual stitching process described in Section 4.1, is implemented in the program Picturenaut. This program is a frontend to the CLI utility MKHDRI which gives the user easy access to both generating the response curve of the camera as well as doing both image alignment and ghost removal. Also color balancing can be inserted. In our setup we used the following settings: Exposure

correction, automatic image alignment and ghost removal. All other settings were set to their default settings. In the case of color balancing this turns off any color processing. Exposure correction means that the program uses the EXIF data from the different image exposures, and uses these to produce a correct gamma correction for each image.

It is worth noticing that even though movement of the camera were kept to a absolute minimum, some movement is unavoidable when touching the camera between each exposure. Also, with the ghost removal setting turned off, the dynamic range of the image dropped significantly on the advanced scene when processing the same image set. Why this is the case is unknown as the estimated dynamic range before this step is the same, but it may be that the strong reflection could be the problem. To keep both the scenes the same and have the maximal dynamic range, this setting was therefore left on.

The resulting image was then stored using the OpenEXR format without compression, which gives us the most compatible format as different revisions of the reader/writer for OpenEXR exists.

5.3 Tone Mapping

For doing the tone mapping of the image captured of the two scenes, we used the functions bundled with the book High Dynamic Range Imaging, [4] on its companion DVD in addition to the tool Luminance 2.0.2, or previously QTPFSGUI [56]

As we had to limit the number of algorithms due to the length of the psychophysical experiment, we decided to use five different local algorithms. We omitted both clipping and global algorithms due to the fact that the reproductions from these would have been considered as outliers, and this would affect the outcome of the psychophysical experiment.

After some initial testing and based on previous studies like Kuang et al [58], we chose to use the following algorithms:

- Fast Bilateral Filtering for the Display of HDR Images by Durand
- Gradient Domain High Dynamic Range Compression by Fattal
- A Perceptual Framework for Contrast Processing of High Dynamic Range Images by Mantiuk
- Photographic Tone Reproduction for Digital Images by Reinhard
- STRESS: A Framework for Spatial Color Algorithms by Kolås

These first four were chosen as they gave the subjective best reproductions of the given scenes in the eyes of the author and had been proved to given good reproductions in earlier research. These TMOs are considered state of the art metrics in the field of tone mapping. The last algorithm STRESS were chosen to test this on a real world scene, and to check how this would score against established tone mappers in a real world experiment as this has not yet been done.

As the both the images of the scene should be whitepoint corrected by the camera, we therefore

use the same parameter settings for the algorithm on both the scenes. The settings found is based on subjective measures of evaluation of the author of this thesis, and they may be improved, especially for the simple scene.

All output from Luminance were output in the original resolution of the HDR input. No processing were performed. The output from the tone mapping were stored in the PNG format which is lossless and supported by the viewing software. After the tone mapping were performed, we resized the images in Matlab using bicubic interpolation to the correct size to be shown in the experiment. A image size of 850×565 pixels were chosen as this was the biggest possible to be shown on the monitor when using the software to perform the perceptual experiment.

Output of the images are presented in Appendix A, Figure 19 to Figure 23 for the simple scene, and Figure 24 to Figure 28 for the advanced scene.

Durand Fast Bilateral Filtering

The operator proposed by Durand is very fast, and even our large HDR images were processed in very short time. The details in an image is a little bit blurred, a, but the overall photorealistic appearance looks to be quite good.

The implementation in Luminance, this operator has tree main parameters: Spatial kernel size, range kernel sigma and base contrast.

The spatial kernel size is connected to the size of the kernel of the bilateral filter, the range kernel sigma is the influence of the intensity. The last parameter defines the level of the base layer.

The parameters used are presented in Table 4.

Table 4: Parameters used for Durand Fast Bilateral Filtering tone mapper

Parameter	Settings
Spatial Kernel Sigma	75.0
Range Kernel Sigma	7.30
Base Contrast	3.5

A output of the tonemapper may be seen in Figure 19 for the simple scene, and Figure 24 for the advanced scene available in Appendix A

Fattal Gradient Domain

Fattal et al. operator has two different input parameters α and β which controls the tone mapper as described in section 4.2.2. These two are implemented in Luminance with scales from [0-2]. In addition a parameter for color saturation and noise reduction with ranges [0-1] is implemented. The color saturation is mapped to the exponent s described in Equation (4.54). The noise reduction is not described in the original paper, but test indicate that the parameter is included to combat any noise that is included by the gradient compression set by the β parameter.

The parameters used are presented in Table 5.

A output of the tonemapper may be seen in Figure 20 for the simple scene, and Figure 25 for the advanced scene available in Appendix A.

Table 5: Settings used for Fattal Gradient Domain tone mapper

Parameter	Settings
α	0.5
β	0.945
Color Saturation	0.275
Noise Reduction	0.5

Mantiuk Framework for Contrast Processing

Mantiuk et al. operator has a very intuitive set of intuitive variables in Luminance. These are a contrast factor, a saturation factor and a detail factor.

The contrast parameter controls the contrast reduction factor l in Equation (4.75), and the saturation operator controls the parameter s in Equation (4.79).

The last parameter called detail factor is not clearly mapped to any of the functions, and we have not looked into the code to see the actual implementation, but this is likely connected to the coefficient $p^k_{i,j}$ in Equation (4.76) which controls the weighting factor for a mismatching contrast.

There is also another parameter called contrast equalization, but this is not used as this equalizes the contrast for the whole image and introduced a lot of artifacts in our tests.

For Mantiuk et al.'s operator, the parameters used are presented in Table 6.

Table 6: Settings used for Mantiuk Framework for Contrast Processing tone mapper

Parameter	Settings
Contrast Factor	0.45
Saturation Factor	0.75
Detail Factor	1.0

A output of the tonemapper may be seen in Figure 21 for the simple scene, and Figure 26 for the advanced scene available in Appendix A.

Reinhard Photographic Operator

Reinhard et al.'s operator is also implemented in Luminance, with two different methods for setting the range for an image. One allows you to set the range and a lower and upper scale for the dodge-and-burn settings, or you can follow the automatic method proposed to calculate these values. In our test, the automatic method produced very good results, so these were chosen instead of tweaking these values manually.

In addition to these, there are two main variables, the key value and the phi. The key is mapped to the key k found in Equation (4.65). The second parameter phi is mapped to the Φ factor in Equation (4.69).

In our experiment, we used the parameters given in Table 7.

This gave the best representation for both the simple and the advanced scene, and the output

Table 7: Settings used for Reinhard Photographic Operator tone mapper

Parameter	Settings
Key Value	0.1
Phi	5.0

may be seen in Figure 22 for the simple scene, and Figure 27 for the advanced scene available in Appendix A.

Kolås STRESS

Kolås et al.'s algorithm is not implemented in any graphical tool yet, and was generated using the original implementation of the tone mapper with only setting the number of iterations each sample point is evaluated and the input luminance value to make the image fit better to the perceived scene. All other parameters were set to their standard values.

The settings used is shown in Table 8:

Table 8: Parameters used for Kolås STRESS tone mapper

Parameter	Settings
Iterations	400
Luminous normalization	8.0

These settings gave us the best rendering of the scene, and is shown in Figure 23 for the simple scene and Figure 23 for the advanced scene, Appendix A. It should be noted that the equalization of the image data, which is a trademark of this TMO, is clearly visible.

5.4 Perceptual Experiment

The perceptual experiment was conducted in the room described in Section 5.1. A setup of the full scene is given in Figure 10.

In the perceptual experiment, we decided to check for the following attributes from Figure 1:

- Overall reproduction
- Color reproduction
- Detail reproduction
- Artifacts

To view the images, a Dell UltraSharp 2407WFP 24 inch monitor with has a resolution of 1920 x 1080 was used. This monitor is based on a S-VPA panel that offers very good color reproduction and a high contrast ratio.

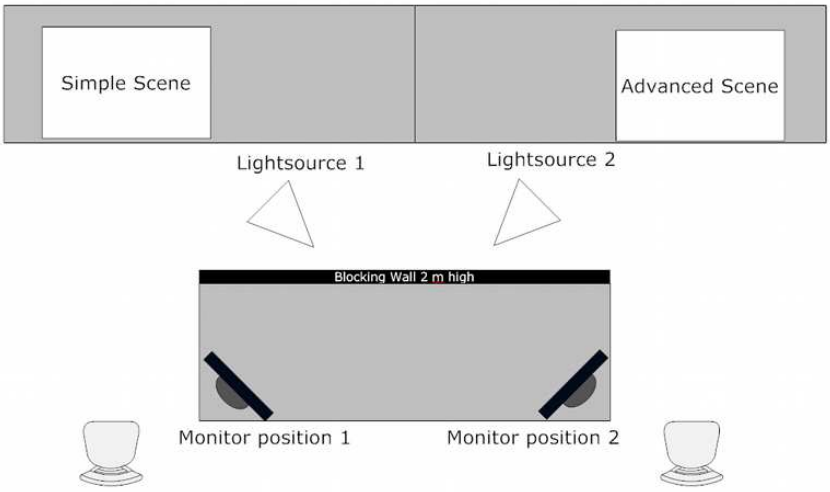


Figure 10: Sketch of perceptual experiment setup

However, this panel can show some color shifts due to viewing angle on the monitor, so the monitor were always aligned to a 90 degree angle to the participants eye to minimize these shifts.

The monitor was then calibrated to a whitepoint of D65 using a EyeOne Match 3 with the software version 3.6.3 and the luminance were set to as close a value to 110 which is the preferred value for viewing a LCD monitor. The calibration data were stored in a ICC profile which was applied for the monitor.

For comparison, we set up a paired comparison using the software package QuickEval 1.0 developed at GUC by Malakauskas and Montvilas [59] for the three first attributes. This software displays all possible combination of the images. As case V of Thurstones law (4.80) requires that the data are normally distributed and that the variance is equal between the reproductions, which is not the case. Other models have been developed for other types of distributions, but later publications such as Andrich [60] shows that these newer models yields the same results as the Thurstone scale with case V when applied in a empirical context. Thurstones law is therefore accepted to ease the comparison between two reproductions and an original.

The software also adheres to the viewing conditions set up in CIE article 18 on viewing conditions [61]. The border around the monitor was black, and the black wall present in Figure 7 eliminates any glare from the light sources used to light the scene. To further reduce surrounding light in the room, the light in the room was dimmed, and only the light source directed the evaluated scene were fully lit during the experiment.

To store the recorded data for each attribute, QuickEval creates a summary matrix after each subject has finished the experiment. From this a proportional matrix as well as a z-score matrix is calculated. When the next subject, the data from the previous participants is added and new matrices are calculated. The matrices for each person is stored to track changes.

The final set of matrices for the different scenes are presented in Appendix C.

All of the subjects that participated in the experiment were given a set of written instructions of

each of the different attributes in addition of a short introduction if any of the written notes were questioned. The written text given to each participant is given in Appendix B. In this document the order of the test is also given. The experiment were started with the overall attribute so that it would not be affected by earlier comparisons. To also try to minimize the influence of previous comparisons, the test subjects were asked to take a small break and walk away from the setup between each part of the experiment. When the simple part was finished, the monitor was moved from the left position where the simple scene is set up, to the right position to evaluate the advanced scene. The position of all the movable objects such as the chairs, the table and the lightsources were marked up so any unintentional movement could be corrected, and to ensure a equal viewing conditions for all participants.

The appearance of artifacts were collected by that each test subject got one sheet of paper for each tone mapper with a black and white representation of the scene, and were asked to compare each of the tone mapped reproductions against the scene. The scenes were presented in fullscreen one at a time with the same viewing conditions as used by QuickEval. As the result from such a experiment varied a lot, only a overall comparison of the number of details at certain areas that shows problems were taken into account. These are discussed in Section 6.1.

5.5 Encoding Framework

As described in the introduction, to compare HDR and LDR image is no trivial task. In Figure 11 we propose the following conversion that enables us to do use most of the existing image difference metrics that exists by the usage of PU coding, and it can easily be extended if other colorspaces need to be used.

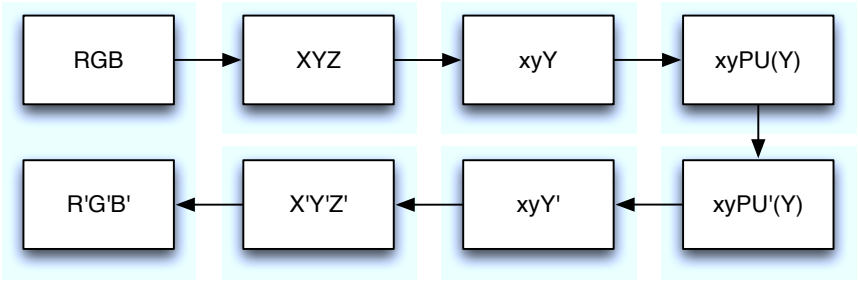


Figure 11: Forward and backward framework for PU coding luminance

The first step will be to move the image data which in most cases are in RGB to XYZ coordinates. In this step we take the whitepoint and any viewing data in the form of the calculated ICC profile of the monitor into account.

In our experiment, we compare a real scene against a calibrated monitor, so it is not possible to get an correct profile for both of them. As noted in the capturing process, the images that builds up the HDR is calibrated to the D65 whitepoint of the scene and as the image monitor is also set up to a whitepoint of D65, we can presume that the two sources have an equal whitepoint. Therefore a

standard matrix for transforming sRGB D65 to XYZ values were used on the HDR.

From XYZ we can easily transfer the data into the xyY color space. By extracting the Y component into a single channel for both the original and the reproduction, we can derive the PU representation of both images.

Even if they are perceptually equal, the ranges of the HDR and the LDR reproduction are not the same. We therefore pursue two different methods of normalizing these: Multiplication by a factor and addition by a factor.

We chose to pursuit both of these to check for different types of scaling of the luminance would affect the result from the image quality metrics.

The multiplication with a factor will do a scaling of the image it is applied to. We choose to do all scaling or manipulation on the LDR because the HDR is considered a perfect reproduction of the perceived scene, and should therefore not be altered.

The PU$'(L)*$ is defined as follows:

$$f = \frac{\max\left(\mathrm{PU}_{HDR}\right)}{\max\left(\mathrm{PU}_{LDR}\right)}$$
$$\mathrm{PU}'(L)_{LDR}* = f \cdot \mathrm{PU}_{LDR} \tag{5.9}$$

where PU_{HDR} is the result of the perceptual coding of the original HDR image, PU_{LDR} is the perceptual coding of the reproduction LDR image, and PU$'(L)$ is the resulting encoding of the LDR data so that it is perceptually equal to the HDR data.

The resulting LDR data is then combined with the chromaticity coordinates from the original LDR image. This results in a vertical shift in the chromaticity.

For the addition we follow the same way, but instead of doing the scaling, we only move $\max(\mathrm{PU}_{LDR})$ up to the same level as the $\max(\mathrm{PU}_{HDR})$. To find this factor we use the following formula for calculating PU'(L):

$$f = \max(\mathrm{PU}_{HDR}) - \max(\mathrm{PU}_{LDR})$$
$$\mathrm{PU}'(L)_{LDR} += f + \mathrm{PU}_{LDR} \tag{5.10}$$

The same vertical shift as observed for PU$'L*$ will also occur here.

The next step will then to invert PU$'L$ to Y-values for the LDR, and replace the Y-channel for the LDR with the calculated data. This can then be transferred back to the XYZ colorspace or processed directly with SSIM or other metrics that only take the luminance information into account.

5.6 Image Comparison

In the field of IDM's, there is a wide variety of algorithms designed to check for different types of image attributes. We have taken a small selection from the very broad survey of current state of the art image difference metrics available done by Pedersen and Hardeberg, [62], but have also included some new metrics that have been developed at GUC.

The different metrics have all been run both through our proposed framework and using direct comparison by standard normalization to the maximal luminance in the HDR.

SSIM

The SSIM metric gives us a very simple and understandable range of quality from [-1,1], where -1 is a total reversal and 1 is a perfect reproduction. We have chosen to use this as our baseline for comparison as it is well tested and easy to understand. The implementation is available in form of a set of Matlab scripts, so all processing of the results were done in Matlab.

We have altered the metric so that only the luminance part of the input images are used. In addition we normalize these values to the maximal value of the HDR original so the range of the luminance is [0,1]. From this we set the parameter L which corresponds to the dynamic range to 1.

For the other parameters, we use the recommended values as defined in the implementation by Wang. This leads that all the different features described in Section 4.5.2 is weighted equally.

The implementation by Wang outputs a single score as well as a map of the image differences, but in our tests we will only using the score as a measure for our comparison.

The resulting scores for each reproduction is available in Appendix D, Table 22.

SSIM_IPT

SSIM_IPT gives us a very simple extension of SSIM to color images with only minor alterations to the original code. It has not showed very good correlations in earlier test, but was used as it is simple to understand and easy to use. The source code for the metric is also available in the form of Matlab source files and all processing was therefore done in Matlab.

The metric SSIM_IPT needs the images that are to be evaluated to be in the IPT color space. To accomplish this, we transferred the resulting XYZ values to LMS cone responses and from there to IPT color space by usage of the transformation functions available from Reinhard et al. [52].

A problem with this colorspace is that the dynamic range of each of the components are not equal for all channels. The I-channel has a range from [0,1], whereas the P and T chromatic channels have a range from [-1,1]. To account for this, the data was normalized to this range for the both HDR original and all LDR reproductions. Each channel is scored using a standard SSIM metric, and then summed using geometrical mean.

It should also be noted that due to problems with conversion from RGB to XYZ, pixels with the value [0,0,0] in standard range [0-255] for RGB which corresponds to pure black, were set to [1,1,1] because a blackpoint calibration was not obtained for the monitor. This should not affect the results as this alteration should not be noticeable in the viewing conditions used.

The output from the metric is four different parameters: geometrical score, geometrical min, geometrical median and geometrical variance. In our tests we have only used the geometrical score as this is directly comparable against a standard SSIM score.

The raw output from the metric is available in Appendix D, Table 23.

S-CIELAB

S-CIELAB is calculating the color difference between two images. The implementation we used comes from X. Zhang [63] which is implemented in Matlab.

S-CIELAB expects input in CIE XYZ coordinates, so as an input here, we use our framework to produce X'Y'Z' outputs for all LDR images as well as an XYZ representation for the HDR original. The same correction to original RGB LDR values as noted in SSIM_IPT Section 5.6 was done to overcome the problems with the blackpoint.

The S-CIELAB outputs a map with a distance score for each pixel. Therefore we calculated a average and mean score to get a single value representing the quality for each LDR reproduction.

The raw output from the metric is available in Appendix D, Table 24.

TVD

The TVD metric is not yet published, but is included as it is newly developed at GUC, and that it is expected to give good measures on image details even if the dynamic range is different. The implementation currently available in a set of Matlab scripts.

The metric expects the input to be in XYZ coordinates, so the same method as for S-CIELAB to produce X'Y'Z' outputs for all LDR reproductions to compare against the HDR original. In addition the metrics expects four different parameters, λ, nrm, TVopt and PixelPerDegree. The parameter λ is discussed in Section 4.5.2, Equation 4.121. We calculate the TVD index for two different values of λ, 0 and 1, to check which of the different components of the metric which corresponds best to the data from the perceptual experiment. The nrm and TVopt parameters were left to default 1 and PixelPerDegree were calculated using the same values as used in S-CIELAB.

Output from the TVD metric is a single score for the combine TV and Fidelity term.

The output from the metric with the different settings described in Appendix D, Table 25.

HDR-VDP

HDR-VDP is based on PFStools [64], which is a software package especially designed for reading, writing and manipulating HDR images and video frames. The VDP script was compiled on a Ubuntu linux system using version 1.8.1 of the HDR-VDP source code against PFStools version 1.7.

The HDR-VDP script has a wide variety of inputs, but only the default settings were used in our setup. The format of the input images needs to be the same, so for all the LDR reproductions, the original output were converted to OpenEXR.

The output from the metrics is a map of the differences. In addition a numerical value for a detection threshold of 75% and 95% is output. In our tests we only use these scores as measures.

The resulting scores for the different scenes is available in Appendix D, Table 26.

DRI

The DRI metric is only implemented online on the webpage of the project [65]. It is available both for images and for video, but we have only used the image part.

The metric is implemented in a wizard consisting of three steps. The first step is to upload a original and a reproduction. It is also possible to run different types of metrics on this page, but in our tests, only the DRI-metric was selected. The second step has the photometric calibration for the LDR. Here the setting for typical LCD with a maximal luminance of 110 and a gamma of 2.2 as this matches our calibration data for the monitor. The third step is the viewing condition, and as measured the viewing distance was set to 0.3 meters which results in a pixel per visual degree of 19.4493, the same as used for S-CIELAB. The peak contrast were set to its default value.

The output from the metric consists of a map where the RGB channels are mapped to different distortions as described in Section 4.5.2. The R-channel is mapped to the reversal of contrast, G-channel to the loss of contrast and B-channel to amplification of contrast. Only errors above 75% detection probability is taken into account. An example of the mapping is available in Figure 12.

The map is presented as four layers, one for each RGB-channel, and one for the gray-level background. We therefore extracted the different errors and summed them as a standard mean. Formally this can be expressed as

$$E = \frac{\text{sum}(R) + \text{sum}(G) + \text{sum}(G)}{3 * h * w} \tag{5.11}$$

where R, G, B is the channels with the different error maps, h is the height of the image and w is the width of the image.

This score is used as a difference score the same way as HDR-VDP.

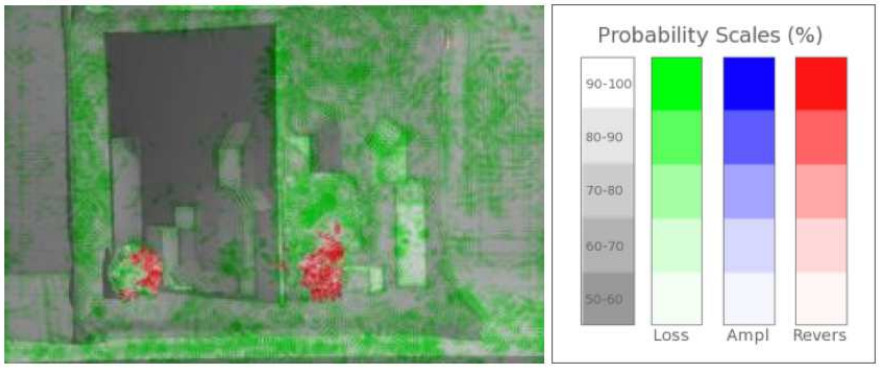


Figure 12: DRI metric channel mapping

The resulting scores for each scene is available in Appendix D, Table 27.

6 Results

In this chapter, the results from the different parts done in the experimental setup will be given.

First, the data from the perceptual experiment is given. Here the different quality attributes are compared to the tone mappers. Also a visualization using a errorbar diagram is presented.

The second part shows the output from the different IDMs chosen to be tested for using the different normalization factors of our framework and using direct comparison.

Last, we present a set of different correlation scores where we compare the different parts against each other. These scores give us an indication of dependency between the different attributes and the tone mappers. Secondly, a comparison of what is important in the overall quality is presented. Last, a comparison of the different metrics are done, and how these compare against each other.

6.1 Results from Perceptual Experiment

From the perceptual experiment, we get a summation matrix for how many times a certain scene is chosen over another. These are then summed up to compute a proportional matrix. From these, a standard Z-score matrix is calculated with a 95 % confidence interval for all the different scenes. The different matrices from the simple scene is shown in Appendix C.

The Z-scores averaged to get a single score for each scene. The resulting scores for the simple scene are as follows:

Table 9: Simple scene, average Z-scores

	Durand	Fattal	Mantiuk	Reinhard	Kolås
Overall quality:	0.4101	-0.5382	-0.2216	0.2970	0.0527
Detail quality:	0.1450	0.3864	0.1067	-0.7494	0.1112
Color quality:	0.3087	-1.2450	0.4088	0.5311	-0.0035

Table 10: Advanced scene, average Z-scores

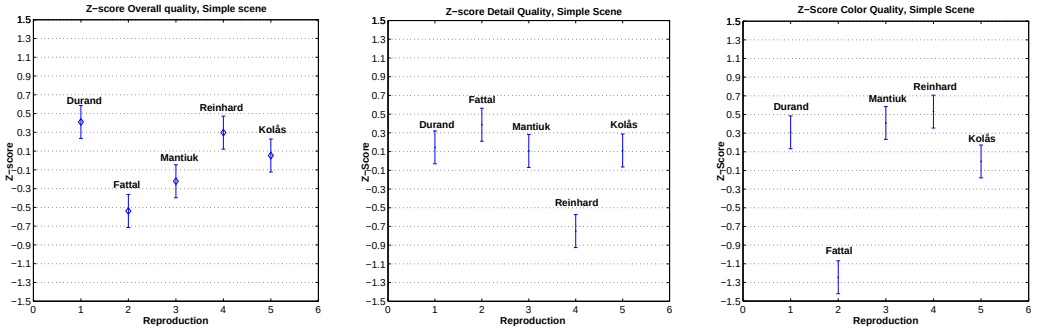
	Durand	Fattal	Mantiuk	Reinhard	Kolås
Overall quality:	0.8835	-0.0755	0.2397	-0.9769	-0.0708
Detail quality:	0.9152	0.5557	0.0291	-0.8716	-0.6284
Color quality:	0.8677	-0.6695	0.6256	-0.6222	-0.2016

These scores can be visualized using a errorbar diagram. To calculate the 95% confidence intervals, we have used the formula proposed by E. D. Montag [66] which gives more correct deviances for experiments based on paired comparisons. As we have 17 participants, and 5 reproductions that

are viewed which results in the following confidence interval (CI):

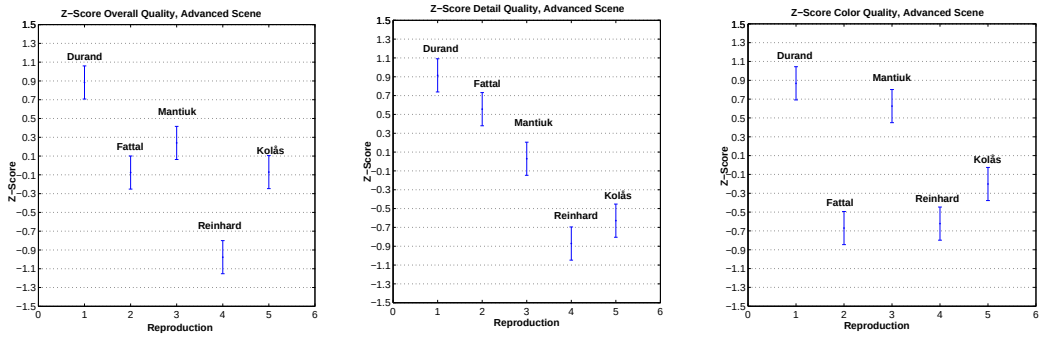
$$CI = \pm 1.96 \times 0.0899 \approx 0.1762 \tag{6.1}$$

For the simple scene, the errorbar plots of the different parameters is shown in Figure 13, and Figure 14 for the advanced scene.



(a) Overall Quality (b) Detail Quality (c) Color Quality

Figure 13: Errorbars, simple scene



(a) Overall Quality (b) Detail Quality (c) Color Quality

Figure 14: Errorbars, advanced cene

For the simple scene, the overall quality seems to overlap for almost all the TMOs. Fattal and Mantiuk are scored as the two worst, Reinhard and Durand are scored as the two best, and Kolås overlaps both the two best as well as Mantiuk. For the detail quality, all the different tone mappers except Reinhard overlap. For color, Durand, Mantiuk and Reinhard are scored about the same level. Kolås becomes second, and Fattal is scored as much worse.

For the advanced scene, the settings are a little bit different. The overall quality there is a clear winner in Durand, Fattal, Mantiuk and Kolås is regarded as overlapping on the middle, and Reinhard

becomes last. For the detail attribute, Durand once again is scored the best with a little overlapping with Fattal. Mantiuk comes at third place, and Kolås and Reinhard overlaps at the bottom. For color Durand once again comes out on top with a overlapping of Mantiuk. Quite a bit under these comes Kolås, with a overlapping between Reinhard and Fattal on the bottom.

From this it looks like the best score is the TMO from Durand as it's in the uppermost part of all the test. For the simple scene the Reinhard is scored much better than in the advanced scene, but this could be to the problems with the whitepoint of the simple scene where Durands colorful representation seems to perform best of all TMOs. This finding is also found by Kuang et al. [58], but in their test Reinhards photographic tone operator performed better than what is seen in our tests where this operator is scored as the worst for the advanced scene.

For the artifacts, the results are not easy to quantize, so we have gathered certain areas of each scenes where problems are detected into groups. For the simple scene, these are summed up in Table 11. The same for the advanced scene is summed up in Table 12.

Table 11: Artifacts, simple scene

Problem area	Durand	Fattal	Mantiuk	Reinhard	Kolås
Dark side, details	3	12	3	2	4
Light side, details	8	9	5	9	11
Left glareball	6	4	5	7	5
Right glareball	7	5	5	7	4
Sum	24	30	18	25	24

Table 12: Artifacts, advanced scene

Problem area	Durand	Fattal	Mantiuk	Reinhard	Kolås
Dark side, details	2	9	8	5	6
Light side, detains	6	8	0	14	14
Glare Lens	6	5	5	10	5
Cola Light part	0	0	2	9	0
Glare ball, top	0	0	3	0	3
Sum	14	22	18	38	28

As seen from these results, the fewest artifacts are found for Mantiuk for the simple scene, and Durand is the best for the advanced scene.

6.2 Results from Image Quality Metrics

In this section we present the output from the different settings from the image quality metrics used.

SSIM

As a perfect reproduction would yield a score of 1, direct inspection of the results seems that the tone mapping process has resulted in some structural changes in the images. Initial inspections shows that the values for the simple scene using our framework lies in the range of 0.44 - 0.57 for the multiplication factor, and 0.35 - 0.43 for the addition factor.

For the advanced scene, we see the same tendencies, but the scores are somewhat higher; multiplication ranges from 0.59 - 0.78, and 0.39 - 0.45 for the addition.

Direct comparison scores for multiplication is in the range 0.10 - 0.29 for the simple scene, 0.10 - 0.48 for the advanced scene.

These initial numbers seems to indicate that running the metric through our framework yields a more correct score than when doing a direct comparison as the ranges are smaller and more consistent. These findings corresponds to earlier findings by Aydın et al. [35].

The output from the SSIM metric for both of the scenes are found in Table 22.

SSIM-IPT

SSIM-IPT has the same scoring range as standard SSIM as the same algorithm is used to generate the score for each channel. Therefore we can directly compare the scores between these two metrics.

The initial output from this metric shows that although the scores are quite low, the simple scene gets a higher score for most of the tone mappers for both multiplication and addition normalization. The scores also vary very little, so initially this metric seems not to be able to produce believable scores.

The output from the SSIM-IPT metric for both scenes can be found in Table 23.

S-CIELAB

The metric S-CIELAB calculates a color difference map. This makes it not directly comparable to the above metrics. This is because SSIM we have an increasing number to give better quality, but metrics that outputs difference maps an increase in number means a bigger difference. In this case a score of 0 means a perfect reproduction.

Scores from the S-CIELAB seems to vary quite a bit. Using our framework with the multiplication factor shows an mean error from 0.75 - 3.06, the median 0.03 - 1.48. Using the addition factor with the framework shows even greater differences, mean ranging from 3.28 - 17.39. For the median we seem the same, a range of 3.02 - 15.28. From these results it is indicated that the addition factor introduces more errors than the multiplication which is to be expected.

Compared to the direct comparison, we see that for the multiplication the range of errors for the different scenes range from 0.64 to 6.88 for the mean, and for the median the range is 0.45 - 2.34.

From these two comparisons, it looks like the framework predicts a quality for both the mean and median measures for both scenes as the error difference is lower.

The full output of the S-CIELAB metric for both scenes is available in Table 24.

TVD

Output from the TVD metric is an error difference map, and therefore the same things applies to S-CIELAB also applies to TVD. Lower error rates means higher quality. Also the scores may vary quite a bit when $\lambda = 1$ because the fidelity part of the metric often can output high numbers.

For $\lambda = 1$ where this is the case, we see a range of 28.50 - 59.14 for the simple scene and 19.26 - 65.50 for the advanced scene using multiplication with our framework. With addition the scores for the simple scene lies between 59.14 - 65.61, the advanced 65.50 - 70.29.

With $\lambda = 0$ where only the TV term is considered the scores are quite a bit lower. The multiplication for the simple scene range from 0.99 - 1.63 and the advanced scene from 0.76 - 1.91. Addition for the simple scene range from 0.97 - 1.25 and advanced 0.76 - 1.29.

Comparing these to direct comparison, the scores for $\lambda = 1$ multiplication gives us the range 30.34 - 59.14 for the simple scene and 31.93 - 65.51 for the advanced. $\lambda = 0$ multiplication results in scores in the range 0.99 - 1.93 for the simple scene, and 0.76 - 1.98 for the advanced scene.

Initial inspection here shows that for $\lambda = 1$, the numbers for are quite similar. Only the addition factor for the seems to predict a bit worse quality for the simple scene, and also the scores for the advanced scene are a little bit higher.

For the $\lambda = 0$, the scores are almost the same for the multiplication factor, with a little smaller max using our framework.

The full output of the metric can be seen in Figure 25.

HDR-VDP

HDR-VDP is at the name implies a difference metric as well. As the metric is implemented in the form of a compiled program, the inner workings of the program is not tested, but we presume that the default settings gives us scores that are representative for the metric. As discussed in Section 5.6 only the 75% and the 95% detection probabilities are used for measurements for the quality of the reproduction. The 75% output for the simple scene ranges from 0.07 - 0.12 and 0.11 - 0.20 for the advanced scene. For the 95% threshold, the values range from 0.05 to 0.09 for the simple scene and 0.07 - 0.17 for the advanced scene. These numbers are direct error measures that runs through its own processing pipeline, and therefore our framework is not used. From these numbers we can see that it looks like the simple scene has fewer errors than the advanced scene where the number of predicted errors is higher for both of the values.

The full overview of the HDR-VDP metric is available in Table 26.

DRI

The DRI metric is the only metric developed that claims actually to be able to determine differences between LDR and HDR. The online implementation has the same limitations as the HDR-VDP, and as it only outputs error difference maps.

For the simple scene the calculated errors from the error maps as described in Section 5.6 are in the range of 0.13 - 0.44 for the simple scene, and 0.33 - 0.43 for the advanced scene. As was the case with HDR-VDP we can see that the minimal error seems to be less on the simple scene, but for the advanced scene the errors have a much closer scale. This seems to contradict the SSIM scores which the metrics internal quality measures are based on.

For the DRI metric, the results are available in Table 27.

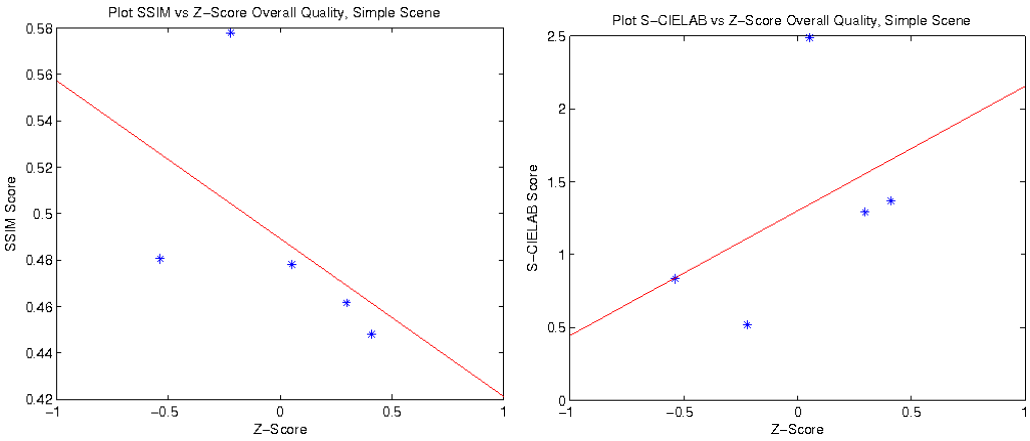
6.3 Correlation of Findings

To check for any dependence between the datasets, a correlation check is often used. A correlation is defined as any departure of two or more random variables from independence. The two types of

correlations we are going to check for is the Pearson correlation which is only sensitive to a linear relationship between the two datasets and the Spearman correlation which is a rank correlation.

When the comparison is done, it is important to determine what basis is to be used. As we compare SSIM scores with a range [-1,1] where 1 is a perfect image, and error difference metrics like S-CIELAB where a larger number means larger difference. This can easily be visualized by a plot of the two functions in Figure 15. As it is shown, the linear correlation is inverted because of the different ranges. By inverting the range of all the metrics that are difference based, a comparison is made possible.

A plot showing this alterations to the S-CIELAB data is available in Figure 16 where a higher score in error metrics are mapped the same way. This results in correlation scores that are directly comparable.



(a) SSIM plot (b) S-CIELAB plot

Figure 15: Plots of correlation, Overall Quality, Simple Scene. Red line is the linear correlation.

All the results presented in the following sections have been corrected so they are comparable to SSIM as a basis.

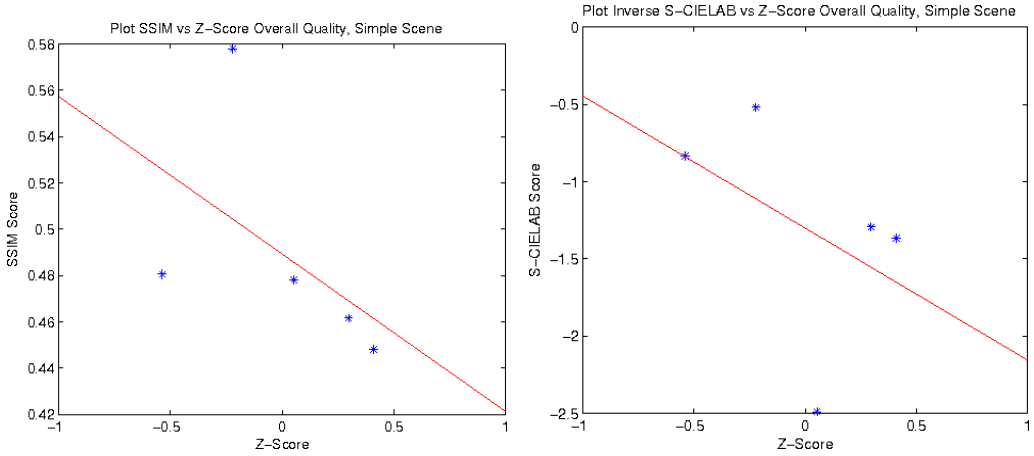
Overall Quality

In Table 13 we present as list of results for correlation for the overall quality for the simple scene sorted by Pearson Correlation.

One score in particular stands out, and that is the TVD PU $\lambda = 0+$. This also gets a perfect Spearman correlation, but this is mainly due to the small number of sample points in our dataset. It looks to be a good measure as the p-value is only 8%, but is still above the 95% interval which is often used to be statistically significant. Therefore could the correlation we see here be due to randomness.

In Table 14 we present the same overall scores for the advanced scene.

The best score is achieved by the HDR-VDP 95% metric, but as shown by the p-value, this result is not significant as it is over a 30% chance this correlation could be random. Another interesting thing is that the metric TVD PU $\lambda = 0+$ which performed very good on the simple scene is the second to



(a) SSIM plot (b) Inverted S-CIELAB plot

Figure 16: Plots of correlation, Overall Quality, Simple Scene. S-CIELAB inverted. Red line is the linear correlation.

last with a very high probability for randomness.

Detail Quality

When comparing the detail quality presented in Table 15, TVD $\lambda = 1+$ gives the best Pearson correlation score. This is interesting as earlier test have shown a better correlation of detail quality when only the TV term, $\lambda = 0$ was evaluated. Although the values are good, the p-values are still to high to be considered significant. Another thing to note is that the metric S-CIELAB S-CIELAB metric comes out as the second and third best, as this metric is created to measure color quality. These result does also have very high p-values, and can therefore be explained by random chance.

For the advanced scene, the correlation for the detail quality is presented in Table 16. Here the best result is interestingly the S-CIELAB noPU median *. The p-values are however not low enough to say that this is a significant result. As S-CIELAB also comes high up here, the p-values are to high to be considered significant.

Color Quality

For the color quality, the results from the simple scene is presented in Table 17. Once again the measure TVD PU $\lambda = 0 +$ is the one that stands out and gives a very good correlation. The same metric does not yield a Spearman score as good, and combined with the Spearman p-value this may indicate that the measure may be due to random chance.

For the advanced scene, the correlation results are presented in Table 18. Here once again the metric HDR-VDP seems to give the best correlation, but as was the case with the overall quality, the p-values are over 30% so it cannot be considered a significant result. The same can be said about TVD $\lambda = 0 +$ which produces the best result for the simple scene.

Table 13: Correlation between metrics and overall quality z-scores, simple scene. Sorted by Pearson Correlation

Metric	Pearson ρ	Pearson p-value	Spearman ρ	Spearman p-value
TVD PU $\lambda = 0$ +	0.8298	0.0821	1.0000	0.0167
HDR-VDP95%	0.6565	0.2288	0.7000	0.2333
TVD noPU $\lambda = 0$ *	0.6062	0.2785	0.1000	0.9500
HDR-VDP 75%	0.5845	0.3007	0.8000	0.1333
TVD PU $\lambda = 0$ *	0.502	0.3887	0.1000	0.9500
DRI	0.2302	0.7096	0.6000	0.3500
SSIM noPU *	-0.1493	0.8107	-0.9000	0.0833
S-CIELAB PU mean +	-0.3242	0.5946	-0.5000	0.4500
S-CIELAB PU median +	-0.347	0.5672	-0.5000	0.4500
S-CIELAB noPU mean *	-0.3493	0.5645	-0.5000	0.4500
TVD PU $\lambda = 1$ *	-0.3599	0.5518	-0.6000	0.3500
TVD no PU$\lambda = 1$ *	-0.3797	0.5284	-0.5000	0.4500
SSIM-IPT PU +	-0.4057	0.498	-0.5643	0.3667
S-CIELAB PU mean *	-0.4414	0.4568	-0.6000	0.3500
S-CIELAB noPU median *	-0.5046	0.3859	-0.5000	0.4500
SSIM PU *	-0.5127	0.3771	-0.9000	0.0833
TVD PU $\lambda = 1$ +	-0.5258	0.3628	-0.5000	0.4500
SSIM-IPT noPU *	-0.5299	0.3584	-0.6000	0.3500
SSIM-IPT PU *	-0.5458	0.3413	-0.6000	0.3500
SSIM PU +	-0.5644	0.3216	-0.5000	0.4500
S-CIELAB PU median *	-0.6019	0.2828	-0.5000	0.4500

Overall Quality correlation with specific parameters

In addition to test for just the correlations of the different metrics, we wanted to test for how good correlation of the different components were when compared to the overall quality for the scenes. This gives us an indication of what is considered important in the different scenes.

In Table 19 the output of the correlation scores are given. We see that for the simple scene, the color quality is the most important, but still not significantly. The detail factor on the other hand gets a very poor result, but also here the p-value for both Pearson and Spearman tests are too high to give any definitive results.

It may be noted that the reproductions of the simple scene has a very yellowish tint due to white balancing of the scene illuminant, and this may affect the result.

For the advanced scene where the white balancing is more correctly adjusted by the camera, both the color and the detail gets almost even correlation scores for Pearson correlations. With the resulting p-values the results still cannot be said to be significant. The Spearman correlations however get a very low correlation with high p-values, so the results here are not significant.

As for the artifacts, the comparison between overall scores and artifacts has to be considered. A higher number of artifacts should result in a higher degree of error the same ways as for error difference metrics considered above, and the same correction test is performed. This results in that

Table 14: Correlation between metrics and overall quality z-scores, Advanced scene. Sorted by Pearson Correlation

Metric	Pearson ρ	Pearson p-value	Spearman ρ	Spearman p-value
HDR-VDP 95%	0.5686	0.3172	0.8000	0.1333
HDR-VDP 75%	0.4886	0.4036	0.3000	0.6833
S-CIELAB noPU median *	0.2656	0.6659	0.3000	0.6833
SSIM PU +	0.1902	0.7593	0.3000	0.6833
SSIM-IPT noPU *	0.1776	0.7751	-0.1000	0.9500
S-CIELAB noPU mean *	0.1763	0.7766	0.3000	0.6833
S-CIELAB PU median *	0.1749	0.7784	0.3000	0.6833
S-CIELAB PU median +	0.1701	0.7845	0.3000	0.6833
SSIM-IPT PU +	0.1646	0.7913	0.1026	0.9000
TVD noPU $\lambda = 1$ *	0.1562	0.802	0.3000	0.6833
S-CIELAB PU mean +	0.1385	0.8242	0.3000	0.6833
SSIM-IPT PU *	0.137	0.8262	-0.1000	0.9500
SSIM PU *	0.097	0.8767	-0.1000	0.9500
TVD PU $\lambda = 1$ *	0.0914	0.8837	-0.1000	0.9500
S-CIELAB PU mean *	0.0836	0.8937	-0.1000	0.9500
TVD PU $\lambda = 0$ +	0.0732	0.9069	0.3000	0.6833
TVD PU $\lambda = 1$ +	0.071	0.9097	0.3000	0.6833
TVD noPU $\lambda = 0$ *	-0.0462	0.9412	0.1000	0.9500
DRI	-0.0589	0.925	-0.1000	0.9500
SSIM noPU*	-0.0899	0.8857	-0.1000	0.9500
TVD PU $\lambda = 0$ *	-0.1216	0.8456	0.1000	0.9500

for the simple scene, the score for both Pearson and Spearman indicates a low correlation, but due to high p-values cannot be said to be not significant. For the advanced scene, the values indicate a very high correlation, and here the p-values are very low, actually as low as 1% for Pearson correlation, and the same tendency is seen for Spearman. This indicates that the artifact attribute is considered very important for the overall quality of this scene, even more so than what is the case of both color and details.

The conclusion for this test must then be that overall quality is highly dependent on the scene content and the rendering. As the simple scene has a problem with the color, this may affect the detail scores obtained last during the perceptual experiment. For the advanced scene, it shows an almost equal correlation between color and detail, and this indicates that if there is no strong color casts present in the scene, the color and details are considered with about the same importance. For artifacts, this gets an even higher score for the advanced scene, and from the results that are statistically significant with p-values of 0.01, we can conclude that this may be the most important attribute i the tone mapping process. However, for the simple scene were there is a colorcast present, the detail factor does not show to be considered important, and neither is the artifacts. The p-values are to high, but also for the color the values are to low to be statistically significant. From this we can

Table 15: Correlation between metrics and detail quality z-scores, Simple scene. Sorted by Pearson Correlation

Metric	Pearson ρ	Pearson p-value	Spearman ρ	Spearman p-value
TVD PU $\lambda = 1$ +	0.4711	0.4232	0.2000	0.7833
S-CIELAB PU median *	0.2928	0.6326	0.2000	0.7833
S-CIELAB noPU median *	0.2917	0.6339	0.2000	0.7833
SSIM PU +	0.2617	0.6707	0.2000	0.7833
SSIM-IPT noPU *	0.2485	0.6869	-0.1000	0.9500
SSIM-IPT PU *	0.2262	0.7144	-0.1000	0.9500
SSIM PU *	0.2203	0.7218	0.0000	1.0000
HDR-VDP 75%	0.2078	0.7373	0.2000	0.7833
HDR-VDP 95%	0.1549	0.8036	0.0000	1.0000
SSIM noPU *	0.1196	0.8481	0.0000	1.0000
SSIM-IPT PU +	0.1100	0.8602	0.0513	1.0000
S-CIELAB PU median +	0.0996	0.8734	0.2000	0.7833
TVD noPU $\lambda = 1$ *	0.0963	0.8775	0.2000	0.7833
S-CIELAB noPU mean *	0.0951	0.8791	0.2000	0.7833
S-CIELAB PU mean *	0.0881	0.8880	-0.1000	0.9500
S-CIELAB PU mean +	0.0556	0.9292	0.2000	0.7833
TVD PU $\lambda = 1$ *	0.0409	0.9480	-0.1000	0.9500
DRI	0.0269	0.9658	-0.5000	0.4500
TVD noPU $\lambda = 0$ *	-0.4416	0.4565	-0.6000	0.3500
TVD PU $\lambda = 0$ *	-0.4433	0.4547	-0.6000	0.3500
TVD PU $\lambda = 0$ +	-0.5213	0.3677	-0.3000	0.6833

conclude that for "problem images" with color casts, we are not able to say if there is any dependency between overall quality and the different quality attributes. From this

6.3.1 Metric vs Metric Correlation

By comparing all the metrics against each other, we can get an indication of how the different metrics compares against each other. In Table 34, Appendix E we have calculated the correlation for all the metrics for the simple scene in a cross matrix, and the same is done for the advanced scene in Table 39. To save space, the metric names have been replaced with numbers, the mapping of these are available in Table 44.

From the metric, we see that the best metric that is proposed to be a truly dynamic range metric (DRI) has a ok correlation with SSIM-IPT PU + which has a correlation factor of 0.7484. with a p-value of 0.7484 for the simple scene, and for the advanced scene the best correlation is with TVD $\lambda = 0$ noPU * with the scores 0.8856 and p-value 0.0456. For other HDR enabled metrics, like HDR-VDP, the best results are with the different detection thresholds defined for each scene.

The color part of the metric gives results between 0.9990 - 0.8531 with p-values 0.0096 - 0.8935 for all the different usages of the S-CIELAB metric for the simple scene. For the advanced scene the same values range from 0.9965 - 0.0837. The reason for this is not clear, but it is believed that the

Table 16: Correlation between metrics and detail quality z-scores, Advanced scene. Sorted by Pearson Correlation

Metric	Pearson ρ	Pearson p-value	Spearman ρ	Spearman p-value
S-CIELAB noPU median *	0.5823	0.3029	0.5000	0.4500
S-CIELAB PU median +	0.5228	0.3661	0.5000	0.4500
S-CIELAB PU median *	0.5117	0.3782	0.5000	0.4500
S-CIELAB PU mean +	0.4895	0.4026	0.5000	0.4500
S-CIELAB noPU mean *	0.4875	0.4048	0.5000	0.4500
TVD noPU $\lambda = 1$ *	0.4546	0.4418	0.5000	0.4500
TVD PU $\lambda = 1$ *	0.4469	0.4505	0.1000	0.9500
S-CIELAB PU mean *	0.4219	0.4792	0.1000	0.9500
SSIM-IPT PU +	0.3280	0.5900	0.3078	0.6667
SSIM PU +	0.2989	0.6251	0.5000	0.4500
HDR-VDP 75%	0.1588	0.7987	0.2000	0.7833
SSIM-IPT noPU *	0.1481	0.8121	0.1000	0.9500
SSIM-IPT PU *	0.1240	0.8425	0.1000	0.9500
HDR-VDP 95%	0.1136	0.8557	0.3000	0.6833
SSIM PU *	-0.0568	0.9277	-0.1000	0.9500
TVD PU $\lambda = 1$ +	-0.2539	0.6802	0.0000	1.0000
TVD PU $\lambda = 0$ +	-0.3743	0.5348	0.0000	1.0000
DRI	-0.5147	0.3748	-0.6000	0.3500
SSIM noPU *	-0.5516	0.3352	-0.1000	0.9500
TVD noPU $\lambda = 0$ *	-0.5704	0.3153	-0.6000	0.3500
TVD PU $\lambda = 0$ *	-0.6346	0.2501	-0.6000	0.3500

metric is highly influenced by the whitepoint settings for the simple scene as discussed before, and therefore the result is highly affected by this.

SSIM that looks at detail in images seem to have a very low correlation. The SSIM PU * gets its best correlation against SSIM-IPT PU + with the scores 0.9592, for the simple scene, and SSIM-IPT PU * for the advanced scene with score 0.9252. The same is also the case with SSIM PU +, the best correlation is with the SSIM-IPT. Using direct comparison for the simple scene For direct comparison, there is no good correlation to any of the other metrics.

TVD with a setting of $\lambda = 0$ also only looks at detail in the image, and it shows that there is a good correlation between the TVD noPU * and TVD PU * for the simple scene. This is also apparent in the advanced scene with subsequent scores of over 0.9.

The last metric is TVD with $\lambda = 1$. For the simple scene, the variations noPU * and PU * shows a very good correlation with the S-CIELAB scores with values over 0.95. Interesting is it that the variation PU + shows a very poor overall correlation, especially for color.

On the advanced scene we see the total opposite. The variation PU + gives the best correlations against the S-CIELAB metric with scores up to 0.9768,0.1311. for S-CIELAB noPU * median, and all the other values are under 0.9.

From these results that is so spread, it is very hard to draw any definitive conclusions. The metrics

Table 17: Correlation between metrics and color quality z-scores, Simple scene. Sorted by Pearson Correlation

Metric	Pearson ρ	Pearson p-value	Spearman ρ	Spearman p-value
TVD PU $\lambda = 0$ +	0.9668	0.0072	0.5000	0.4500
TVD noPU $\lambda = 0$ *	0.8782	0.0501	0.3000	0.6833
TVD PU $\lambda = 0$ *	0.8416	0.0738	0.3000	0.6833
HDR-VDP 95%	0.4228	0.4782	0.1000	0.9500
SSIM noPU *	0.1781	0.7745	-0.2000	0.7833
DRI	0.1679	0.7873	0.3000	0.6833
HDR-VDP 75%	0.1636	0.7926	-0.1000	0.9500
SSIM PU *	0.1326	0.8316	-0.2000	0.7833
SSIM-IPT noPU *	0.0721	0.9083	0.2000	0.7833
SSIM-IPT PU +	0.0650	0.9173	0.1026	0.900
SSIM-IPT PU *	0.0595	0.9243	0.2000	0.7833
SSIM PU +	-0.0421	0.9465	0.0000	1.0000
TVD noPU $\lambda = 1$ *	-0.0553	0.9297	0.0000	1.0000
S-CIELAB noPU mean *	-0.0579	0.9263	0.0000	1.0000
S-CIELAB PU mean +	-0.0797	0.8986	0.0000	1.0000
TVD PU $\lambda = 1$ *	-0.0917	0.8834	0.2	0.7833
S-CIELAB PU median +	-0.1116	0.8581	0.0000	1.0000
S-CIELAB PU mean *	-0.1255	0.8406	0.2000	0.7833
TVD PU $\lambda = 1$ +	-0.1855	0.7652	-0.5000	0.4500
S-CIELAB noPU median *	-0.2287	0.7114	0.0000	1.0000
S-CIELAB PU median *	-0.2992	0.6249	0.0000	1.0000

that are considered dynamic range independent like the HDR-VPD and DRI shows very low correlation for almost all measures given. The color part shows very promising results for a single scene, but is totally reversed for the advanced scene. Therefore we cannot conclude that the metric tested will yield good results for a set of different scenes.

Table 18: Correlation between metrics and color quality z-scores, Advanced scene. Sorted by Pearson Correlation

Metric	Pearson ρ	Pearson p-value	Spearman ρ	Spearman p-value
HDR-VDP 95%	0.5839	0.3013	0.9000	0.0830
TVD PU $\lambda = 0$ +	0.5305	0.3577	0.6000	0.3500
TVD noPU $\lambda = 0$ *	0.3975	0.5076	0.3000	0.6833
SSIM-IPT noPU *	0.3625	0.5488	-0.2000	0.7833
SSIM-IPT PU +	0.3490	0.5648	-0.0513	1.0000
HDR-VDP 75%	0.3458	0.5687	0.4000	0.5167
TVD PU $\lambda = 0$ *	0.3308	0.5866	0.3000	0.6833
SSIM-IPT PU *	0.3232	0.5958	-0.2000	0.7833
SSIM PU +	0.3185	0.6014	0.1000	0.9500
SSIM PU *	0.3088	0.6132	-0.3000	0.6833
S-CIELAB noPU mean *	0.2735	0.6561	0.1000	0.9500
TVD noPU $\lambda = 1$ *	0.2633	0.6686	0.1000	0.9500
S-CIELAB noPU median *	0.2601	0.6726	0.1000	0.9500
S-CIELAB PU median +	0.2318	0.7075	0.1000	0.9500
S-CIELAB PU mean +	0.2224	0.7192	0.1000	0.9500
TVD PU $\lambda = 1$ *	0.1809	0.7709	-0.2000	0.7833
S-CIELAB PU mean *	0.1657	0.7900	-0.2000	0.7833
S-CIELAB PU median *	0.1518	0.8075	0.1000	0.9500
SSIM noPU *	0.0122	0.9844	-0.3000	0.6833
TVD PU $\lambda = 1$ +	-0.0047	0.9941	0.1000	0.9500
DRI	-0.0733	0.9068	0.2000	0.7833

Table 19: Correlation between overall quality and color/detail z-scores.

Correlation test	Pearson ρ	Pearson p-value	Spearman ρ	Spearman p-value
Overall/Color Simple Scene	0.7717	0.1264	0.5000	0.4500
Overall/Color Advanced Scene	0.8327	0.0801	-0.1000	0.9500
Overall/Detail Simple Scene	-0.5785	0.3069	-0.3000	0.6833
Overall/Detail Advanced Scene	0.8195	0.0895	0.3000	0.6833
Overall/Artifacts Simple Scene	0.2429	0.6938	0.2052	0.7667
Overall/Artifacts Advanced Scene	0.9559	0.0110	0.9000	0.0833

7 Conclusion and Further Research

In this chapter, a set of conclusions that can be drawn from the gathered result will be presented. In addition we will present a set of different ideas and possibilities for future research in the field that that this thesis touches.

7.1 Conclusion

From the results gathered from the perceptual experiment, we are able to answer the first research question. When looking at the results in Figure 13 for the simple scene, it looks like the TMO STRESS by Kolås performs around average for both overall and detail quality attributes. For the color quality, it performs a little bit under the other metrics, but this is to be expected as a contrast equalization is the trait of this TMO . For the advanced scene as presented in Figure 14, we see another pattern. For the overall quality, the TMO seems to score in the middle of the tested tone mappers, but for both detail and color quality it is scored among the lower. The color quality may be affected by the equalization mentioned above, but the detail quality is another matter. It however seems that there is a strong dependency between the color and detail quality when compared to the overall quality of the advanced scene, and it is believed that this is affect the score.

As for the number of artifacts found on this TMO, it places itself in the middle. The main problem with it the tested scenes is the dark background which gets a much brighter appearance due to the contrast stretching and color adjustments. These are altered to such an extreme degree that by many of the observers the changes are considered an artifact. This is also noticed in the simple scene with the yellowish light that can be seen in Appendix A, Figure 23 almost looks like is lit by the same lamp as the advanced scene depicted in Appendix A, Figure 28.

From the discussion above, the TMO STRESS certainly has a great potential, but it has some problems especially with scenes that have a non uniform color cast when compared to other state of the art TMOs. Some measures to correct for this has been implemented in other versions of STRESS as those developed by A.B.M. Tariqul Islam [67] where a preprocessing stage and post-processing stage enables us to adjust the color balance. The drawback with such a solutions is that you get another parameter which you adjust for each original that you want to make a reproduction of.

We therefore must conclude that further studies using the implementation above should be done, but in it's current form the STRESS algorithm performs around average when compared to the scenes produced, and it is believed that better results could be obtained using a version which could correct for color balance.

As discussed in Section 5.4, we have looked only at four of the different attributes which are considered important in a HDR scene. By analyzing the scores from the perceptual experiment we can try to answer the second research question.

The results from these shows us that there is a very strong and significant Pearson correlation between artifacts and overall quality for our advanced scene, but for the simple scene the same score is shown to be much lower. The same is the case with both color and details for the advanced scene, whereas in the simple scene the results does not show any strong correlation. An extent of this is the fact that in our findings for the simple scene, the details seems to be considered not important, but due to high p-values, this result is as discussed not significant. When looking at the Spearman correlation, another pattern emerges. Artifacts are still considered vitally important for the advanced scene, but both color and detail get much lower scores with higher degrees of error. This may indicate that the data has some outliers. By plotting these data as seen in Figure 17, we see this outlier very clearly. This observation does indicate that the Pearson correlation may not yield correct measures.

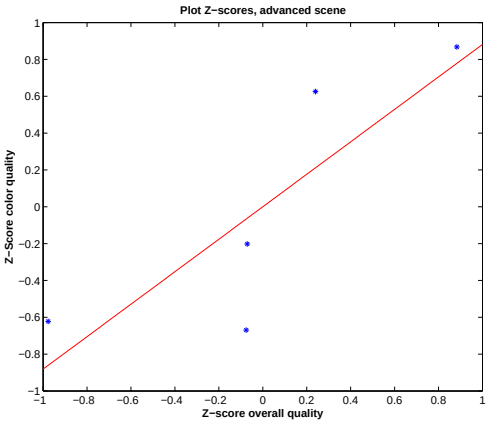


Figure 17: Plot Z-score overall quality/Z-score color quality for the advanced scene

To be able to improve these scores, we need to test more TMOs on the same setup, but due to time constraints, this was not done.

Another reason why the detail is disregarded for the simple scene is believed to originate in the whitepoint settings of the camera. The whitepoint settings obtained with the camera is when comparing the "best effort" shot of the scene scene shown in Figure 8, this shows a strong yellowish color cast which is not apparent in the perceived scene. This also is apparent in the different tone mapped versions of the scene. As the color part of the simple scene shows a correlation about the same as for the advanced scene, this indicates that they are considered of almost equal importance. In our experiment the detail part was asked after the color, and it could lead to that this may then be an affect of this as the color cast is so strong that is remembered and affects the ratings for the details. To be able to answer this hypothesis, a small experiment where the order is changed could be performed.

From this we can draw the conclusion that for a scene which does not have a strong color cast, the all the different attributes tested indicate that they are important in a HDR tone mapping process, but due to low Spearman scores, this must be evaluated further. The artifacts attribute is shown to be more important, as both the Pearson and Spearman correlations for this are high, and from this

we can conclude that this may be the most important attribute in such scenes.

When strong color casts are present, other attributes such as details and artifacts seems to be overlooked if they are evaluated after the colors. Further research into the order can as stated can give more insight into this and should be done before we can conclude how the different attributes are linked together in such scenes.

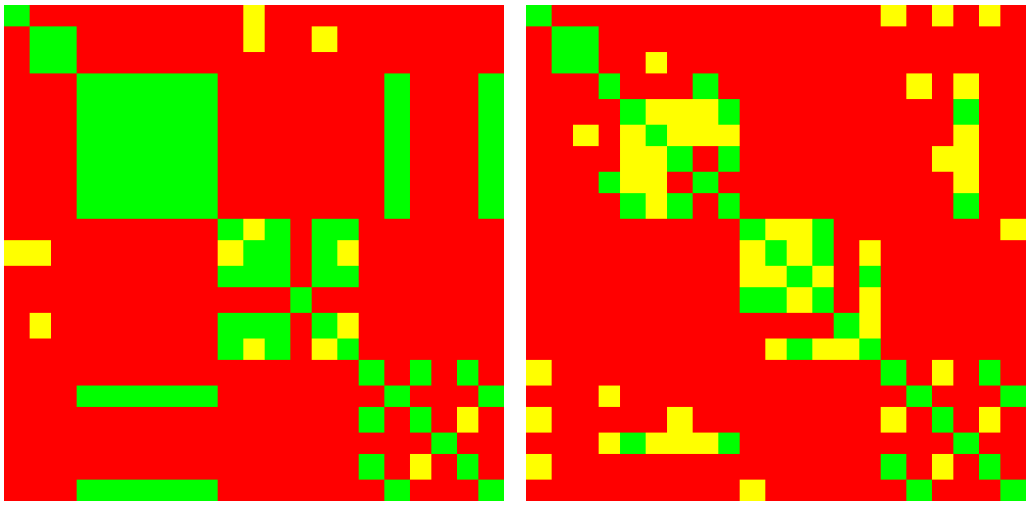
By the usage of our framework, we have proposed an extension that enables the usage of current state of the art image quality metrics. For luminance only metrics like SSIM, the results show that even with the usage of our metric, there is a very low correlation with the results from the perceptual experiment, even for detail qualities which the metric is developed for. Interestingly enough, the best score for details is achieved by a S-CIELAB metric which is developed for color, but overall the high p-values in the correlations makes these findings irrelevant. The results with the new metric TVD gives promising results on overall quality for the simple scene, but falls through on the advanced with a correlation score of around 0. Other metrics such as the DRI scores low on the overall quality on the simple scene, and even lower on the advanced scene. This is also seen for details which should be one of the strong points for this metric. HDR-VDP is the metric that seems to work best for overall quality on the advanced scene, and also give decent scores for the simple, indicating that this is the best overall metric. S-CIELAB gives the best correlation for color for the advanced scene which is to be expected. From the gathered results, the difference between direct comparison and our framework seems to yield in the favor of direct comparison for the median, but for all other measures our framework comes out better. This may be due to randomness as the p-values are from 0.30 to 0.40. For the simple scene, the whitepoint once again makes the data hard to compare. TVD used equal weighting of the TV and fidelity term with our framework a seems to do best, but cannot be considered significant. A final interesting observation is that TVD PU $\lambda = 0 +$ gives a very high correlation against the simple color score and also performs quite well on the advanced scene.

It has to be pointed out, that in the setup that we have done, we have presumed that the HDR image is a perfect reproduction of the scene. As we have no way of doing a visual inspection of the HDR, it is lightly that this could introduce some errors, however these should be small compared to the TMO process.

From this it shows that on a best effort scene there is no single metric that is shown to give good scores. The scores vary a lot between the different scenes, and because of the problems with the simple scene, a conclusion cannot be drawn in either way. It seems to be giving an indication that the framework is working, and comparison is possible, but further research with a wider selection of scenes and TMOs are required to verify these results. For the quality of the reproductions, further study is required as the correlation scores obtained have a to large degree of randomness to be considered significant.

The different image difference metrics have been well tested when comparing standard LDR images. From the data mapped in Table 34 for the simple scene, and Table 39 for the advanced scene, a better vizualization score can be built in form of a set of images as displayed in Figure 18.

Here the each metric is represented by 25 pixels horizontally and vertically, and every metric is checked for a correlation with all other metrics. This results in a diagonal with a score of 1 where



(a) Simple Scene (b) Advanced Scene

Figure 18: Image of correlation of different metrics in the form of a cross table. Green > 0.9, yello > 0.7, red < 0.7

the same metrics are compared. Also a mirroring around this diagonal can be observed. Metrics that match with a score over 0.9 is given a green score, yellow for > 0.7 and red for the rest. P-values are not taken into account as this is only for visualization. The ordering is the same for both scenes, and is given in Table 44. As can bee seen from the first vertial line, which is the DRI metric, this has a very low correlation with almost all of the metrics it compares to except to SSIM-IPT for the simple scene and TVD PU $\lambda = 0$ *. For the S-CIELAB metrics there is a overall good correlation with scores above 0.9, and they also have a good correlation with TVD PU $\lambda = 1$ + and TVD noPU $\lambda = 1$ * for the simple scene. For the advanced scene, we see the same tendencies, but with a little lower correlation. It should be noted that TVD noPU $\lambda = 1$ * here drops down to the red range.

For the next block which is the SSIM-IPT, we see a good correlation between all the different variations of the metric. In addition we see a strong correlation with standard SSIM scores. The same tendencies, but with weaker correlation is also apparent in the advanced scene.

The only metric that should be mentioned is TVD, where we see a strong correlation in the λ values, but no significant other changes on the different scenes.

From this it can be assumed that the proposed framework is quite stable for a wide variety of metrics, and the low variations of correlation on the different metric settings in different scenes indicate that image content does not hugely affect the scores obtained. However, there are some problem areas such as the color where we see a total reversal. Therefore further research should be done to check these findings on a wider settings of images.

The last research question is difficult to answer. As referred to many times in this master thesis, the setup seems to be working for the most part, but no very good results are obtained. Also the

problem with the whitepoint in the simple scene which gives radically different results than for the advanced scene.

From the results seen in the perceptual experiment, the output from some tone mappers seem to work fine on the tested scenes, and some does not produce as good results. This may be due to operator settings used, but as shown the results vary to a very high degree that no direct conclusion can be drawn. To be able to fully test if using the scene as a reference, a display that is able to accurately display the HDR data can be used as a reference for a similar experiment. If there is a high correlation between the results from such an experiment and one using the original scene as the reference, the question could be answered.

In our case as no technology that is able to display the original data was available to do the testing described above were not performed. Further research into this is therefore required.

7.2 Further Research

As said in the introduction of this chapter, we will present some possibilities for further research.

The first thing is to extend our test by introducing more tone mappers. We have only tested a very small set of the different tone mappers available, and new ones are under constant development. By testing for this, a better basis for comparison can be obtained and correlation tests will show a lower degree of error(p-values). Also other types of test such as Spearman and Kendall-test can be used, as Spearman tend to give very coarse results when comparing few samples, and Kendall Tau test can help us determine if we should reject the null hypotheses.

An extension of using a wider set of TMOs is of corse to use other image difference metrics. More and more metrics appear for testing the quality, and although many of them are primarily developed for LDR, they could be tested by using our proposed framework. Hopefully a set of metrics that could work for a wider range of images could be connected, and a simple scoring could be obtained.

An extension of this, is that if a consistent and stable scoring system could be developed. This could be narrowed down to be used to automate the time consuming tweaking of different variables that many of the proposed TMOs operate with, and in the end help to create better reproductions using less time.

As discussed in the conclusion, the need to do test if using a scene is a viable solution is something that should be done. This can be done without any large alterations to both the framework or the test procedures as the same values that are measured on the scenes in our case can be either obtained through measurement or by design of the HDR display. One such display that has been used is the BrightSide DR37-P which has now been acquired by Dolby and is included in their Dolby Vision Solar Series displays, but no publicly available information is available at this time.

Another way is to try to edit the framework, so instead of normalizing the values on the lightness scale, one can try to alter it on either the chromaticity or the hue values. This will of corse introduce other types of shifts in the images. This could also result in better color scores as the chroma change will then be avoided.

A final thing to check for is that in the field of contrast sensitivity functions (CSF), there has begun some research by using LEDs to create monochromatic light, and deriving the CSFs for these [68].

By determining this for certain values such as RGB, the basis used in our framework can be extended to also work for colors without introducing any shift by doing a PU coding directly on each channel. Another way to go is to try to use the works by Mullen [69] who mapped the contrast sensitivity of opponent chromatic gratings, and to check if this can be successfully integrated.

All in all, as HDR becomes more and more commonly used, new research and new ways of integrating it into different situations will become important. Therefore I predict that this is a field that will receive both a lot of commercial as well as academical interest in the coming years.

A Output from Tone Mappers

Figure 19: Output Durand Fast Bilateral Filtering tone mapping operator, simple scene

Figure 20: Output Fattal Gradient Domain tone mapping operator, simple scene

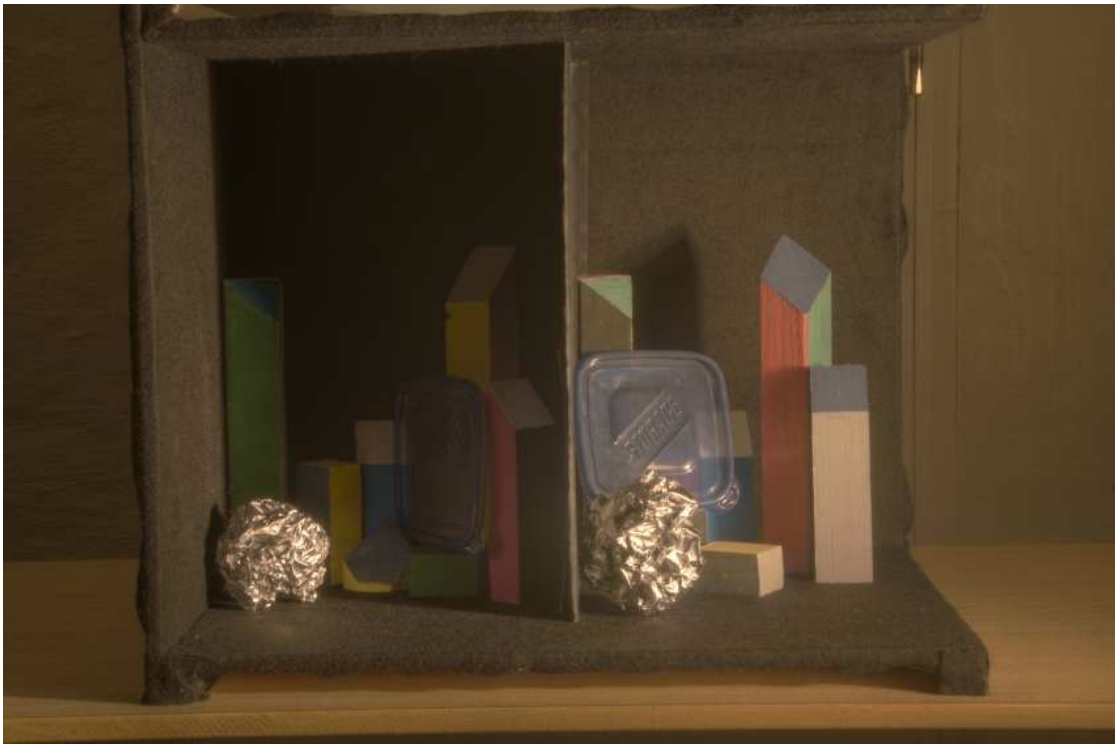


Figure 21: Output Mantiuk Contrast Framework tone mapping operator, simple scene

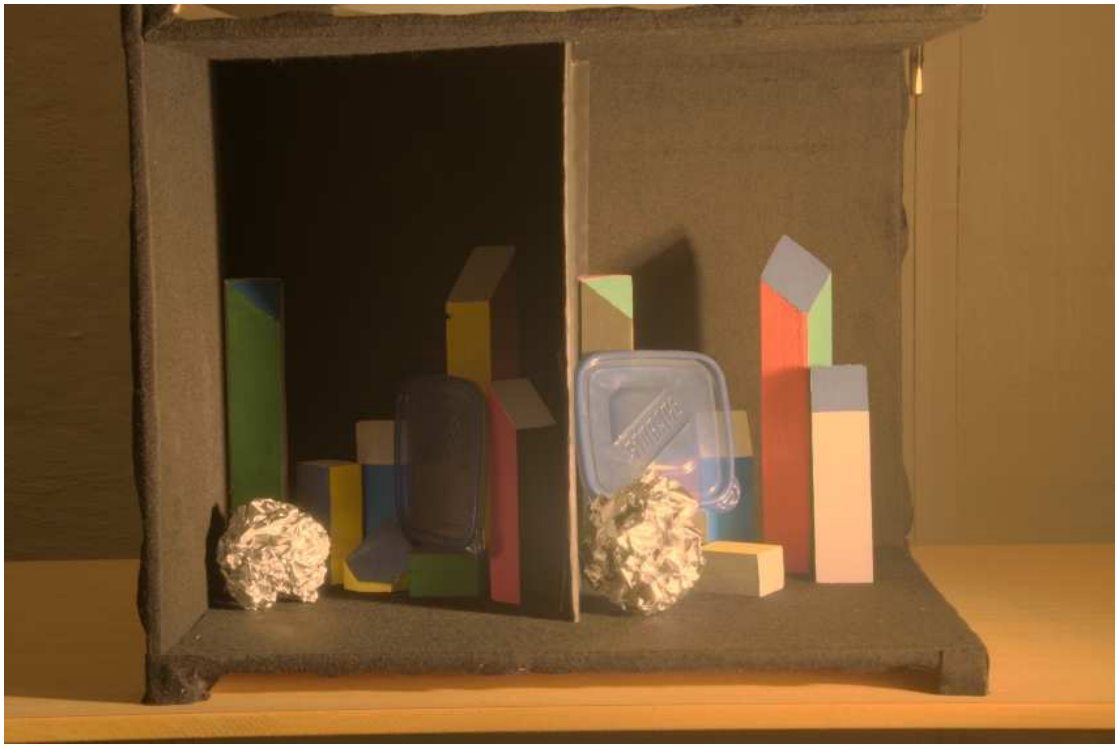


Figure 22: Output Reinard Photographic tone mapping operator, simple scene

Figure 23: Output Kolås STRESS tone mapping operator, simple scene

Figure 24: Output Durand Fast Bilateral Filtering tone mapping operator, advanced scene

Figure 25: Output Fattal Gradient Domain tone mapping operator, advanced scene

Figure 26: Output Mantiuk Contrast Framework tone mapping operator, advanced scene

Figure 27: Output Reinard Photographic tone mapping operator, advanced scene

Figure 28: Output Kolås STRESS tone mapping operator, advanced scene

B Protocol Used for Perceptual Experiment

B.1 Simple Scene

In the following sections, the first part of the written instructions for the perceptual experiment which was presented to all the participants are given.

Overall Quality

Part 1 – Simple Scene – Overall Quality

What is Image Quality ?

There are a lot of different measures of image quality, but these are highly subjective. In this first task we are interested in your opinion on the given set of images, and what image is the most correct.

In this part of the experiment, we are interested in your opinion on the following for the given samples:

Which of the given images is the most correct reproduction of the given scene with the respect to overall quality?

Please click on the image that answers the question above best.
If you think the reproductions are equal, you still have to choose on of them.

Color Quality

Part 2 – Simple Scene – Color

What is color?

Color is a sensation caused by light of different wavelengths, which is reflected from a material, reaches the eye, converted to an electrochemical signal and interpreted by the brain. Colors may be described in a lot of ways “full”/saturated or “dull”/understaturated. The “base color”, hue, may be incorrect or a lot of other errors may occur in the process of creating a reproduction of a scene.

In this part of the experiment, we are interested in your opinion on the following on the given samples:

Which of the given images is the most correct reproduction of the given scene with the respect to the aspect of colors?

Please click on the image that answers the question above best.
If you think the reproductions are equal, you still have to choose on of them.

Detail Quality

Part 3 – Simple Scene – Details

What are details?

Details in imaging, is the correct reproduction of all small variations that we can see in the original scene. These may be lost due to image resolution, but also to different types of noise both in the capturing and processing phase of a reproduction of a scene.

In this part of the experiment, we are interested in your opinion on the following on the given samples:

Which of the given images is the most correct reproduction of the given scene with the respect of the aspect of details?

Please click on the image that answers the question above best.

If you think both reproductions are equal, you still have to choose on of them.

Artifacts

Part 4 – Simple Scene – Artifacts

What are artifacts?

Artifacts in imaging, is unwanted elements or noise in an image. This may occur due to imperfections in optical equipment, the image sensor or in the processing of image data. Some artifacts that may occur are the following:

- Ringing effects – Areas around objects boundary become blurry.
- Halos – A “glowing” ring around objects and their borders.
- Flares – Diffraction of light that hits either barriers or other objects and creates bright spots on an image.
- Other disturbing elements which is not a part of the original scene.

In this part of the experiment, we are interested in your opinion on the following on the given samples:

On the given sheet, please mark in any artifact appearing in the reproduction, but not in the original scene.

When you have marked them all down, you will get a new sheet and do the same for next reproduction.

There are 5 reproductions for the given scene.

B.2 Advanced Scene

Overall Quality

Part 5 – Advanced Scene – Overall Quality

What is Image Quality ?

There are a lot of different measures of image quality, but these are highly subjective. In this first task we are interested in your opinion on the given set of images, and what image is the most correct.

In this part of the experiment, we are interested in your opinion on the following for the given samples:

Which of the given images is the most correct reproduction of the given scene with the respect to overall quality?

Please click on the image that answers the question above best.
If you think the reproductions are equal, you still have to choose on of them.

Color Quality

Part 6 – Advanced Scene – Color

What is color?

Color is a sensation caused by light of different wavelengths, which is reflected from a material, reaches the eye, converted to an electrochemical signal and interpreted by the brain. Colors may be described in a lot of ways “full”/saturated or “dull”/understaturated. The “base color”, hue, may be incorrect or a lot of other errors may occur in the process of creating a reproduction of a scene.

In this part of the experiment, we are interested in your opinion on the following on the given samples:

Which of the given images is the most correct reproduction of the given scene with the respect to the aspect of colors?

Please click on the image that answers the question above best.
If you think the reproductions are equal, you still have to choose on of them.

Detail Quality

Part 7 – Advanced Scene – Details

What are details?

Details in imaging, is the correct reproduction of all small variations that we can see in the

original scene. These may be lost due to image resolution, but also to different types of noise both in the capturing and processing phase of a reproduction of a scene.

In this part of the experiment, we are interested in your opinion on the following on the given samples:

Which of the given images is the most correct reproduction of the given scene with the respect of the aspect of details?

Please click on the image that answers the question above best.
If you think both reproductions are equal, you still have to choose on of them.

Artifacts

Part 8 – Advanced Scene – Artifacts

What are artifacts?

Artifacts in imaging, is unwanted elements or noise in an image. This may occur due to imperfections in optical equipment, the image sensor or in the processing of image data. Some artifacts that may occur are the following:

- Ringing effects – Areas around objects boundary become blurry.
- Halos – A “glowing” ring around objects and their borders.
- Flares – Diffraction of light that hits either barriers or other objects and creates bright spots on an image.
- Other disturbing elements which is not a part of the original scene.

In this part of the experiment, we are interested in your opinion on the following on the given samples:

On the given sheet, please mark in any artifact appearing in the reproduction, but not in the original scene.

When you have marked them all down, you will get a new sheet and do the same for next reproduction.

There are 5 reproductions for the given scene.

C Output from Perceptual Matrices

Under are the raw data from the perceptual experiment presented.

Table 20: Results from perceptual experiments, simple scene

(a) Summation matrix, overall quality

0	6	9	16	16
28	0	21	28	18
25	13	0	22	20
18	6	12	0	17
18	16	14	17	0

(b) Z-score matrix, overall quality

0.0000	-0.8998	-0.6011	-0.0697	-0.0697
0.8998	0.0000	0.2833	0.8998	0.0697
0.6011	-0.2833	0.0000	0.3578	0.2108
0.0697	-0.8998	-0.3578	0.0000	0.0000
0.0697	-0.0697	-0.2108	0.0000	0.0000

(c) Summation matrix, color quality

0	4	16	19	15
30	0	32	32	26
18	2	0	19	13
15	2	15	0	13
19	8	21	21	0

(d) Z-score Matrix, color quality

0.0000	-1.1650	-0.0697	0.1398	-0.1398
1.1650	0.0000	1.5615	1.5615	0.6922
0.0697	-1.5615	0.0000	0.1398	-0.2833
-0.1398	-1.5615	-0.1398	0.0000	-0.2833
0.1398	-0.6922	0.2833	0.2833	0.0000

(e) Summation matrix, detail quality

0	20	15	7	19
14	0	11	6	17
19	23	0	7	14
27	28	27	0	24
15	17	20	10	0

(f) Z-score Matrix, detail quality

0.0000	-0.8998	-0.6011	-0.0697	-0.0697
0.8998	0.0000	0.2833	0.8998	0.0697
0.6011	-0.2833	0.0000	0.3578	0.2108
0.0697	-0.8998	-0.3578	0.0000	0.0000
0.0697	-0.0697	-0.2108	0.0000	0.0000

Table 21: Results from perceptual experiments, advanced scene

(a) Summation Matrix, overall quality

0	8	11	1	10
26	0	20	9	17
23	14	0	6	13
33	25	28	0	24
24	17	21	10	0

(b) Z-score matrix, overall quality

0.0000	-0.6922	-0.4351	-1.8909	-0.5158
0.6922	0.0000	0.2108	-0.6011	0.0000
0.4351	-0.2108	0.0000	-0.8998	-0.2833
1.8909	0.6011	0.8998	0.0000	0.5158
0.5158	0.0000	0.2833	-0.5158	0.0000

(c) Summation matrix, color quality

0	3	14	5	6
31	0	29	13	25
20	5	0	6	7
29	21	28	0	21
28	9	27	13	0

(d) Z-score matrix, color quality

0.0000	-1.3376	-0.2108	-1.0225	-0.8998
1.3376	0.0000	1.0225	-0.2833	0.6011
0.2108	-1.0225	0.0000	-0.8998	-0.7910
1.0225	0.2833	0.8998	0.0000	0.2833
0.8998	-0.6011	0.7910	-0.2833	0.0000

(e) Summation matrix, detail quality

0	16	8	2	3
18	0	8	4	11
26	26	0	6	9
32	30	28	0	15
31	23	25	19	0

(f) Z-score matrix, detail quality

0.0000	-0.0697	-0.6922	-1.5615	-1.3376
0.0697	0.0000	-0.6922	-1.1650	-0.4351
0.6922	0.6922	0.0000	-0.8998	-0.6011
1.5615	1.1650	0.8998	0.0000	-0.1398
1.3376	0.4351	0.6011	0.1398	0.0000

D Output from Metrics

Table 22: Output from SSIM metric

(a) Perceptual coding, multiplication factor

Scene	Durand	Fattal	Mantiuk	Reinhard	STRESS
Simple	0.4481	0.4806	0.5780	0.4616	0.4782
Advanced	0.6990	0.6558	0.7863	0.5906	0.7432

(b) Perceptual coding, addition factor

Scene	Durand	Fattal	Mantiuk	Reinhard	STRESS
Simple	0.3758	0.3959	0.4304	0.3744	0.3579
Advanced	0.4345	0.4365	0.4383	0.3968	0.4508

(c) Direct comparison, multiplication factor

Scene	Durand	Fattal	Mantiuk	Reinhard	STRESS
Simple	0.1006	0.1226	0.2270	0.1361	0.2960
Advanced	0.1812	0.1892	0.3091	0.1001	0.4859

Table 23: Output from SSIM-IPT metric

(a) Perceptual coding, multiplication factor

Scene	Durand	Fattal	Mantiuk	Reinhard	STRESS
Simple	0.0750	0.0797	0.0909	0.0760	0.0744
Advanced	0.0261	0.0261	0.0284	0.0218	0.0266

(b) Perceptual coding, addition factor

Scene	Durand	Fattal	Mantiuk	Reinhard	STRESS
Simple	0.0636	0.0659	0.0733	0.0636	0.0566
Advanced	0.0156	0.0156	0.0164	0.0139	0.0146

(c) Direct comparison, multiplication factor

Scene	Durand	Fattal	Mantiuk	Reinhard	STRESS
Simple	0.0695	0.0731	0.0833	0.0697	0.0687
Advanced	0.0184	0.0184	0.0200	0.0156	0.0215

Table 24: Output from S-CIELAB metric

(a) Perceptual coding, multiplication factor

Scene	Durand	Fattal	Mantiuk	Reinhard	STRESS
Simple Mean	1.3675	0.8344	0.5184	1.2913	2.4889
Simple Median	0.9113	0.4559	0.3765	0.9892	1.4821
Advanced Mean	1.2669	0.9583	0.7580	3.0608	2.2553
Advanced Median	0.1564	0.1667	0.0856	0.3699	0.0348

(b) Perceptual coding, addition factor

Scene	Durand	Fattal	Mantiuk	Reinhard	STRESS
Simple Mean	4.5365	3.8703	3.2815	4.6983	7.6720
Simple Median	4.1508	3.4990	3.0391	4.4485	7.0543
Advanced Mean	14.4791	11.5305	13.0258	17.3960	12.4847
Advanced Median	12.5918	10.8336	11.3888	15.2878	7.9259

(c) Direct comparison, multiplication factor

Scene	Durand	Fattal	Mantiuk	Reinhard	STRESS
Simple Mean	1.6599	1.2467	0.6411	1.8431	3.6377
Simple Median	1.1560	0.6701	0.4590	1.4730	2.3119
Advanced Mean	3.3901	2.8857	2.4402	6.8893	6.0835
Advanced Median	1.1724	1.1791	0.7440	2.3378	0.8580

Table 25: Output from TVD metric

(a) Perceptual coding, multiplication factor, $\lambda = 1$

Scene	Durand	Fattal	Mantiuk	Reinhard	STRESS
Simple	38.9158	32.6805	28.5003	38.3373	59.1485
Advanced	19.9819	19.2676	15.3060	30.1060	65.5097

(b) Perceptual fcoding, multiplication factor, $\lambda = 0$

Scene	Durand	Fattal	Mantiuk	Reinhard	STRESS
Simple	1.2689	1.6329	1.0246	1.1135	0.9967
Advanced	1.4264	1.9088	1.1213	1.4187	0.7624

(c) Perceptual coding, addition factor, $\lambda = 1$

Scene	Durand	Fattal	Mantiuk	Reinhard	STRESS
Simple	65.6175	62.6634	60.4351	65.6186	59.1485
Advanced	66.0078	65.5887	63.7031	70.2917	65.5097

(d) Perceptual coding, addition factor, $\lambda = 0$

Scene	Durand	Fattal	Mantiuk	Reinhard	STRESS
Simple	0.9739	1.2505	1.0000	0.9810	0.9967
Advanced	1.2693	1.2899	1.2914	1.2779	0.7624

(e) Direct comparison, multiplication factor, $\lambda = 1$

Scene	Durand	Fattal	Mantiuk	Reinhard	STRESS
Simple	41.7049	36.9395	30.3482	42.6052	59.1485
Advanced	36.9919	36.9790	31.9358	46.7693	65.5097

(f) Direct comparison, multiplication factor, $\lambda = 0$

Scene	Durand	Fattal	Mantiuk	Reinhard	STRESS
Simple	1.2930	1.9313	1.0799	1.1655	0.9967
Advanced	1.4286	1.9829	1.2380	1.3944	0.7624

Table 26: Output from HDR-VDP metric

(a) HDR-VDP with detection threshold of 75%

Scene	Durand	Fattal	Mantiuk	Reinhard	STRESS
Simple	0.0639	0.1123	0.1197	0.1102	0.0659
Advanced	0.1093	0.1089	0.1805	0.2005	0.1058

(b) HDR-VDP with detection threshold of 95%

Scene	Durand	Fattal	Mantiuk	Reinhard	STRESS
Simple	0.0473	0.0887	0.0758	0.0798	0.0503
Advanced	0.0723	0.0704	0.1367	0.1736	0.0785

Table 27: Output from DRI metric

Scene	Durand	Fattal	Mantiuk	Reinhard	STRESS
Simple	0.3865	0.4358	0.3959	0.3837	0.1371
Advanced	0.3785	0.4283	0.3989	0.4109	0.3319

E Output from Correlation

In this chapter we present the raw output from the correlation scores for the different attributes checked for.

Table 28: Correlation between metrics and overall quality z-scores, simple scene

Metric	Pearson λ	Pearson p-value	Spearman λ	Spearman p-value
SSIM PU *	-0.5127	0.3771	-0.9000	0.0833
SSIM PU +	-0.5644	0.3216	-0.5000	0.4500
SSIM noPU *	-0.1493	0.8107	-0.9000	0.0833
SSIM-IPT PU *	-0.5458	0.3413	-0.6000	0.3500
SSIM-IPT PU +	-0.4057	0.4980	-0.5643	0.3667
SSIM-IPT noPU *	-0.5299	0.3584	-0.6000	0.3500
S-CIELAB PU mean *	-0.4414	0.4568	-0.6000	0.3500
S-CIELAB PU median *	-0.6019	0.2828	-0.5000	0.4500
S-CIELAB PU mean +	-0.3242	0.5946	-0.5000	0.4500
S-CIELAB PU median +	-0.3470	0.5672	-0.5000	0.4500
S-CIELAB noPU mean *	-0.3493	0.5645	-0.5000	0.4500
S-CIELAB noPU median *	-0.5046	0.3859	-0.5000	0.4500
TVD PU $\lambda = 1$ *	-0.3599	0.5518	-0.6000	0.3500
TVD PU $\lambda = 1$ +	-0.5258	0.3628	-0.5000	0.4500
TVD noPU $\lambda = 1$ *	-0.3797	0.5284	-0.5000	0.4500
TVD PU $\lambda = 0$ *	0.5020	0.3887	0.1000	0.9500
TVD PU $\lambda = 0$ +	0.8298	0.0821	1.0000	0.0167
TVD noPU $\lambda = 0$ *	0.6062	0.2785	0.1000	0.9500
HDR-VDP 75%	0.5845	0.3007	0.8000	0.1333
HDR-VDP 95%	0.6565	0.2288	0.7000	0.2333
DRI	0.2302	0.7096	0.6000	0.3500

Table 29: Correlation between metrics and overall quality z-scores, advanced scene

Metric	Pearson λ	Pearson p-value	Spearman λ	Spearman p-value
SSIM PU *	0.0970	0.8767	-0.1000	0.9500
SSIM PU +	0.1902	0.7593	0.3000	0.6833
SSIM noPU *	-0.0899	0.8857	-0.1000	0.9500
SSIM-IPT PU *	0.1370	0.8262	-0.1000	0.9500
SSIM-IPT PU +	0.1646	0.7913	0.1026	0.9000
SSIM-IPT noPU *	0.1776	0.7751	-0.1000	0.9500
S-CIELAB PU mean *	0.0836	0.8937	-0.1000	0.9500
S-CIELAB PU median *	0.1749	0.7784	0.3000	0.6833
S-CIELAB PU mean +	0.1385	0.8242	0.3000	0.6833
S-CIELAB PU median +	0.1701	0.7845	0.3000	0.6833
S-CIELAB noPU mean *	0.1763	0.7766	0.3000	0.6833
S-CIELAB noPU median *	0.2656	0.6659	0.3000	0.6833
TVD PU $\lambda = 1$ *	0.0914	0.8837	-0.1000	0.9500
TVD PU $\lambda = 1$ +	0.0710	0.9097	0.3000	0.6833
TVD noPU $\lambda = 1$ *	0.1562	0.8020	0.3000	0.6833
TVD PU $\lambda = 0$ *	-0.1216	0.8456	0.1000	0.9500
TVD PU $\lambda = 0$ +	0.0732	0.9069	0.3000	0.6833
TVD noPU $\lambda = 0$ *	-0.0462	0.9412	0.1000	0.9500
HDR-VDP 75%	0.4886	0.4036	0.3000	0.6833
HDR-VDP 95%	0.5686	0.3172	0.8000	0.1333
DRI	-0.0589	0.9250	-0.1000	0.9500

Table 30: Correlation between metrics and color quality z-scores, simple scene

Metric	Pearson λ	Pearson p-value	Spearman λ	Spearman p-value
SSIM PU *	0.1326	0.8316	-0.2000	0.7833
SSIM PU +	-0.0421	0.9465	0.0000	1.0000
SSIM noPU *	0.1781	0.7745	-0.2000	0.7833
SSIM-IPT PU *	0.0595	0.9243	0.2000	0.7833
SSIM-IPT PU +	0.0650	0.9173	0.1026	0.9000
SSIM-IPT noPU *	0.0721	0.9083	0.2000	0.7833
S-CIELAB PU mean *	-0.1255	0.8406	0.2000	0.7833
S-CIELAB PU median *	-0.2992	0.6249	0.0000	1.0000
S-CIELAB PU mean +	-0.0797	0.8986	0.0000	1.0000
S-CIELAB PU median +	-0.1116	0.8581	0.0000	1.0000
S-CIELAB noPU mean *	-0.0579	0.9263	0.0000	1.0000
S-CIELAB noPU median *	-0.2287	0.7114	0.0000	1.0000
TVD PU $\lambda = 1$ *	-0.0917	0.8834	0.2000	0.7833
TVD PU $\lambda = 1$ +	-0.1855	0.7652	-0.5000	0.4500
TVD noPU $\lambda = 1$ *	-0.0553	0.9297	0.0000	1.0000
TVD PU $\lambda = 0$ *	0.8416	0.0738	0.3000	0.6833
TVD PU $\lambda = 0$ +	0.9668	0.0072	0.5000	0.4500
TVD noPU $\lambda = 0$ *	0.8782	0.0501	0.3000	0.6833
HDR-VDP 75%	0.1636	0.7926	-0.1000	0.9500
HDR-VDP 95%	0.4228	0.4782	0.1000	0.9500
DRI	0.1679	0.7873	0.3000	0.6833

Table 31: Correlation between metrics and color quality z-scores, advanced scene

Metric	Pearson λ	Pearson p-value	Spearman λ	Spearman p-value
SSIM PU *	0.3088	0.6132	-0.3000	0.6833
SSIM PU +	0.3185	0.6014	0.1000	0.9500
SSIM noPU *	0.0122	0.9844	-0.3000	0.6833
SSIM-IPT PU *	0.3232	0.5958	-0.2000	0.7833
SSIM-IPT PU +	0.3490	0.5648	-0.0513	1.0000
SSIM-IPT noPU *	0.3625	0.5488	-0.2000	0.7833
S-CIELAB PU mean *	0.1657	0.7900	-0.2000	0.7833
S-CIELAB PU median *	0.1518	0.8075	0.1000	0.9500
S-CIELAB PU mean +	0.2224	0.7192	0.1000	0.9500
S-CIELAB PU median +	0.2318	0.7075	0.1000	0.9500
S-CIELAB noPU mean *	0.2735	0.6561	0.1000	0.9500
S-CIELAB noPU median *	0.2601	0.6726	0.1000	0.9500
TVD PU $\lambda = 1$ *	0.1809	0.7709	-0.2000	0.7833
TVD PU $\lambda = 1$ +	-0.0047	0.9941	0.1000	0.9500
TVD noPU $\lambda = 1$ *	0.2633	0.6686	0.1000	0.9500
TVD PU $\lambda = 0$ *	0.3308	0.5866	0.3000	0.6833
TVD PU $\lambda = 0$ +	0.5305	0.3577	0.6000	0.3500
TVD noPU $\lambda = 0$ *	0.3975	0.5076	0.3000	0.6833
HDR-VDP 75%	0.3458	0.5687	0.4000	0.5167
HDR-VDP 95%	0.5839	0.3013	0.9000	0.0833
DRI	-0.0733	0.9068	0.2000	0.7833

Table 32: Correlation between metrics and detail quality z-scores, simple scene

Metric	Pearson λ	Pearson p-value	Spearman λ	Spearman p-value
SSIM PU *	0.2203	0.7218	0.0000	1.0000
SSIM noPU *	0.1196	0.8481	0.0000	1.0000
SSIM-IPT PU *	0.2262	0.7144	-0.1000	0.9500
SSIM-IPT PU +	0.1100	0.8602	0.0513	1.0000
SSIM-IPT noPU *	0.2485	0.6869	-0.1000	0.9500
S-CIELAB PU mean *	0.0881	0.8880	-0.1000	0.9500
S-CIELAB PU median *	0.2928	0.6326	0.2000	0.7833
S-CIELAB PU mean +	0.0556	0.9292	0.2000	0.7833
S-CIELAB PU median +	0.0996	0.8734	0.2000	0.7833
S-CIELAB noPU mean *	0.0951	0.8791	0.2000	0.7833
S-CIELAB noPU median *	0.2917	0.6339	0.2000	0.7833
TVD PU $\lambda = 1$ *	0.0409	0.9480	-0.1000	0.9500
TVD PU $\lambda = 1$ +	0.4711	0.4232	0.2000	0.7833
TVD noPU $\lambda = 1$ *	0.0963	0.8775	0.2000	0.7833
TVD PU $\lambda = 0$ *	-0.4433	0.4547	-0.6000	0.3500
TVD PU $\lambda = 0$ +	-0.5213	0.3677	-0.3000	0.6833
TVD noPU $\lambda = 0$ *	-0.4416	0.4565	-0.6000	0.3500
HDR-VDP 75%	0.2078	0.7373	0.2000	0.7833
HDR-VDP 95%	0.1549	0.8036	0.0000	1.0000
DRI	0.0269	0.9658	-0.5000	0.4500

Table 33: Correlation between metrics and detail quality z-scores, advanced scene

Metric	Pearson λ	Pearson p-value	Spearman λ	Spearman p-value
SSIM PU *	-0.0568	0.9277	-0.1000	0.9500
SSIM PU +	0.2989	0.6251	0.5000	0.4500
SSIM noPU *	-0.5516	0.3352	-0.1000	0.9500
SSIM-IPT PU *	0.1240	0.8425	0.1000	0.9500
SSIM-IPT PU +	0.3280	0.5900	0.3078	0.6667
SSIM-IPT noPU *	0.1481	0.8121	0.1000	0.9500
S-CIELAB PU mean *	0.4219	0.4792	0.1000	0.9500
S-CIELAB PU median *	0.5117	0.3782	0.5000	0.4500
S-CIELAB PU mean +	0.4895	0.4026	0.5000	0.4500
S-CIELAB PU median +	0.5228	0.3661	0.5000	0.4500
S-CIELAB noPU mean *	0.4875	0.4048	0.5000	0.4500
S-CIELAB noPU median *	0.5823	0.3029	0.5000	0.4500
TVD PU $\lambda = 1$ *	0.4469	0.4505	0.1000	0.9500
TVD PU $\lambda = 1$ +	-0.2539	0.6802	0.0000	1.0000
TVD noPU $\lambda = 1$ *	0.4546	0.4418	0.5000	0.4500
TVD PU $\lambda = 0$ *	-0.6346	0.2501	-0.6000	0.3500
TVD PU $\lambda = 0$ +	-0.3743	0.5348	0.0000	1.0000
TVD noPU $\lambda = 0$ *	-0.5704	0.3153	-0.6000	0.3500
HDR-VDP 75%	0.1588	0.7987	0.2000	0.7833
HDR-VDP 95%	0.1136	0.8557	0.3000	0.6833
DRI	-0.5147	0.3748	-0.6000	0.3500

Table 34: Correlation between metrics, simple scene.
Mapping numbers are available in Table 44.

	1		2		3		4		5	
	Rho	P-val	Rho	P-val	Rho	P-val	Rho	P-val	Rho	P-val
1	1.0000	0.0000	0.6295	0.2551	0.6267	0.2579	-0.9219	0.0259	-0.8810	0.0484
2	0.6295	0.2551	1.0000	0.0000	0.9399	0.0175	-0.7097	0.1793	-0.6824	0.2043
3	0.6267	0.2579	0.9399	0.0175	1.0000	0.0000	-0.5968	0.2880	-0.5968	0.2880
4	-0.9219	0.0259	-0.7097	0.1793	-0.5968	0.2880	1.0000	0.0000	0.9772	0.0041
5	-0.8810	0.0484	-0.6824	0.2043	-0.5968	0.2880	0.9772	0.0041	1.0000	0.0000
6	-0.9539	0.0118	-0.7082	0.1807	-0.6191	0.2655	0.9956	0.0004	0.9661	0.0074
7	-0.9522	0.0125	-0.6939	0.1937	-0.6103	0.2743	0.9955	0.0004	0.9749	0.0048
8	-0.9114	0.0312	-0.7883	0.1132	-0.6902	0.1971	0.9915	0.0009	0.9769	0.0042
9	-0.8554	0.0646	-0.7546	0.1404	-0.6835	0.2033	0.9573	0.0105	0.9907	0.0011
10	0.3978	0.5072	0.6657	0.2200	0.4276	0.4726	-0.7159	0.1739	-0.7425	0.1506
11	0.7484	0.1457	0.7471	0.1468	0.5535	0.3331	-0.9444	0.0156	-0.9279	0.0230
12	0.4086	0.4946	0.6969	0.1910	0.4645	0.4306	-0.7207	0.1695	-0.7455	0.1481
13	-0.8090	0.0973	-0.1582	0.7994	-0.2859	0.6410	0.5399	0.3476	0.4826	0.4103
14	0.6099	0.2747	0.7376	0.1549	0.5447	0.3425	-0.8616	0.0605	-0.8866	0.0450
15	0.1324	0.8319	0.5781	0.3073	0.3250	0.5936	-0.4936	0.3981	-0.5274	0.3611
16	0.5812	0.3040	0.3074	0.6149	0.5536	0.3330	-0.3914	0.5147	-0.4987	0.3925
17	-0.9023	0.0361	-0.7375	0.1549	-0.6154	0.2692	0.9982	0.0001	0.9778	0.0039
18	0.3730	0.5364	0.3959	0.5094	0.6281	0.2565	-0.2698	0.6607	-0.4151	0.4871
19	0.6052	0.2795	0.0012	0.9984	0.0847	0.8922	-0.3143	0.6065	-0.1834	0.7678
20	0.5781	0.3073	0.1991	0.7482	0.4528	0.4439	-0.3617	0.5497	-0.4625	0.4329
21	-0.9461	0.0149	-0.7505	0.1439	-0.6610	0.2245	0.9947	0.0005	0.9662	0.0074

	6		7		8		9		10	
	Rho	P-val	Rho	P-val	Rho	P-val	Rho	P-val	Rho	P-val
1	-0.9539	0.0118	-0.9522	0.0125	-0.9114	0.0312	-0.8554	0.0646	0.3978	0.5072
2	-0.7082	0.1807	-0.6939	0.1937	-0.7883	0.1132	-0.7546	0.1404	0.6657	0.2200
3	-0.6191	0.2655	-0.6103	0.2743	-0.6902	0.1971	-0.6835	0.2033	0.4276	0.4726
4	0.9956	0.0004	0.9955	0.0004	0.9915	0.0009	0.9573	0.0105	-0.7159	0.1739
5	0.9661	0.0074	0.9749	0.0048	0.9769	0.0042	0.9907	0.0011	-0.7425	0.1506
6	1.0000	0.0000	0.9990	0.0000	0.9874	0.0017	0.9454	0.0152	-0.6509	0.2342
7	0.9990	0.0000	1.0000	0.0000	0.9868	0.0018	0.9541	0.0117	-0.6526	0.2325
8	0.9874	0.0017	0.9868	0.0018	1.0000	0.0000	0.9747	0.0048	-0.7368	0.1555
9	0.9454	0.0152	0.9541	0.0117	0.9747	0.0048	1.0000	0.0000	-0.7640	0.1326
10	-0.6509	0.2342	-0.6526	0.2325	-0.7368	0.1555	-0.7640	0.1326	1.0000	0.0000
11	-0.9120	0.0309	-0.9098	0.0321	-0.9460	0.0149	-0.9222	0.0257	0.9000	0.0374
12	-0.6577	0.2276	-0.6585	0.2269	-0.7464	0.1473	-0.7715	0.1266	0.9990	0.0000
13	0.6101	0.2745	0.6100	0.2746	0.4980	0.3932	0.4243	0.4764	0.1998	0.7473
14	-0.8140	0.0936	-0.8176	0.0909	-0.8795	0.0493	-0.9014	0.0366	0.9673	0.0071
15	-0.4161	0.4859	-0.4160	0.4860	-0.5268	0.3617	-0.5655	0.3205	0.9592	0.0098
16	-0.4378	0.4609	-0.4622	0.4332	-0.4308	0.4689	-0.5322	0.3559	-0.0481	0.9388
17	0.9898	0.0012	0.9894	0.0013	0.9947	0.0005	0.9636	0.0083	-0.7522	0.1425
18	-0.2950	0.6299	-0.3210	0.5984	-0.3456	0.5689	-0.4906	0.4013	0.0773	0.9017
19	-0.3897	0.5167	-0.3742	0.5349	-0.2561	0.6775	-0.1085	0.8621	-0.3895	0.5170
20	-0.4131	0.4894	-0.4383	0.4603	-0.3852	0.5220	-0.4793	0.4140	-0.1326	0.8316
21	0.9980	0.0001	0.9961	0.0003	0.9941	0.0005	0.9534	0.0120	-0.6726	0.2135

	11		12		13		14		15	
	Rho	P-val	Rho	P-val	Rho	P-val	Rho	P-val	Rho	P-val
1	0.7484	0.1457	0.4086	0.4946	-0.8090	0.0973	0.6099	0.2747	0.1324	0.8319
2	0.7471	0.1468	0.6969	0.1910	-0.1582	0.7994	0.7376	0.1549	0.5781	0.3073
3	0.5535	0.3331	0.4645	0.4306	-0.2859	0.6410	0.5447	0.3425	0.3250	0.5936
4	-0.9444	0.0156	-0.7207	0.1695	0.5399	0.3476	-0.8616	0.0605	-0.4936	0.3981
5	-0.9279	0.0230	-0.7455	0.1481	0.4826	0.4103	-0.8866	0.0450	-0.5274	0.3611
6	-0.9120	0.0309	-0.6577	0.2276	0.6101	0.2745	-0.8140	0.0936	-0.4161	0.4859
7	-0.9098	0.0321	-0.6585	0.2269	0.6100	0.2746	-0.8176	0.0909	-0.4160	0.4860
8	-0.9460	0.0149	-0.7464	0.1473	0.4980	0.3932	-0.8795	0.0493	-0.5268	0.3617
9	-0.9222	0.0257	-0.7715	0.1266	0.4243	0.4764	-0.9014	0.0366	-0.5655	0.3205
10	0.9000	0.0374	0.9990	0.0000	0.1998	0.7473	0.9673	0.0071	0.9592	0.0098
11	1.0000	0.0000	0.9028	0.0359	-0.2385	0.6992	0.9696	0.0063	0.7471	0.1467
12	0.9028	0.0359	1.0000	0.0000	0.1941	0.7545	0.9692	0.0065	0.9581	0.0102
13	-0.2385	0.6992	0.1941	0.7545	1.0000	0.0000	-0.0444	0.9435	0.4655	0.4295
14	0.9696	0.0063	0.9692	0.0065	-0.0444	0.9435	1.0000	0.0000	0.8594	0.0619
15	0.7471	0.1467	0.9581	0.0102	0.4655	0.4295	0.8594	0.0619	1.0000	0.0000
16	0.1649	0.7910	-0.0336	0.9572	-0.6136	0.2710	0.1605	0.7965	-0.2303	0.7094
17	-0.9601	0.0095	-0.7578	0.1378	0.4924	0.3994	-0.8876	0.0445	-0.5406	0.3468
18	0.1406	0.8216	0.0963	0.8776	-0.2867	0.6400	0.2196	0.7226	-0.0224	0.9715
19	0.0400	0.9491	-0.3841	0.5232	-0.9021	0.0362	-0.1973	0.7504	-0.5985	0.2863
20	0.1091	0.8613	-0.1221	0.8449	-0.6784	0.2080	0.0855	0.8912	-0.3275	0.5906
21	-0.9215	0.0261	-0.6816	0.2051	0.5801	0.3052	-0.8306	0.0815	-0.4459	0.4517

	16		17		18		19		20	
	Rho	P-val	Rho	P-val	Rho	P-val	Rho	P-val	Rho	P-val
1	0.5812	0.3040	-0.9023	0.0361	0.3730	0.5364	0.6052	0.2795	0.5781	0.3073
2	0.3074	0.6149	-0.7375	0.1549	0.3959	0.5094	0.0012	0.9984	0.1991	0.7482
3	0.5536	0.3330	-0.6154	0.2692	0.6281	0.2565	0.0847	0.8922	0.4528	0.4439
4	-0.3914	0.5147	0.9982	0.0001	-0.2698	0.6607	-0.3143	0.6065	-0.3617	0.5497
5	-0.4987	0.3925	0.9778	0.0039	-0.4151	0.4871	-0.1834	0.7678	-0.4625	0.4329
6	-0.4378	0.4609	0.9898	0.0012	-0.2950	0.6299	-0.3897	0.5167	-0.4131	0.4894
7	-0.4622	0.4332	0.9894	0.0013	-0.3210	0.5984	-0.3742	0.5349	-0.4383	0.4603
8	-0.4308	0.4689	0.9947	0.0005	-0.3456	0.5689	-0.2561	0.6775	-0.3852	0.5220
9	-0.5322	0.3559	0.9636	0.0083	-0.4906	0.4013	-0.1085	0.8621	-0.4793	0.4140
10	-0.0481	0.9388	-0.7522	0.1425	0.0773	0.9017	-0.3895	0.5170	-0.1326	0.8316
11	0.1649	0.7910	-0.9601	0.0095	0.1406	0.8216	0.0400	0.9491	0.1091	0.8613
12	-0.0336	0.9572	-0.7578	0.1378	0.0963	0.8776	-0.3841	0.5232	-0.1221	0.8449
13	-0.6136	0.2710	0.4924	0.3994	-0.2867	0.6400	-0.9021	0.0362	-0.6784	0.2080
14	0.1605	0.7965	-0.8876	0.0445	0.2196	0.7226	-0.1973	0.7504	0.0855	0.8912
15	-0.2303	0.7094	-0.5406	0.3468	-0.0224	0.9715	-0.5985	0.2863	-0.3275	0.5906
16	1.0000	0.0000	-0.3707	0.5391	0.9274	0.0232	0.2589	0.6740	0.9901	0.0012
17	-0.3707	0.5391	1.0000	0.0000	-0.2697	0.6608	-0.2655	0.6660	-0.3337	0.5832
18	0.9274	0.0232	-0.2697	0.6608	1.0000	0.0000	-0.0923	0.8827	0.8746	0.0523
19	0.2589	0.6740	-0.2655	0.6660	-0.0923	0.8827	1.0000	0.0000	0.3424	0.5727
20	0.9901	0.0012	-0.3337	0.5832	0.8746	0.0523	0.3424	0.5727	1.0000	0.0000
21	-0.4369	0.4620	0.9921	0.0008	-0.3129	0.6082	-0.3578	0.5544	-0.4038	0.5002

	21	
	Rho	P-val
1	-0.9461	0.0149
2	-0.7505	0.1439
3	-0.6610	0.2245
4	0.9947	0.0005
5	0.9662	0.0074
6	0.9980	0.0001
7	0.9961	0.0003
8	0.9941	0.0005
9	0.9534	0.0120
10	-0.6726	0.2135
11	-0.9215	0.0261
13	0.5801	0.3052
14	-0.8306	0.0815
15	-0.4459	0.4517
16	-0.4369	0.4620
17	0.9921	0.0008
18	-0.3129	0.6082
19	-0.3578	0.5544
20	-0.4038	0.5002
21	1.0000	0.0000

Table 39: Correlation between metrics, advanced scene.
Mapping numbers are available in Table 44.

	1		2		3		4		5	
	Rho	P-val	Rho	P-val	Rho	P-val	Rho	P-val	Rho	P-val
1	1.0000	0.0000	0.4092	0.4939	0.3345	0.5822	-0.3681	0.5421	0.4178	0.4839
2	0.4092	0.4939	1.0000	0.0000	0.9874	0.0017	0.2502	0.6848	0.5460	0.3411
3	0.3345	0.5822	0.9874	0.0017	1.0000	0.0000	0.4002	0.5043	0.6267	0.2579
4	-0.3681	0.5421	0.2502	0.6848	0.4002	0.5043	1.0000	0.0000	0.6340	0.2507
5	0.4178	0.4839	0.5460	0.3411	0.6267	0.2579	0.6340	0.2507	1.0000	0.0000
6	0.1475	0.8128	0.6721	0.2139	0.7368	0.1556	0.5989	0.2859	0.8518	0.0669
7	0.6244	0.2602	0.6802	0.2063	0.6768	0.2096	0.1657	0.7900	0.8158	0.0923
8	-0.2477	0.6878	0.3472	0.5670	0.4910	0.4009	0.9875	0.0017	0.7424	0.1508
9	0.5970	0.2878	0.5926	0.2924	0.6400	0.2448	0.4398	0.4586	0.9727	0.0054
10	-0.8029	0.1019	-0.3918	0.5142	-0.3697	0.5402	0.0586	0.9254	-0.7263	0.1647
11	-0.5403	0.3472	-0.7332	0.1587	-0.7774	0.1218	-0.4757	0.4180	-0.9571	0.0106
12	-0.5077	0.3825	-0.2279	0.7124	-0.3051	0.6176	-0.4947	0.3969	-0.9152	0.0293
13	-0.6879	0.1992	-0.5288	0.3596	-0.5526	0.3341	-0.2701	0.6604	-0.9145	0.0296
14	0.2012	0.7455	-0.1923	0.7566	-0.3428	0.5722	-0.9688	0.0066	-0.7471	0.1467
15	-0.2750	0.6543	-0.4109	0.4919	-0.5161	0.3733	-0.7408	0.1521	-0.9798	0.0034
16	0.8856	0.0456	-0.0488	0.9379	-0.1163	0.8523	-0.4813	0.4118	0.2561	0.6775
17	-0.7517	0.1429	-0.2542	0.6799	-0.1158	0.8529	0.7853	0.1155	0.0515	0.9345
18	0.8842	0.0465	0.4361	0.4629	0.3371	0.5791	-0.4998	0.3912	0.3381	0.5779
19	0.2273	0.7131	0.4359	0.4632	0.5433	0.3440	0.7659	0.1311	0.9768	0.0042
20	0.8607	0.0611	-0.0596	0.9241	-0.1097	0.8607	-0.3817	0.5261	0.3506	0.5629
21	-0.8160	0.0921	-0.3086	0.6134	-0.1829	0.7684	0.6966	0.1912	-0.0928	0.8821

	6		7		8		9		10	
	Rho	P-val	Rho	P-val	Rho	P-val	Rho	P-val	Rho	P-val
1	0.1475	0.8128	0.6244	0.2602	-0.2477	0.6878	0.5970	0.2878	-0.8029	0.1019
2	0.6721	0.2139	0.6802	0.2063	0.3472	0.5670	0.5926	0.2924	-0.3918	0.5142
3	0.7368	0.1556	0.6768	0.2096	0.4910	0.4009	0.6400	0.2448	-0.3697	0.5402
4	0.5989	0.2859	0.1657	0.7900	0.9875	0.0017	0.4398	0.4586	0.0586	0.9254
5	0.8518	0.0669	0.8158	0.0923	0.7424	0.1508	0.9727	0.0054	-0.7263	0.1647
6	1.0000	0.0000	0.8444	0.0720	0.7071	0.1817	0.8221	0.0876	-0.5863	0.2988
7	0.8444	0.0720	1.0000	0.0000	0.3185	0.6014	0.9064	0.0339	-0.9043	0.0350
8	0.7071	0.1817	0.3185	0.6014	1.0000	0.0000	0.5705	0.3152	-0.0912	0.8841
9	0.8221	0.0876	0.9064	0.0339	0.5705	0.3152	1.0000	0.0000	-0.8558	0.0643
10	-0.5863	0.2988	-0.9043	0.0350	-0.0912	0.8841	-0.8558	0.0643	1.0000	0.0000
11	-0.8860	0.0454	-0.9223	0.0257	-0.6060	0.2786	-0.9789	0.0037	0.7901	0.1117
12	-0.6016	0.2831	-0.6719	0.2141	-0.5888	0.2962	-0.9015	0.0366	0.7570	0.1384
13	-0.7652	0.1317	-0.9375	0.0186	-0.4124	0.4902	-0.9802	0.0033	0.9409	0.0171
14	-0.6112	0.2734	-0.2633	0.6686	-0.9752	0.0047	-0.5737	0.3119	0.1232	0.8436
15	-0.8081	0.0980	-0.6984	0.1896	-0.8266	0.0844	-0.9125	0.0306	0.6200	0.2646
16	-0.0907	0.8846	0.4134	0.4890	-0.3890	0.5175	0.4249	0.4757	-0.7464	0.1474
17	0.0176	0.9776	-0.4738	0.4201	0.6789	0.2076	-0.1804	0.7716	0.6065	0.2781
18	0.2946	0.6304	0.7526	0.1421	-0.3585	0.5535	0.5467	0.3404	-0.8556	0.0644
19	0.8430	0.0729	0.7041	0.1844	0.8508	0.0676	0.9032	0.0356	-0.5962	0.2886
20	-0.0055	0.9930	0.4645	0.4306	-0.2851	0.6420	0.5024	0.3883	-0.7920	0.1102
21	-0.0925	0.8823	-0.5837	0.3014	0.5755	0.3100	-0.3203	0.5993	0.7158	0.1739

	11		12		13		14		15	
	Rho	P-val	Rho	P-val	Rho	P-val	Rho	P-val	Rho	P-val
1	-0.5403	0.3472	-0.5077	0.3825	-0.6879	0.1992	0.2012	0.7455	-0.2750	0.6543
2	-0.7332	0.1587	-0.2279	0.7124	-0.5288	0.3596	-0.1923	0.7566	-0.4109	0.4919
3	-0.7774	0.1218	-0.3051	0.6176	-0.5526	0.3341	-0.3428	0.5722	-0.5161	0.3733
4	-0.4757	0.4180	-0.4947	0.3969	-0.2701	0.6604	-0.9688	0.0066	-0.7408	0.1521
5	-0.9571	0.0106	-0.9152	0.0293	-0.9145	0.0296	-0.7471	0.1467	-0.9798	0.0034
6	-0.8860	0.0454	-0.6016	0.2831	-0.7652	0.1317	-0.6112	0.2734	-0.8081	0.0980
7	-0.9223	0.0257	-0.6719	0.2141	-0.9375	0.0186	-0.2633	0.6686	-0.6984	0.1896
8	-0.6060	0.2786	-0.5888	0.2962	-0.4124	0.4902	-0.9752	0.0047	-0.8266	0.0844
9	-0.9789	0.0037	-0.9015	0.0366	-0.9802	0.0033	-0.5737	0.3119	-0.9125	0.0306
10	0.7901	0.1117	0.7570	0.1384	0.9409	0.0171	0.1232	0.8436	0.6200	0.2646
11	1.0000	0.0000	0.8064	0.0992	0.9404	0.0173	0.5653	0.3206	0.8820	0.0478
12	0.8064	0.0992	1.0000	0.0000	0.8805	0.0487	0.6819	0.2048	0.9252	0.0243
13	0.9404	0.0173	0.8805	0.0487	1.0000	0.0000	0.4273	0.4730	0.8365	0.0774
14	0.5653	0.3206	0.6819	0.2048	0.4273	0.4730	1.0000	0.0000	0.8527	0.0664
15	0.8820	0.0478	0.9252	0.0243	0.8365	0.0774	0.8527	0.0664	1.0000	0.0000
16	-0.2874	0.6392	-0.5087	0.3815	-0.5585	0.3279	0.2599	0.6728	-0.1707	0.7837
17	0.1661	0.7894	-0.0263	0.9666	0.3436	0.5713	-0.7013	0.1869	-0.2270	0.7135
18	-0.5219	0.3670	-0.3412	0.5742	-0.6789	0.2076	0.3710	0.5387	-0.1666	0.7889
19	-0.8847	0.0462	-0.8958	0.0397	-0.8218	0.0879	-0.8628	0.0597	-0.9971	0.0002
20	-0.3558	0.5567	-0.6054	0.2793	-0.6281	0.2565	0.1483	0.8119	-0.2793	0.6491
21	0.2974	0.6270	0.1183	0.8498	0.4771	0.4164	-0.5902	0.2948	-0.0823	0.8953

	16		17		18		19		20	
	Rho	P-val	Rho	P-val	Rho	P-val	Rho	P-val	Rho	P-val
1	0.8856	0.0456	-0.7517	0.1429	0.8842	0.0465	0.2273	0.7131	0.8607	0.0611
2	-0.0488	0.9379	-0.2542	0.6799	0.4361	0.4629	0.4359	0.4632	-0.0596	0.9241
3	-0.1163	0.8523	-0.1158	0.8529	0.3371	0.5791	0.5433	0.3440	-0.1097	0.8607
4	-0.4813	0.4118	0.7853	0.1155	-0.4998	0.3912	0.7659	0.1311	-0.3817	0.5261
5	0.2561	0.6775	0.0515	0.9345	0.3381	0.5779	0.9768	0.0042	0.3506	0.5629
6	-0.0907	0.8846	0.0176	0.9776	0.2946	0.6304	0.8430	0.0729	-0.0055	0.9930
7	0.4134	0.4890	-0.4738	0.4201	0.7526	0.1421	0.7041	0.1844	0.4645	0.4306
8	-0.3890	0.5175	0.6789	0.2076	-0.3585	0.5535	0.8508	0.0676	-0.2851	0.6420
9	0.4249	0.4757	-0.1804	0.7716	0.5467	0.3404	0.9032	0.0356	0.5024	0.3883
10	-0.7464	0.1474	0.6065	0.2781	-0.8556	0.0644	-0.5962	0.2886	-0.7920	0.1102
11	-0.2874	0.6392	0.1661	0.7894	-0.5219	0.3670	-0.8847	0.0462	-0.3558	0.5567
12	-0.5087	0.3815	-0.0263	0.9666	-0.3412	0.5742	-0.8958	0.0397	-0.6054	0.2793
13	-0.5585	0.3279	0.3436	0.5713	-0.6789	0.2076	-0.8218	0.0879	-0.6281	0.2565
14	0.2599	0.6728	-0.7013	0.1869	0.3710	0.5387	-0.8628	0.0597	0.1483	0.8119
15	-0.1707	0.7837	-0.2270	0.7135	-0.1666	0.7889	-0.9971	0.0002	-0.2793	0.6491
16	1.0000	0.0000	-0.6903	0.1970	0.7711	0.1269	0.1085	0.8621	0.9911	0.0010
17	-0.6903	0.1970	1.0000	0.0000	-0.9218	0.0259	0.2474	0.6883	-0.6269	0.2577
18	0.7711	0.1269	-0.9218	0.0259	1.0000	0.0000	0.1437	0.8177	0.7476	0.1463
19	0.1085	0.8621	0.2474	0.6883	0.1437	0.8177	1.0000	0.0000	0.2184	0.7242
20	0.9911	0.0010	-0.6269	0.2577	0.7476	0.1463	0.2184	0.7242	1.0000	0.0000
21	-0.7443	0.1491	0.9891	0.0014	-0.9682	0.0068	0.1050	0.8666	-0.6959	0.1918

	21	
	Rho	P-val
1	-0.8160	0.0921
2	-0.3086	0.6134
3	-0.1829	0.7684
4	0.6966	0.1912
5	-0.0928	0.8821
6	-0.0925	0.8823
7	-0.5837	0.3014
8	0.5755	0.3100
9	-0.3203	0.5993
10	0.7158	0.1739
11	0.2974	0.6270
12	0.1183	0.8498
13	0.4771	0.4164
14	-0.5902	0.2948
15	-0.0823	0.8953
16	-0.7443	0.1491
17	0.9891	0.0014
18	-0.9682	0.0068
19	0.1050	0.8666
20	-0.6959	0.1918
21	1.0000	0.0000

Table 44: Mapping of names for metric correlations.

Metric	Value
SSIM PU *	15
SSIM PU +	14
SSIM noPU *	13
SSIM-IPT PU *	12
SSIM-IPT PU +	11
SSIM-IPT noPU *	10
S-CIELAB PU mean *	8
S-CIELAB PU median *	9
S-CIELAB PU mean +	6
S-CIELAB PU median +	7
S-CIELAB noPU mean *	4
S-CIELAB noPU median *	5
TVD PU $\lambda = 0$ *	20
TVD PU $\lambda = 0$ +	18
TVD noPU $\lambda = 0$ *	16
TVD noPU $\lambda = 1$ *	21
TVD PU $\lambda = 1$ *	19
TVD PU $\lambda = 1$ +	17
HDR-VDP 75%	2
HDR-VDP 95%	3
DRI	1

Bibliography

[1] Debevec, P. E. & Malik, J. 1997. Recovering high dynamic range radiance maps from photographs. In *Proceedings of the 24th annual conference on Computer graphics and interactive techniques*, SIGGRAPH '97, 369–378, New York, NY, USA. ACM Press/Addison-Wesley Publishing Co.

[2] Mann, S. & Picard, R. W. 1995. On being 'undigital' with digital cameras: Extending dynamic range by combining differently exposed pictures. In *Proceedings of IST*, 442–448.

[3] Mitsunaga, T. & Nayar, S. 1999. Radiometric self calibration. In *Computer Vision and Pattern Recognition, 1999. IEEE Computer Society Conference on.*, volume 1, 2 vol. (xxiii+637+663).

[4] Reinhard, E., Ward, G., Pattanaik, S., Debevec, P., Heindrich, W., & Myszkowski, K. 2010. *High Dynamic Range Imaging 2nd Edition, Acquisition, Display and Image-Based Lighting*. Morgan Kaufman, 2nd edition.

[5] Kang, S. B., Uyttendaele, M., Winder, S., & Szeliski, R. July 2003. High dynamic range video. *ACM Trans. Graph.*, 22, 319–325.

[6] Khan, E., Akyuz, A., & Reinhard, E. oct. 2006. Ghost removal in high dynamic range images. In *Image Processing, 2006 IEEE International Conference on*, 2005 –2008.

[7] Ward, G. 1995. Making global illumination user-friendly. In *Proceedings of Eurographics Workshop on Rendering*. Citeseer.

[8] Kainz, F. & Bogart, R. 2009. Technical introduction to OpenEXR. *Industrial light and magic*.

[9] Tumblin, J. & Rushmeier, H. E. 1339. Tone reproduction for computer generated images. In *IEEE Computer Graphics and Applications*, volume Vol. 13, No. 6, pp. 42–48.

[10] Miller, G. & Hoffman, C. 1984. Illumination and reflection maps: Simulated objects in simulated and real environments. In *SIGGRAPH 84 Advanced Computer Graphics Animation seminar notes*, volume 190.

[11] Ward, G. *A contrast-based scalefactor for luminance display*, 415–421. Academic Press Professional, Inc., San Diego, CA, USA, 1994.

[12] Reinhard, E. & Devlin, K. January 2005. Dynamic range reduction inspired by photoreceptor physiology. *IEEE Transactions on Visualization and Computer Graphics*, 11, 13–24.

[13] Chiu, K., Herf, M., Shirley, P., Swamy, S., Wang, C., & Zimmerman, K. 1993. Spatially nonuniform scaling functions for high contrast images. In *In Proceedings of Graphics Interface '93*, 245–253.

[14] Reinhard, E., Stark, M., Shirley, P., & Ferwerda, J. 2002. Photographic tone reproduction for digital images. In *Proceedings of the 29th annual conference on Computer graphics and interactive techniques*, SIGGRAPH '02, 267–276, New York, NY, USA. ACM.

[15] Durand, F. & Dorsey, J. 2002. Fast bilateral filtering for the display of high-dynamic-range images. In *Proceedings of the 29th annual conference on Computer graphics and interactive techniques*, SIGGRAPH '02, 257–266, New York, NY, USA. ACM.

[16] Kolås, Ø., Farup, I., & Rizzi, A. 2011 (accepted). Stress: A framework for spatial color algorithms. *Journal of Imaging Science and Technology*.

[17] Jobson, D. J., Zia-ur, R., & Glenn A., W. 1996. Retinex image processing: Improved fidelity to direct visual observation.

[18] Fattal, R., Lischinski, D., & Werman, M. July 2002. Gradient domain high dynamic range compression. *ACM Trans. Graph.*, 21, 249–256.

[19] Fairchild, M. D. & Johnson, G. M. 2002. Meet icam: A next-generation color appearance model. In *IST/SID 10 th Color Imaging Conference*, 33–38.

[20] Pattanaik, S. N., Ferwerda, J. A., Fairchild, M. D., & Greenberg, D. P. 1998. A multiscale model of adaptation and spatial vision for realistic image display. In *Proceedings of the 25th annual conference on Computer graphics and interactive techniques*, SIGGRAPH '98, 287–298, New York, NY, USA. ACM.

[21] Mantiuk, R., Myszkowski, K., & Seidel, H.-P. July 2006. A perceptual framework for contrast processing of high dynamic range images. *ACM Trans. Appl. Percept.*, 3, 286–308.

[22] Stevens, J. C. & Stevens, S. S. Mar 1963. Brightness function : Effects of adaptation. *J. Opt. Soc. Am.*, 53(3), 375–385.

[23] CIE. An analytic model for describing the influence of lighting parameters upon visual performance. Technical Report Vol. 1, CIE, 1981.

[24] Reinhard, E., Ward, G., Pattanaik, S., & Debevec, P. December 2005. *High Dynamic Range Imaging: Acquisition, Display and Image-Based Lighting*. Morgan Kaufmann Publishers.

[25] Naka, K. I. & Rushton, W. A. H. 1966. S-potential and dark adaptation in fish. *Journal of Physiology*, 185, 587–599.

[26] Hood, D. C., Finkelstein, M. A., & Buckingham, E. 1979. Psychophysical tests of models of the response function. *Vision Research*, 19(4), 401 – 406. Visual Sensitivity and Adaption, The British Photobiology Society and The Association for Research in Vision and Opthalmology Inc.

[27] Tomasi, C. & Manduchi, R. 1998. Bilateral filtering for gray and color images. *Computer Vision, IEEE International Conference on*, 0, 839.

[28] Tumblin, J., Hodgins, J. K., & Guenter, B. K. January 1999. Two methods for display of high contrast images. *ACM Trans. Graph.*, 18, 56–94.

[29] Schlick, C. 1994. Quantization techniques for visualization of high dynamic range pictures. In *Fifth Eurographics Workshop on Rendering*.

[30] Ashikhmin, M. 2002. A tone mapping algorithm for high contrast images. In *Proceedings of the 13th Eurographics workshop on Rendering*, EGRW '02, 145–156, Aire-la-Ville, Switzerland, Switzerland. Eurographics Association.

[31] Peli, E. 1990. Contrast in complex images. *JOSA A*, 7(10), 2032–2040.

[32] Wilson, H. R. 1980. A transducer function for threshold and suprathreshold human vision. *Biological Cybernetics*, 38, 171–178. 10.1007/BF00337406.

[33] Press, W. H., Vetterling, W. T., Teukolsky, S. A., & Flannery, B. P. 2002. *Numerical Recipes in C++: the art of scientific computing*. Cambridge University Press, New York, NY, USA, 2nd edition.

[34] Thurstone, L. 1927. A law of comparative judgment. *Psychological review*, 34(4), 273.

[35] Aydın, T. O., Mantiuk, R., & Seidel, H.-P. January 2008. Extending quality metrics to full dynamic range images. In *Human Vision and Electronic Imaging XIII*, Proceedings of SPIE, 6806–10, San Jose, USA.

[36] Daly, S. 1993. The visible differences predictor: an algorithm for the assessment of image fidelity.

[37] Irawan, P., Ferwerda, J. A., & Marschner, S. R. 2005. Perceptually based tone mapping of high dynamic range image streams. In *Proceedings of the Eurographics Symposium on Rendering*, 231–242.

[38] CadÌk, M., Wimmer, M., Neumann, L., & Artusi, A. 2008. Evaluation of hdr tone mapping methods using essential perceptual attributes. *Computers Graphics*, 32(3), 330–349.

[39] Lehmann, E. & Casella, G. 1998. *Theory of Point Estimation*. Springer New York, 2nd edition.

[40] Huynh-Thu, Q. & Ghanbari, M. 19 2008. Scope of validity of psnr in image/video quality assessment. *Electronics Letters*, 44(13), 800 –801.

[41] CIE. Industrial colour-difference evaluation. Technical Report 116, CIE, 1995.

[42] Wang, Z. & Bovik, A. March 2002. A universal image quality index. *Signal Processing Letters, IEEE*, 9(3), 81 –84.

[43] Wang, Z., Bovik, A., Sheikh, H., & Simoncelli, E. 2004. Image quality assessment: from error visibility to structural similarity. *Image Processing, IEEE Transactions on*, 13(4), 600 –612.

[44] Mantiuk, R., Daly, S., Myszkowski, K., & Seidel, H. 2005. Predicting visible differences in high dynamic range images: model and its calibration. In *Proc. SPIE*, volume 5666, 204–214. Citeseer.

[45] Deeley, R., Drasdo, N., & Charman, W. 1991. A simple parametric model of the human ocular modulation transfer function. *Ophthalmic and Physiological Optics*, 11(1), 91–93.

[46] Zhang, X. & Wandell, B. 1996. A spatial extension of cielab for digital color image reproduction. *Soc. Inform. Display 96 Digest*, 731 – 734.

[47] Pedersen, M., Simone, G., Gong, M., & Farup, I. A total variation based color image quality metric with perceptual contrast filtering. To be published in The International Conference on Image Processing 2011.

[48] Rudin, L., Osher, S., & Fatemi, E. 1992. Nonlinear total variation based noise removal algorithms. *Physica D: Nonlinear Phenomena*, 60(1-4), 259–268.

[49] Peli, E. 2001. Contrast sensitivity function and image discrimination. *JOSA A*, 18(2), 283–293.

[50] Bonnier, N., Schmitt, F., Brettel, H., & Berche, S. 2006. Evaluation of spatial gamut mapping algorithms. In *14th Color Imaging Conference*, volume 14, 56–61. IST/SID.

[51] Ebner, F. & Fairchild, M. 1998. Development and testing of a color space (ipt) with improved hue uniformity. In *IS&T/SID 6th Color Imaging Conference*, 8–13.

[52] Reinhard, E., Kahn, E. A., Akyüz, A. O., & Johnson, G. 2008. *Color Imaging: Fundamentals and Applications*. A. K. Peters.

[53] Aydin, T., Mantiuk, R., Myszkowski, K., & Seidel, H. 2008. Dynamic range independent image quality assessment. In *ACM SIGGRAPH 2008 papers*, 1–10. ACM.

[54] CIE. Guidelines for the evaluation of gamut mapping algorithms. Technical Report 156, Commision internationale de l'eclairage, 2004.

[55] Light Craft Workshop. Light Craft Workshop - Fader ND adjustable ND Filter (ND2 400) mark II. http://www.lightcraftworkshop.com/site/page1000.aspx. Last Checked: July 2011.

[56] Anastasia, D. Luminance HDR aka Qtpfsgui: an Open Source Workflow for HDR Imaging. http://qtpfsgui.sourceforge.net/. Last Checked: July 2011.

[57] Marc Mels, C. B. Picturenaut hdri. http://www.hdrlabs.com/picturenaut/. Last Checked: July 2011.

[58] Kuang, J., Yamaguchi, H., Johnson, G., & Fairchild, M. 2004. Testing hdr image rendering algorithms. *Society for Imaging Science and Technology*.

[59] Malakauskas, M. & Montvilas, G. Panel testing for image quality. Batchelor Thesis, 05 2003.

[60] Andrich, D. 1978. Relationships between the thurstone and rasch approaches to item scaling. *Applied Psychological Measurement*, 2(3), 451.

[61] Johnson, T. & Green, P. 2002. Viewing conditions. *CIE Draft No: 18*.

[62] Pedersen, M. & Hardeberg, J. Y. Survey of full-reference image quality metrics. Technical Report 5, Gjøvik, Norway, June 2009. ISSN: 1890-520X.

[63] Zhang, X. S-CIELAB: A Spatial Extension to the CIE L*a*b* ΔE Color Difference Metric. http://white.stanford.edu/ brian/scielab/. Last Checked: July 2011.

[64] Mantiuk, R. pfstools - high dynamic range images and video. http://pfstools.sourceforge.net/. Last Checked: July 2011.

[65] Aydın, T. O. Dynamic range independent quality assessment. http://drim.mpi-sb.mpg.de/. Last Checked: July 2011.

[66] Montag, E. D. 2006. Empirical formula for creating error bars for the method of paired comparison. *Journal of Electronic Imaging*, 15(1), 010502.

[67] Islam, A. T. Spatio-temporal colour correction of strongly degraded films. Master's thesis, Høgskolen I Gjøvik, 2010.

[68] Ramamurthy, V., Narendran, N., Freyssinier, J., Raghavan, R., & Boyce, P. 2004. Determining contrast sensitivity functions for monochromatic light emitted by high-brightness leds. In *Conference on Solid State Lighting*, volume 5187, 294–300.

[69] Mullen, K. 1985. The contrast sensitivity of human colour vision to red-green and blue-yellow chromatic gratings. *The Journal of Physiology*, 359(1), 381.