Czech Technical University in Prague
Faculty of Electrical Engineering
Department of Computer Science and Engineering



Perceptually Based Image Quality Assessment and Image Transformations

Martin Čadík

A doctoral thesis submitted to
the Faculty of Electrical Engineering, Czech Technical University in Prague,
in partial fulfilment of the requirements for the degree of Doctor.

Ph.D. programme: Electrical Engineering and Information Technology
Branch of study: Information Science and Computer Engineering

January 2008

Thesis Supervisor:
Prof. Ing. Pavel Slavík, CSc.
Department of Computer Science and Engineering
Faculty of Electrical Engineering
Czech Technical University in Prague
Karlovo nám. 13
121 35 Praha 2
Czech Republic

Copyright © 2008 by Martin Čadík

Abstract

Computational approaches mimicking perceptual properties of the human visual system (HVS) have been successfully applied in various fields of computer graphics and digital imaging. The knowledge about the HVS is continually developing, however there are still many open questions and hypotheses. Naturally, we are far from having an accurate computational model of the HVS. It is therefore of particular importance and necessity to conduct experimental subjective analyses of the methods that incorporate HVS models to validate and evaluate them properly. Besides exposing strengths and weaknesses of inquired methods, the experimental evaluations on human subjects also attain a deeper knowledge of the examined field which can advance current state of the art and which can furthermore result in proposals of new approaches.

The fields that traditionally utilize computational models of the HVS and which are concerned in this thesis are a) image quality assessment, b) high dynamic range tone mapping, and c) color image to grayscale conversions.

a) Image quality assessment metrics aim to predict the difference between images as perceived by human subjects. We present results of an experimental subjective evaluation of two principal approaches to image quality assessment (traditional error sensitivity based approach, and structural similarity based approach). The results show that the structural similarity outperforms the traditional approach for involved input stimuli.

b) The field of tone mapping has witnessed a lot of research effort to solve the problem of displaying images with high dynamic range (HDR) of luminance on ordinary output devices. We present a study about the effect of image attributes (contrast, brightness, etc.) in the HDR tone mapping. Furthermore, we present experimental subjective evaluations of global and local tone mapping approaches. Our results imply that proper global part of a tone mapping method is essential to obtain good perceptual results for typical real world scenes.

Motivated by implications of conducted studies, we propose a novel simple yet powerful general hybrid approach to tone mapping. In our approach, we apply a global tone mapping method first to reproduce overall image attributes correctly and we construct an enhancement map to guide a local operator to the critical areas that deserve an enhancement. The new approach is general and can be easily tailored to miscellaneous goals of tone mapping. An implementation of the proposed approach produces plausible results, it is easy to implement, and fast to compute.

c) Simple color image to grayscale conversions potentially disrupt chromatic informations contained in color images, and existing advanced conversions are eminently computationally intensive. We propose novel perceptually convincing local color to grayscale conversion techniques that operate in gradient domain. The novel methods produce perceptually justifiable results, and they have linear complexity in the number of pixels, which makes them suitable for high-resolution images.

Keywords:

human perception, human visual system, experimental evaluation, image quality assessment, image comparison, HDR, tone mapping, color to gray

Acknowledgements

I would like to express my gratitude to my supervisor prof. Pavel Slavík. His initial encouragement, continuous support and constant pressure were essential to the coming about of this thesis. I would also like to thank to prof. Jiří Žára for leading the computer graphics group at the CTU in Prague and for keeping our lab running. Further, I would like to thank to prof. László Neumann for his constant enthusiasm, encouragement, and for sharing his valuable experience.

Thanks also go to all my colleagues and friends in computer graphics group in Prague, especially to Jiří Bittner, Dan Sýkora, Jaroslav Křivánek, Jaroslav Sloup, Roman Berka, Adam Sporka, Martin Klíma, Vladimír Štěpán, David Ambrož, Vlastimil Havran and others, for their ideas and comments, and for their friendship. I would like to thank to the staff of our department and especially to Josef Kolář and Pavel Tvrdík for providing a pleasant environment for my research. I am much in debt to people from the CG lab at the TU Vienna, namely Werner Purgathofer, Eduard Gröller, Michael Wimmer, Alessandro Artusi and Attila Neumann, for being great hosts during my research visits in Vienna. Thanks also go to prof. Walter Paetzold for his kindness and hospitality during my stage at the HTW Dresden and to prof. Rafael García and his colleagues for their support and nice courteousness during my visit at the UdG in Girona. I have to thank to prof. Jiří Habel and to prof. Přemysl Maydl for their valuable time and for a lot of useful discussions. I would like to thank to lots of people that were kindly and peacefully participating in my subjective experiments. Thanks to all proofreaders of the manuscript.

Special thanks to my family and to my girlfriend Jana without whose infinite patience this thesis would never have been finished, and thanks to God for all the mistakes I have made so far.

My work was supported by the Ministry of Education, Youth and Sports of the Czech Republic under research programs No.: MSM-212300014, MSM-6840770014 (Research in the area of information technologies and communications) and LC-06008 (Center for Computer Graphics), by the Czech Technical University in Prague – grant No. CTU-0408813, and by the Aktion Kontakt OE/CZ grant No. 48p11.

Dedication

To Jana and to my parents for their love and patience.

Contents

1 Introduction **1**

1.1 Motivation 1

1.1.1 Image Quality Assessment 1

1.1.2 High Dynamic Range Tone Mapping 2

1.1.3 Color to Grayscale Conversions 3

1.2 Contributions 4

1.3 Structure of the Thesis 4

2 Human Visual System **7**

2.1 Physical Structure 7

2.1.1 The Human Eye 7

2.1.2 The Retina 8

2.1.3 Visual Cortex 10

2.1.4 The Physiological Pathways Hypothesis 11

2.2 Visual Perception 12

2.2.1 Adaptation 12

2.2.2 Ambiguous Figures 13

2.2.3 Visual Completion 13

2.2.4 Impossible Objects 14

2.2.5 Classification 14

2.2.6 Attention and Consciousness 14

2.3 Image Processing Theories 15

2.3.1 Line and Edge Detection Theory 15

2.3.2 Spatial Frequency Theory 16

3 Perceptual Evaluation of Image Quality Metrics **19**

3.1 Introduction 19

3.2 Traditional Perceptual Image Quality Assessment 20

3.2.1 General Framework of Perceptual Quality Metrics 20

3.2.2 Traditional Perceptual Image Quality Metrics 21

3.2.2.1 Pixel-Based Mathematical Metrics 21

3.2.2.2 Model after Mannos and Sakrison 21

3.2.2.3 Model after Gervais 22

3.2.2.4 Visible Differences Predictor . . . 22
3.2.2.5 Perceptual Distortion Measure by Teo and Heeger . . . 25
3.2.2.6 Visual Discrimination Model . . . 25
3.2.2.7 Gabor Pyramid Model of the HVS . . . 25
3.2.2.8 Stochastic Color Image Difference by Neumann et al. . . . 27
3.2.2.9 Wavelet Visible Difference Predictor . . . 28
3.2.2.10 S-CIELab and ST-CIELab . . . 28
3.2.2.11 Multistage Perceptual Quality Assessment Model . . . 28
3.2.3 Summary . . . 29
3.3 Structural Similarity Based Image Quality Assessment . . . 30
3.3.1 The Structural SIMilarity Index . . . 30
3.3.2 An SVD Based Gray-Scale Image Quality . . . 32
3.3.3 Gradient-Based Structural Similarity . . . 33
3.4 Selection of Models for Comparison . . . 33
3.5 Measurements . . . 34
3.5.1 Subjective Testing . . . 34
3.5.2 Objective Testing . . . 34
3.5.3 Test Results . . . 38
3.6 Discussion . . . 39
3.6.1 Quality Assessment Performances . . . 39
3.6.2 ROI Quality Assessment . . . 40
3.6.3 Advantages of Evaluated Models . . . 41
3.7 Summary . . . 41

4 Perceptual Evaluation of Tone Mapping Methods 43
4.1 Introduction . . . 43
4.2 Previous Work . . . 44
4.2.1 Experimental Evaluations of Tone Mapping Methods . . . 44
4.2.2 Evaluations using HDR Displays . . . 45
4.2.3 Other Related Studies . . . 46
4.2.4 Our Approach . . . 46
4.3 Overall Image Quality . . . 47
4.4 Image Attributes . . . 48
4.4.1 Brightness . . . 48
4.4.2 Contrast . . . 49

4.4.3 Reproduction of Colors . . . 49
4.4.4 Reproduction of Details . . . 50
4.4.5 Artifacts . . . 52
4.4.6 Special Attributes . . . 52
4.5 Attribute Relationships . . . 53
4.6 Subjective Perceptual Studies . . . 55
4.6.1 Subjective Testing Setup . . . 56
4.7 Results and Discussion . . . 58
4.7.1 Effects of Input Scenes and Methods . . . 62
4.7.2 Effect of the Experimental Setup . . . 63
4.7.3 Evaluation of HDR Tone Mapping Methods . . . 65
4.7.3.1 Overall Results . . . 65
4.7.3.2 Comparison of the Two Experiments . . . 67
4.7.3.3 Comparison of the Results for Input Scenes . . . 69
4.7.4 Overall Image Quality and Relationships of Attributes . . . 69
4.7.5 Comparison to Other Studies . . . 72
4.8 Summary . . . 73

5 Hybrid Approach to Tone Mapping 75
5.1 Introduction . . . 75
5.2 Related Work . . . 77
5.2.1 Time-Dependent Tone Mapping . . . 77
5.2.2 Linear Tone Mapping . . . 78
5.3 Hybrid Tone Mapping . . . 78
5.3.1 Construction of Enhancement Map . . . 80
5.4 Use Cases and Results . . . 80
5.4.1 Fast Perceptually Plausible Approach . . . 82
5.4.2 Time-Dependent Hybrid Mapping . . . 82
5.4.3 Cognitive Approach . . . 85
5.4.4 Performance Results . . . 87
5.5 Summary . . . 87

6 Conversion of Color Images to Grayscale 89
6.1 Introduction and Related Work . . . 89
6.2 A Simple New CIELab-Based Formula . . . 91

6.3 Color-to-Grayscale Conversion Based on the Coloroid System 92
6.3.1 The Coloroid System 92
6.3.2 Observation Conditions of the Coloroid Experiments 94
6.3.3 Perceptually-Equivalent Color and Grayscale Attributes 95
6.3.4 The Coloroid-Based Formula 96
6.4 Gradient Inconsistency Correction 97
6.4.1 The Basic Idea 97
6.4.2 Gradient Inconsistency Correction Method 98
6.4.3 The Algorithm 100
6.5 Results and Discussion 102
6.6 Summary 103

7 Conclusions **107**
7.1 Summary of Results 107
7.1.1 Perceptual Evaluation of Image Quality Metrics 107
7.1.2 Perceptual Evaluation of Tone Mapping Methods 108
7.1.3 Hybrid Approach to Tone Mapping 108
7.1.4 Perceptually-Plausible Conversion of Color Images to Grayscale 109
7.2 Suggestions for Further Research 109
7.2.1 Perceptual Evaluation of Image Quality Metrics 109
7.2.2 Perceptual Evaluation of Tone Mapping Methods 110
7.2.3 Hybrid Approach to Tone Mapping 110
7.2.4 Perceptually-Plausible Conversion of Color Images to Grayscale 111

8 Bibliography **113**

A Statistical Tests **125**
A.1 ANOVA 125
A.2 Correlation Coefficients 127
A.2.1 Pearson Product-Moment Correlation Coefficient (r) 127
A.2.2 Kendall Tau Coefficient (t) 127
A.2.3 Spearman's Rho (r_s) 128

B Multiscale Transforms **129**

List of Figures

1.1 Full-reference image quality metrics. 2
1.2 High dynamic range tone mapping. 3
1.3 Color image to grayscale conversion. 3

2.1 A cross section of the human eye. 8
2.2 The human visual system. 9
2.3 The human retina. 10
2.4 Receptive field structure of ganglion cells. 11
2.5 Schematic diagram of the visual pathways hypothesis. 12
2.6 Ambiguous figures. 13
2.7 An impossible object. 14
2.8 Contrast sensitivity function. 17
2.9 A theoretical hypothesis about the architecture of image processing. 18

3.1 Block diagram of typical discrimination/quality metric. 20
3.2 Block diagram of the Visible Differences Predictor. 23
3.3 Cortex transform. 24
3.4 Block diagram of the Visual Discrimination Model. 26
3.5 Block diagram of the Gabor Pyramid Model. 27
3.6 Block diagram of the Wavelet Visible Differences Predictor. 28
3.7 Block diagram of the MPQA model. 29
3.8 Diagram of the structural similarity (SSIM) quality measure. 31
3.9 Image quality assessment using VDP and SSIM (uniform compression). . . . 37
3.10 Image quality assessment using VDP and SSIM (ROI compression). 38
3.11 Quality predictions compared to subjective mean opinion scores (MOS). . . . 39
3.12 Quality assessment performances of the VDP, and the SSIM models. 40
3.13 ROI functions. 41

4.1 Three possible approaches to tone mapping. 48
4.2 Peli's local band-limited contrast on three different spatial resolutions. 50
4.3 Simulation of color sensitivity. 51
4.4 Reproduction of details in a very bright area. 51
4.5 Halo artifacts and oversaturation. 52
4.6 Bloom (veiling luminance) simulation. 53

4.7 The relationships between image attributes. 54
4.8 Example of subjective perceptual experiments setups. 55
4.9 Luminance histograms of the experimental HDR images. 56
4.10 Overall accuracy scores for all examined TM methods. 66
4.11 Average overall image quality and average Mahalanobis distances. 67
4.12 Accuracy scores for rating experiment and for ranking experiment. 68
4.13 Relations of the ranking experiment and rating experiment. 69
4.14 Overall image quality scores for each input scene. 70

5.1 Illustration of strengths and weaknesses of global and local TM methods. . . 76
5.2 Overview of the hybrid tone mapping approach. 79
5.3 Perceptually plausible hybrid approach. 79
5.4 Tone mapping transforms. 81
5.5 Results of the new hybrid approach. 83
5.6 Results of the new hybrid approach. 84
5.7 Cognitive hybrid approach. 85
5.8 Further results of the new hybrid approach. 86

6.1 Comparison of the new method with the CIE Y and Gooch et al. 92
6.2 Comparison of the new method with existing approaches. 93
6.3 Limit-colors of the Coloroid system. 94
6.4 Shape of the Coloroid gamut. 95
6.5 The three steps of the proposed adaptive color-to-grayscale conversion. 98
6.6 Direct integration of inconsistent and consistent gradient fields. 99
6.7 Inconsistent gradient field, orthogonal projection. 99
6.8 Isoluminant image, CIE Y, result of the new method. 103
6.9 Comparison of the results obtained using two proposed formulas. 104
6.10 Result of the adaptive color-to-grayscale conversion (Coloroid formula). . . . 105
6.11 Maximum error selection using the Fibonacci heap. 106

7.1 The simultaneous contrast effect. 111

List of Abbreviations

All abbreviations are defined when first used in the text, with the exception of frequently used ones.

ANOVA	ANalysis Of VAriance
CIE	Commission Internationale de l'Éclairage
CRT	Cathode Ray Tube
CSF	Contrast Sensitivity Function
FFT	Fast Fourier Transform
FOV	Field Of View
FPS	Frames Per Second
GPU	Graphics Processing Unit
HDR	High Dynamic Range
HDRI	High Dynamic Range Imaging
HLS	Hue, Lightness, Saturation
HSV	Hue, Saturation, Value
HVS	Human Visual System
IQA	Image Quality Assessment
IQM	Image Quality Metrics
JND	Just Noticeable Difference
JPEG	Joint Picture Experts Group
LCD	Liquid Crystal Display
LDR	Low Dynamic Range
LGN	Lateral Geniculate Nucleus
LMS	Least Mean Squares
LUT	Look Up Table
MANOVA	Multivariate ANalysis Of VAriance
MOS	Mean Opinion Score
MSE	Mean Squared Error
MSSIM	Mean Structural SIMilarity index
PSF	Point Spread Function
PSNR	Peak Signal-to-Noise Ratio
RGB	Red, Green, Blue
RMSE	Root Mean Squared Error
ROI	Region Of Interest
SVD	Singular Value Decomposition
SNR	Signal-to-Noise Ratio
SSIM	Structural SIMilarity index
TM	Tone Mapping
TMO	Tone Mapping Operator
TRC	Tone Reproduction Curve
UQI	Universal Quality Index
VDM	Visual Discrimination Model
VDP	Visible Differences Predictor

1 Introduction

"All our knowledge has its origins in our perceptions."

(Leonardo da Vinci)

Visual perceptions are remarkably significant for human beings. We resolve various everyday issues, we gain a lot of knowledge, and we amuse ourselves thanks to our vision. Human visual system (HVS) is an extremely evolved part of the nervous system that interprets the information from visible light to enable us to see. Although the complex HVS that constitutes considerable piece of the brain excellently carries out visual tasks, it has its specific features, sometimes referred to as HVS limitations. Since outputs of computer graphics and digital image processing methods are observed by human subjects, such methods can and have to reflect the features of the HVS to generate perceptually correct and plausible images and to improve the performance. The knowledge of the HVS usually takes the form of *computational models* and these models can be incorporated at various areas of computer graphics and digital image processing. The fields where the utilization of HVS models becomes particularly beneficial and which we concern ourselves with in this thesis are image quality assessment, high dynamic range tone mapping, and color image to grayscale conversions.

1.1 Motivation

Even though the knowledge about the HVS is continually developing, many unanswered questions and hypotheses still remain to be solved. On that account, we are quite far from having an accurate computational model of the HVS. It is therefore of particular importance and necessity to conduct *experimental subjective evaluations* of the computational approaches that incorporate HVS models to judge and validate them properly. Moreover, the evaluations give us a deeper insight into the inquired area and besides the evaluation of existing methods, they can lead to *proposals of new ones*.

1.1.1 Image Quality Assessment

The goal of *Image Quality Assessment* (IQA) is to computationally predict human perceptions of image quality. It is known [Teo94] that classical metrics like Root Mean Squared (RMS) error are not adequate to the comparison of images, because they poorly predict the differences between the images as perceived by the human observer. To solve the problem properly, various perceptual *Image Quality Metrics* (IQM) have been proposed. Image quality metrics traditionally comprise a computational HVS model to correctly predict image difference as a human would perceive it. In this thesis, we concern ourselves with *full-reference image quality metrics* that assume that a complete reference image is known. The input to a full-

reference IQM includes two images (e.g. a reference image and a distorted image), whereas the output is a map describing the visible differences between them, see Figure 1.1.

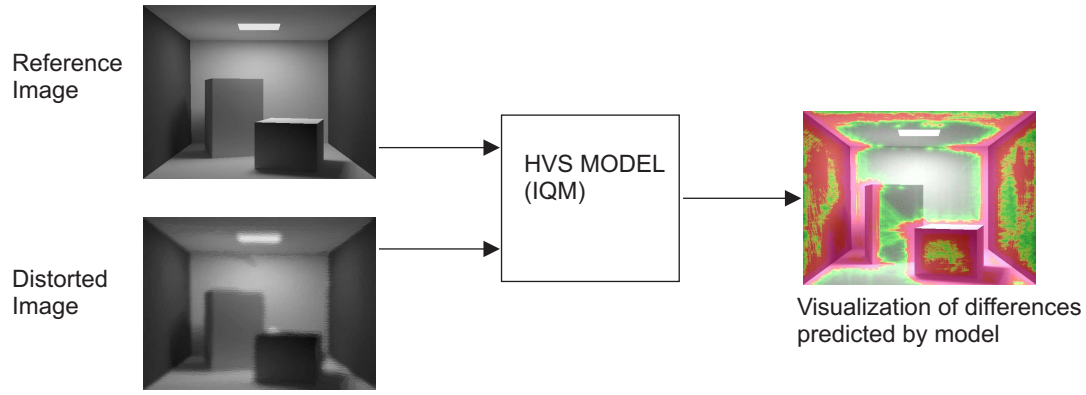


Figure 1.1: Full-reference image quality metrics.

Image quality assessment is very practical in various imaging applications. The main domains of IQA lie in the areas of monitoring of image quality (e.g. in lossy image compression), benchmarking of imaging applications, and optimizing algorithms and their parameter settings. However, image quality metrics have successfully been applied also to image database retrievals, to the evaluation of the perceptual impact of different rendering algorithms, to the perception-guided rendering of animations, etc.

1.1.2 High Dynamic Range Tone Mapping

High Dynamic Range (HDR) images are becoming widespread in many imaging areas thanks to the uniquely high range of values (luminances) and the high numerical precision. One may acquire an HDR image synthetically (e.g. utilizing a physically-based renderer), by means of a specialized HDR camera, or using series of ordinary images captured with varying exposure by a consumer-level digital camera. Since an HDR image can comprise vast range of luminances (typically the whole range of a real-world scene), the challenge is how to display it on devices with limited (Low) Dynamic Range (LDR). This problem is called an *HDR tone mapping* and various methods have been proposed to solve it, an overview can be found in [Devl02, Rein05b].

Tone mapping methods (sometimes called Tone Mapping Operators – TMOs) take an HDR image on input and reduce (transform) its range to an ordinary (LDR) image, see Figure 1.2. The goal of many tone mapping methods is to perform the transformation in a way that corresponds to human perception of a scene captured in the input HDR image. Accordingly, many tone mapping methods mimic the behavior of the human visual system. As we currently register several dozens of tone mapping approaches whose merits and shortcomings are not immanently clear, an experimental validation and evaluation of them is needed.

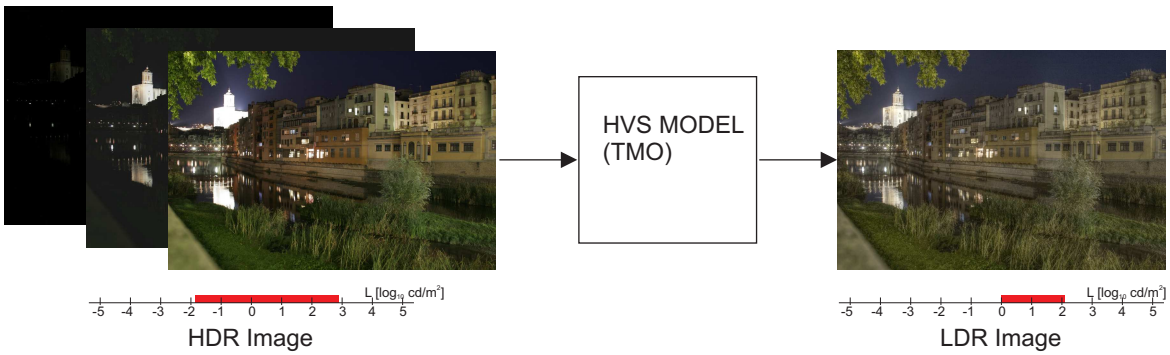


Figure 1.2: High dynamic range tone mapping.

1.1.3 Color to Grayscale Conversions

Color images often have to be converted to grayscale for reproduction or for subsequent processing. Methods performing *color to grayscale conversion* do basically a reduction of three dimensional color data into one dimension, see Figure 1.3. The aim of color to grayscale conversions is usually to produce perceptually plausible grayscale results. Unfortunately, an analogous conversion is not naturally present in the HVS. However, one may measure perceptual differences between colors in subjective psychophysical experiments. Moreover, it is evident that some loss of information due to the conversion is inevitable, so that the other goal is to save as much information contained in the original color image as possible. The complexity of the color to grayscale conversion issue is known, but the existing approaches [Gooc05, Rasc05] are eminently computationally intensive, which does not make them suitable for interactive conversions or conversions of high resolution images.

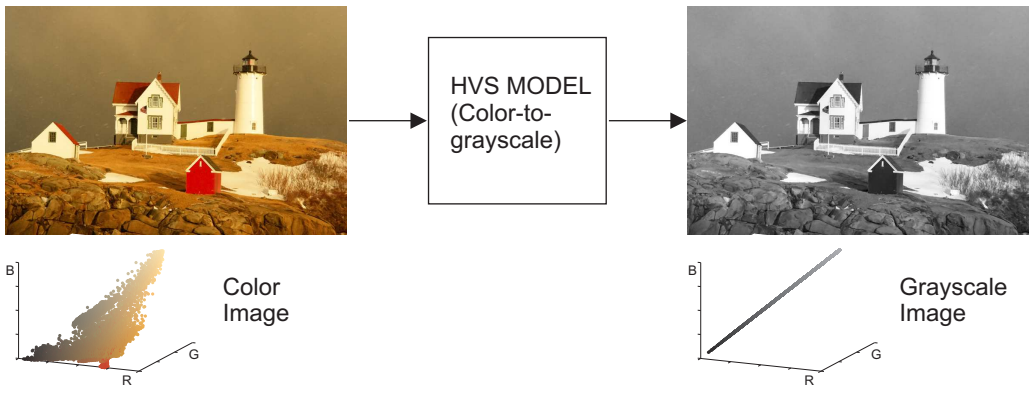


Figure 1.3: Color image to grayscale conversion.

1.2 Contributions

The thesis aims to provide the following contributions:

- Presentation and discussion of the results of subjective experiment that we have performed to *compare and evaluate traditional error sensitivity based and structural similarity based approaches* to the image quality assessment. A survey and the description of existing methods of the two groups (Chapter 3).

- A study about the effects of basic *image attributes for HDR tone mapping* and a survey how different methods reproduce these attributes. Proposal of a scheme of *relationships between essential image attributes*, leading to the definition of an overall image quality measure (Chapter 4).

- Presentation and discussion of the results of two different subjective psychophysical experiments that we have performed to prove the proposed relationship scheme of image attributes. Moreover, the presentation of an *evaluation of fourteen existing tone mapping methods* with regard to these image attributes (Chapter 4).

- A new perceptually plausible, fast and simple yet powerful general *hybrid approach to HDR tone mapping*. The new approach applies a global tone mapping method first to reproduce overall image attributes correctly and constructs an enhancement map to guide a local operator to the critical areas that deserve local enhancement. The new approach is general and can be easily tailored to miscellaneous goals of tone mapping (Chapter 5).

- Two new perceptually convincing local *color to grayscale conversion techniques*. One is based on the CIELab color difference formula, and the other on the Coloroid system and its experimental background (Chapter 6).

- A new and efficient way of converting an inconsistent gradient field into a consistent one, which can be directly converted into an image by a simple 2D integration. The complexity of this method which we call *gradient inconsistency correction* is linear in the number of pixels, making it suitable for high-resolution images (Chapter 6).

1.3 Structure of the Thesis

The thesis is organized as follows[1]:

Chapter 2 covers the anatomical structure and the perceptual behavior of the human visual system. This knowledge is often used in computer graphics and image processing algorithms.

Chapter 3 at first surveys perceptual image quality metrics and then describes our contribution to the field – an experimental evaluation of two principal approaches to the image quality assessment.

[1] Additional materials (full color versions of the images, etc.) are available at the web pages of the author: `http://www.cgg.cvut.cz/~cadikm/diss`

Chapter 4 presents a comprehensive overview and analysis of image attributes for HDR tone mapping and describes our efforts on evaluation of tone mapping methods using the essential image attributes.

Chapter 5 introduces a new fast, easy to implement, and perceptually plausible hybrid approach to HDR tone mapping.

Chapter 6 presents two new local grayscale conversion techniques for perceptually justifiable conversion of color images to grayscale.

Chapter 7 concludes the thesis and gives suggestions for further research.

Finally, the Appendix describes several methods of statistical data analysis and image processing algorithm designs that are used in the thesis.

2 Human Visual System

"One touch of nature makes the whole world kin."

(William Shakespeare)

Since the majority of perceptually based approaches to computer graphics is inspired by properties of the *human visual system* (HVS), we will describe the HVS first. We will begin with the physical structure (Section 2.1) that is quite well established and that can help us to understand rather complex characteristics of the perceptual behavior (Section 2.2). Finally, we will outline several theories of the nature of human image processing (Section 2.3). This part of the thesis was largely acquired from an excellent book on vision science by Stephen E. Palmer [Palm02].

2.1 Physical Structure

In this section, we will describe basic visual anatomy and physiology. This can give us insights into the kinds of information that can be coded by visual mechanisms.

2.1.1 The Human Eye

Human eyes [Palm02] are approximately spherical in shape except for a bulge at the front. Located at about the horizontal midline of the head, they sit in nearly hemispherical holes in the skull, called the eye sockets. Each eye is moved by the coordinated use of six small, strong muscles, called the *extraocular muscles*, which are controlled by specific areas in the brain. Eye movements are necessary for scanning different regions of the visual field without having to turn the entire head and for focusing on objects at different distances.

The eyes gather light reflected from surfaces in the world and focus it in a clear image on the back of the eye. There are many parts of the eye that accomplish different optical functions [Palm02], see Figure 2.1. First, light enters the *cornea*, a transparent bulge on the front of the eye behind which is a cavity filled with a clear liquid, called the *aqueous humor*. Next, light passes through the *pupil*, a variably sized opening in the opaque *iris*, which gives the eye its external color. Just behind the iris, light passes through the *lens*, whose shape is controlled by *ciliary muscles*. The optical properties of the lens can be altered by changing its shape, a process called *accommodation*. The photon then travels through the clear *vitreous humor* that fills the central chamber of the eye. Finally, it reaches the *retina*, the curved surface at the back of the eye. The retina is densely covered with over 100 million light-sensitive *photoreceptors*, which convert light into neural activity.

The information about the light striking the retina is transmitted to the primary visual cortex in the occipital lobe at the back of the head, see Figure 2.2. Some estimates [Palm02] put the

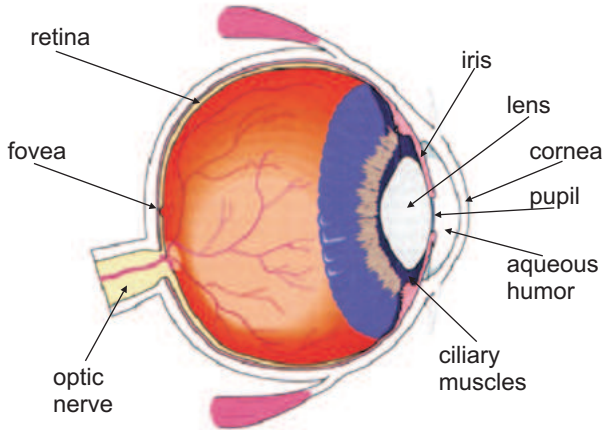


Figure 2.1: A cross section of the human eye. (After Fairchild [Fair98].)

percentage of cortex involved with visual function at more than 50% for the macaque monkey, although it is probably slightly lower for humans. The complete visual system includes much of the brain as well as the eyes, and the whole eye-brain system must function properly for the organism to extract reliable information about the environment.

2.1.2 The Retina

After the optical processing, the next job of the eye is to convert light into neural activity so that the brain can process it. In the visual system, this function is carried out by *photoreceptors* in the retina: photoreceptors are specialized retinal cells that are stimulated by light energy. There are two distinct classes of photoreceptor cells [Palm02]: *rods* and *cones*. Rods are more numerous (about 120 million), extremely sensitive to light, and located everywhere in the retina except at its very center. They are used exclusively for vision at very low light levels (called *scotopic conditions*). Cones are less abundant (about 8 million), much less sensitive to light, and heavily concentrated in the center of the retina, although some are found scattered throughout the periphery. They are responsible for our visual experiences under most normal lighting conditions (called *photopic conditions*) and for all our experiences of color. There is a small region, called the *fovea*, right at the center of the retina that contains nothing but densely packed cones. The visual angle covered by the fovea is only about 2 degrees. Another region exists where the axons of the ganglion cells leave the eye at the optic nerve. This region is called the *optic disk* (also known as the *blind spot*) and it contains no receptor cells at all. However, we do not experience blindness there, except under very special circumstances.

When the optical information is coded into neural responses, some initial processing is accomplished within the retina itself by several other types of neurons, including the *horizontal*, *bipolar*, *amacrine*, and *ganglion cells* [Palm02], all of which integrate responses from many nearby cells, see Figure 2.3. The inputs of the retinal ganglion cells are arranged in an antagonistic, concentric pattern composed of a center and a surround region (the area of the retina which the ganglion cell receives input from is called the *receptive field*). The ganglion cell is

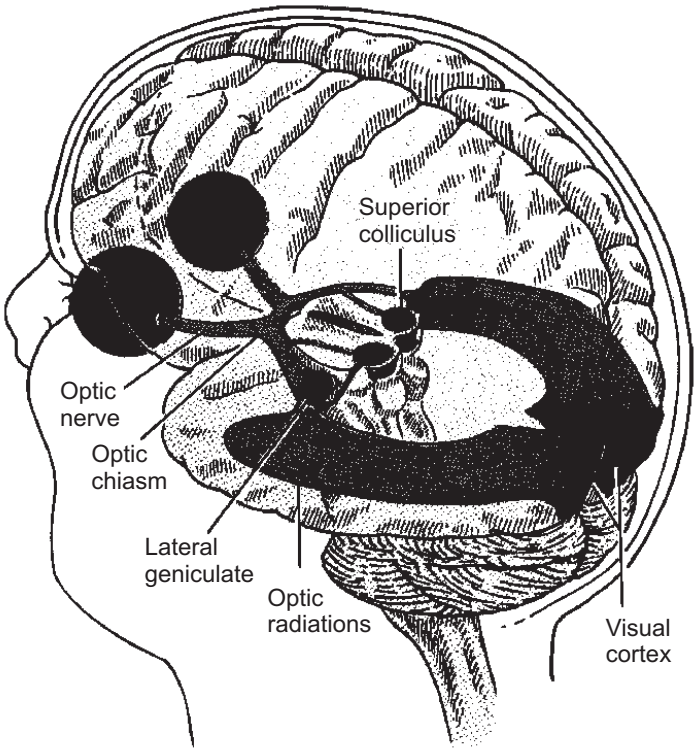


Figure 2.2: The human visual system. (After Palmer [Palm02].)

continually emitting a background signal; however when light strikes the photoreceptors in one region, this stimulates an increased response from the retinal ganglion cell (*on-response*), whereas light falling on the other region will generate a reduced response (*off-response*). There are two distinct types of ganglion cells, the *on-center cells*, where the center region is stimulated by an on-response, and the *off-center cells*, where the center region is stimulated by an off-response, see Figure 2.4.

The axons of the ganglion cells carry information out of the eye through the *optic nerve* to the *optic chiasm* [Palm02]. Here, the fibres from the nasal side of the fovea in each eye cross over to the opposite side of the brain while the others remain on the same side. The result is that the mapping from external visual fields to the cortex is completely crossed – all of the information from the left half of the visual field goes to the right half of the brain, while all the information from the right visual field goes to the left half of the brain. From the optic chiasm, there are two separate pathways into the brain on each side. The smaller one (only a few percent) goes to the *superior colliculus*, a nucleus in the brain stem. This visual center seems to process primarily information about where things are in the world and to be involved in the control of eye movements. The larger pathway goes first to the *lateral geniculate nucleus* (or LGN) of the thalamus and then to the *occipital cortex* (or *primary visual cortex*).

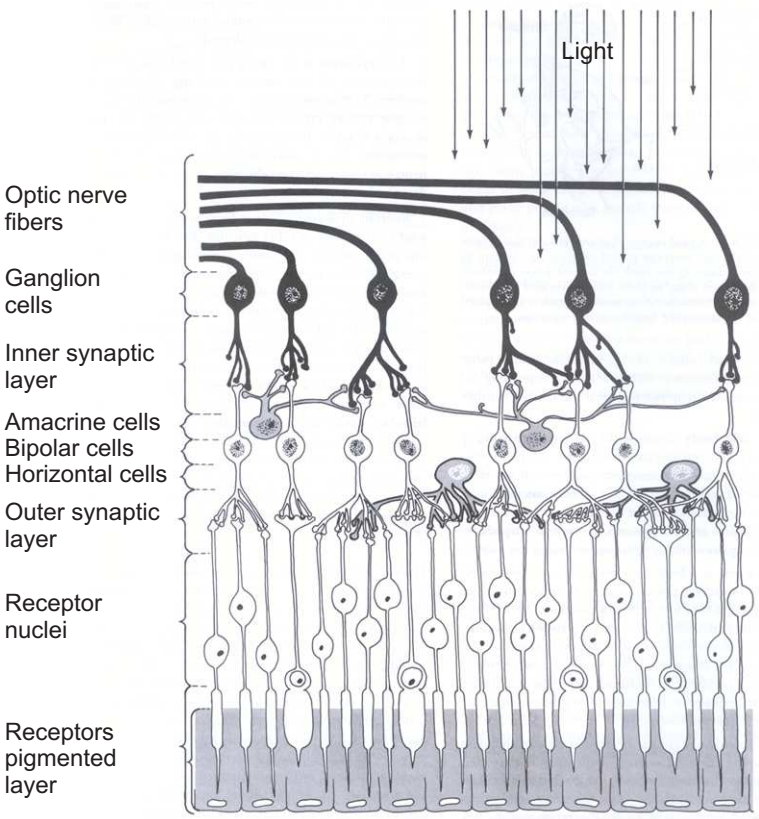


Figure 2.3: The human retina. (After Palmer [Palm02].)

2.1.3 Visual Cortex

The human cortex is divided into two halves, *cerebral hemispheres*, that are approximately symmetrical. As a result of many neuropsychological studies [Palm02], it is now well established that the occipital lobe is the primary cortical receiving area for visual information. Although it would be a gross overstatement to say that vision scientists understand how visual cortex works, they are at least beginning to get some glimmerings of what the assorted pieces might be and how they might fit together.

The first steps in cortical processing of visual information take place in the *striate cortex*, sometimes called *primary visual cortex* or area *V1* [Palm02]. This is the largest part of the occipital lobe and it seems likely that the most complex visual processing occurs there. Striate cortex receives its input from the LGN on the same side of the brain, so the visual input of striate cortex, like that of LGN, is completely crossed. Both sides are activated by the thin central vertical strip, measuring about 1 degree of visual field. The cells that are sensitive to this strip in one side of the brain are connected to the corresponding cells on the other side of the brain through the *corpus callosum*, the large fiber tract that allows communication between the two cerebral hemispheres. The mapping from retina to striate cortex is topological in that nearby regions on the retina project to nearby regions in striate cortex. The central area of the visual field, which falls on or near the fovea, receives proportionally much greater

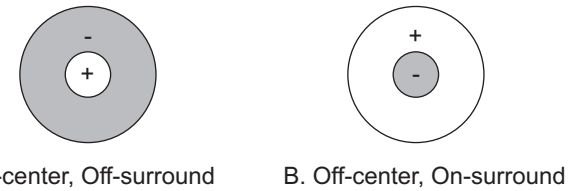


Figure 2.4: Receptive field structure of ganglion cells. On-center, off-surround cells (A) fire to light onset and stop at offset in their excitatory center, but they stop firing to light onset and begin firing at offset in their inhibitory surround. Off-center, on-surround cells (B) exhibit the opposite characteristics. (After Palmer [Palm02].)

representation in the cortex than the periphery does. This is called the *cortical magnification factor*.

The inferior temporal centers in the lower (ventral) system seem to be involved in *identifying objects*, whereas the parietal centers in the upper (dorsal) system seem to be involved in *locating objects* [Palm02]. These two pathways are often called the *"what" system* and the *"where" system*. It seems almost inevitable that these two different kinds of information must get together somewhere in the brain so that the "what-where" connection can be made, but it is not yet known where this happens.

It is abundantly clear that a great deal of visual processing takes place in parallel across different areas, each region projecting fibers to several other areas but by no means to all of them. The connections are generally bidirectional; that is, if the area X projects to the area Y, then Y projects back to X as well.

2.1.4 The Physiological Pathways Hypothesis

One possible relation between anatomical structure and physiological function has begun to emerge during the last decade. The hypothesis is that there are separate neural *pathways* for processing information about different visual properties such as color, shape, depth, and motion. Livingstone and Hubel [Livi88] proposed that these four types of information are processed in different neural pathways from the retina onward. They report evidence that color, form, motion, and stereoscopic depth information are processed in distinct subregions of visual cortical areas V1 and V2, as indicated schematically in Figure 2.5.

These areas are then projected to distinct higher-level areas of cortex: movement and stereoscopic depth information to area *V5* (also called *MT*, Medial Temporal cortex), color to area *V4*, and form through several intermediate centers (including V4) to area *IT* (InferoTemporal cortex) [Palm02], where cells have been found that respond selectively to faces, hands, and other highly complex stimuli. From these areas, the form and color pathways may project to the ventral "what" system for object identification and the depth and motion pathways to the dorsal "where" system for object localization.

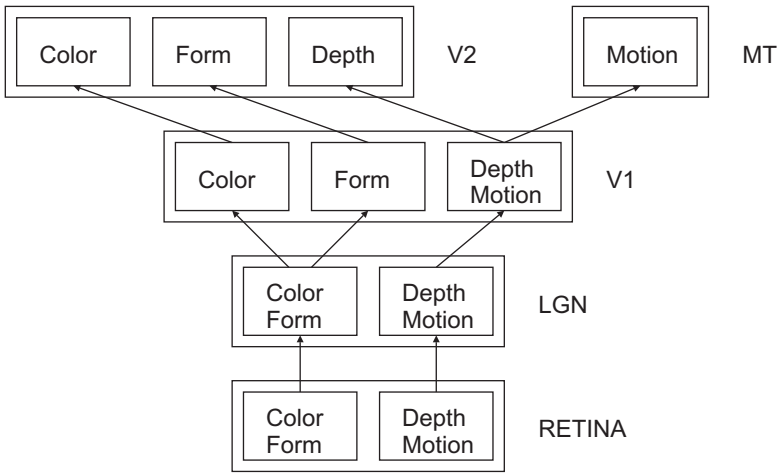


Figure 2.5: Schematic diagram of the visual pathways hypothesis. (After Palmer [Palm02].)

2.2 Visual Perception

Visual perception is the process of acquiring knowledge about environmental objects and events by extracting information from the light they emit or reflect. Visual perception concerns the *acquisition of knowledge* – this means that vision is fundamentally a cognitive activity [Palm02], distinct from purely optical processes such as photographic ones. There are indeed important similarities between eyes and cameras in terms of optical phenomena, but there are no similarities whatever in terms of perceptual phenomena – cameras have no perceptual capabilities at all. The knowledge achieved by visual perception concerns *objects and events in the environment*, perception is not merely about an observer's subjective visual experiences. Visual knowledge about the environment is obtained by *extracting information*, that implies information processing approach to the vision. Finally the information that is processed in visual perception comes from the light that is *emitted or reflected by objects*, optical information is the foundation of all vision.

2.2.1 Adaptation

Visual perception changes over time as it adapts to particular conditions. When we enter a darkened room on a bright afternoon, for instance, we cannot see much. After 20 minutes, however, we can see everything surprisingly well. This increase in sensitivity to light is called *dark adaptation* [Palm02]. Adaptation is a very general phenomenon in visual perception – visual experience may become less intense as a result of prolonged exposure to a wide variety of different kinds of stimulation: color, orientation, size, motion, etc. These changes in visual experience show that visual perception is not always a clear window onto reality because we have different visual experiences of the same physical environment at different stages of adaptation.

2.2.2 Ambiguous Figures

To provide us with information, vision is an *interpretive process* [Palm02] that somehow transforms complex, moving, two-dimensional patterns of light at the back of the eyes into stable perceptions of three-dimensional space. The objects we perceive are actually interpretations based on the structure of images rather than direct registrations of physical reality. Potent demonstrations of the interpretive nature of vision come from *ambiguous figures*, single images that can give rise to two or more distinct perceptions, see for an example Figure 2.6. The interpretations of these ambiguous figures are *mutually exclusive*. We perceive just one of them at a time: a duck or a rabbit, not both. This is consistent with the idea that perception involves the construction of an interpretive model because only one such a model can be fit to the sensory data at one time. If perception was completely determined by the light stimulating the eye, there would be no ambiguous figures because each pattern of stimulation would map onto a unique percept.



Figure 2.6: Ambiguous figures. Figure on the left can be seen as a duck (facing left) or a rabbit (facing right). Figure on the right can be seen as a saxophonist (facing right) or a face of a woman.

2.2.3 Visual Completion

People's perceptions actually correspond to the *models* that their visual systems have constructed [Palm02] rather than to the sensory stimulation on which the models are based. That is why perceptions can be illusory and ambiguous despite the nonillusory and unambiguous status of the raw optical images in which they are based. Perceptual models must be closely coupled to the information in the projected image of the world and must provide reasonably accurate interpretations of this information.

Perhaps the most convincing evidence that visual perception involves the construction of environmental models comes from the fact that our perceptions include portions of surfaces that we cannot actually see. This perceptual filling in of parts of objects that are hidden from view is called *visual completion*. It happens automatically and effortlessly whenever we perceive the environment. Visual perception also includes information about *self-occluded surfaces*, those surfaces of an object that are entirely hidden from view by its own visible surfaces.

2.2.4 Impossible Objects

Impossible objects are two-dimensional line drawings that initially give the clear perception of coherent three-dimensional objects but are physically impossible, see Figure 2.7. Such demonstrations support the idea that vision actively constructs environmental models rather than simply registering what is present. If visual perceptions were merely a reliable reflection of the world, a physically impossible object simply could not be perceived. These kinds of errors that are evident in perceiving impossible objects seem to indicate that at least some visual processes work initially at a local level and only later fit the results into a global framework.

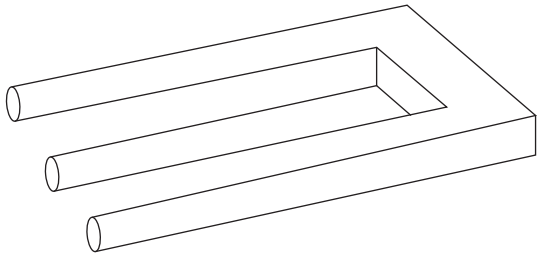


Figure 2.7: An impossible object. The drawing in this figure produces perception of coherent three-dimensional object, but it is physically impossible.

2.2.5 Classification

Our perceptual constructions go even further than completing unseen surfaces [Palm02]. They include information about the meaning or functional significance of objects and situations. Being able to *classify* objects as members of known categories allows us to respond to them in appropriate ways because it gives us access to vast amounts of information that we have stored from previous experiences with a similar object. Previous experience with members of a given category allows us to predict with reasonable certainty what new members of the same class will do. As a consequence we can deal with most new objects at the more abstract level of their category, even though we have never seen that particular object before.

2.2.6 Attention and Consciousness

The visible environment contains much more information than anyone can fully perceive [Palm02]. We must therefore be *selective* in what we attend to, and what we select will depend a great deal on our needs, goals, plans, and desires. Perception is not therefore an entirely stimulus-driven process – perceptions are not determined solely by the nature of the optical information present in sensory stimulation. Our perceptions are also influenced by *cognitive constraints* – higher-level goals, plans, and expectations. We look at different things in our surroundings depending on what we are trying to accomplish, and we may perceive them differently as a result.

One of the functions of *attention* is to bring visual information to consciousness. Certain properties of objects do not seem to be experienced consciously unless they are attended, yet unattended objects are often processed fully enough outside of consciousness to attract our attention. Once the object is attended, we become conscious of its detailed properties and are able to identify it and discern its meaning in the present situation. In general, lower levels of perception do not seem to be accessible to, or modifiable by, conscious knowledge and expectations, whereas higher levels do. However not much is yet known about the role of consciousness in perception.

2.3 Image Processing Theories

Several theories for description of the nature of human image processing have evolved over the years. These theories compete because their functional implications are quite different and no one is universally valid. We will describe the two most common of them: line-edge detection theory, and spatial frequency theory, although yet another approaches exist (connectionistic theory, neural networks, scale-space, etc.). Finally, we will give an overview of contemporary theoretical hypothesis about the architecture of human image processing. This hypothesis integrates in some respect these two fundamentally different theories.

2.3.1 Line and Edge Detection Theory

Hubel and Wiesel [Hube59] were the first to successfully apply the receptive field mapping techniques (described in Section 2.1.2) to striate cortex. Their investigations revealed that there were several different kinds of cortical cells that had different receptive field characteristics. They classified them into three types: *simple cells, complex cells,* and *hypercomplex cells*. For simple cells the responses to complex stimuli can be predicted from their responses to individual spots of light. A simple cell's response to a larger and more complex pattern of stimulation can therefore be roughly predicted by summing its responses to the set of small spots of light that compose it. There were identified several different subtypes of simple cells. The vast majority have an elongated structure, firing most vigorously to a line or an edge at a specific retinal position and orientation. Simple cells that have an area of excitation on one side and an area of inhibition on the other side, respond to a luminance edge in the proper orientation and are called *edge detectors*. Simple cells that have receptive fields with a central elongated region that is either excitatory or inhibitory, with an antagonistic field on both sides of it, respond maximally to bright or dark lines and are called *line detectors* or *bar detectors*.

The view of image processing that has emerged from these findings is that an early step in spatial image processing is to find the lines and edges in the image. Higher-level properties, such as shapes and orientations of objects, might then be constructed by putting together many local edges and lines that have been identified by their detector cells in V1. Whether or not this is the correct view is still an open question, but it has dominated thinking about the initial stages of visual processing for several decades.

About 75% of the cells in striate cortex are *complex cells* [Palm02]. Similar to simple cells, complex cells have elongated receptive fields but they differ from simple cells in several important respects: complex cells are highly *nonlinear*, they tend to be highly responsive to

moving lines or edges anywhere within their receptive field. Complex cells are not very sensitive to the *position* of certain stimuli and they tend to have somewhat *larger receptive fields* than simple cells. Complex cells are thought to receive input from several simple cells whose receptive fields have the same orientation but different positions.

The third type of striate cell is the *hypercomplex cell*. The most striking characteristic of hypercomplex cells is that extending a line or an edge beyond a certain length causes them to fire less vigorously than they do to a shorter line or edge. For this reason, they are often called *end-stopped cells*. Recent quantitative studies suggest that the degree of "end-stopping" is a continuum rather than an all-or-none phenomenon.

2.3.2 Spatial Frequency Theory

The spatial frequency theory dominates psychophysical theories of early spatial vision because it is able to explain a large number of important and surprising results from psychophysical experiments, not only in adult vision, but in infant vision as well [Palm02]. This theory is based on an atomistic assumption: the representation of any image, no matter how complex, is an assemblage of many primitive spatial "atoms". The primitives of spatial frequency theory are spatially extended patterns called *sinusoidal gratings*: two-dimensional patterns whose luminance varies according to a sine wave over one spatial dimension and is constant over the perpendicular dimension. Each primitive sinusoidal grating can be characterized completely by four parameters: its *spatial frequency, orientation, amplitude,* and *phase*. Spatial frequency is usually specified in terms of the number of light/dark cycles per degree of visual angle. The orientation of the grating is specified in degrees counterclockwise from vertical. Phase is specified in degrees, such that the grating whose positive-going inflection point is at the reference point is said to have a phase 0° (called sine phase).

There is a good formal mathematical reason for choosing sinusoidal gratings as primitives: *Fourier analysis* [Smit99]. Fourier analysis is a method, by which any two-dimensional luminance image can be analyzed into the sum of a set of sinusoidal gratings that differ in spatial frequency, orientation, amplitude, and phase. The Fourier analysis of an image consists of two parts: the power spectrum and the phase spectrum. The *power spectrum* specifies the amplitude of each grating at a particular spatial frequency and orientation, whereas the *phase spectrum* specifies the phase of each grating at a particular spatial frequency and orientation. If all of these gratings at the proper phases and amplitudes were added up, they would exactly recreate the original image. Thus, Fourier analysis provides a very general method of decomposing complex images into primitive components, since it has been proven to work for any image. Fourier analysis is also capable of being "inverted" through a process called *Fourier synthesis* so that the original image can be reconstructed from its power and phase spectra.

Spatial Frequency Channels The spatial frequency theory proposes that early visual processing can be understood in terms of a large number of overlapping *psychophysical channels* that are selectively tuned to different ranges of spatial frequencies and orientations. Thanks to many psychophysical studies of people's detection and discrimination of grating stimuli [Blak69], there is now a great deal of evidence to support this view.

The standard measurement of human sensitivity to gratings at different frequencies is called the *contrast sensitivity function (CSF)*. It is determined by finding the lowest contrast at which the observer can just barely detect the difference between a sinusoidal grating and a uniform gray field, that is, the threshold at which a very low-contrast grating stops looking like a uniform gray field and starts to look striped. This threshold is measured for gratings at many different spatial frequencies from low to high. The results can be summarized in a graph in which the contrast sensitivity at threshold is plotted as a function of spatial frequency, as shown in Figure 2.8.

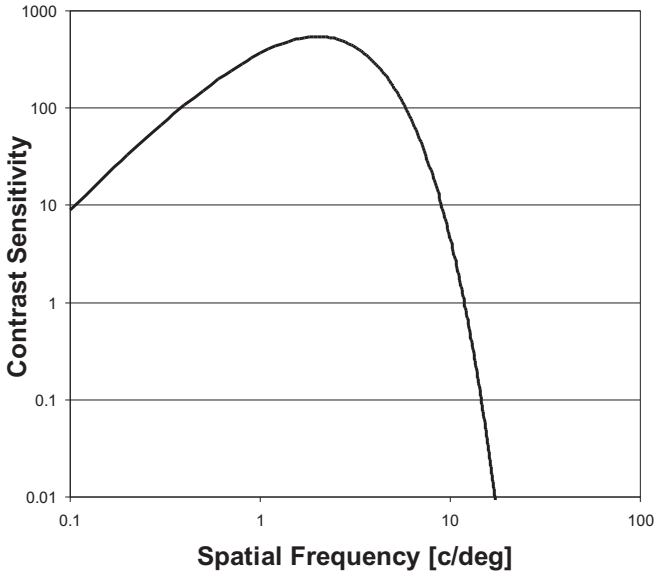


Figure 2.8: Contrast sensitivity function.

The CSF shows [Palm02] that people are most sensitive to intermediate spatial frequencies at about 4–5 cycles per degree of visual angle. If the CSF is measured under low-light (scotopic) conditions for humans, sensitivity to all frequencies drops dramatically, especially at the highest frequencies. This means that at night, when just the rods are operating, human vision lacks the high acuity that it has in daylight.

Several measurements on human subjects have shown the selective adaptation of channels. The extended exposure to the grating caused the subject's visual system to *adapt*, that is, to become less sensitive after the prolonged viewing experience (see Section 2.2.1), but only near the particular spatial frequency and orientation of the adapting grating. Just as gratings of a particular spatial frequency and orientation produce specific adaptation effects, they also produce specific *aftereffects*.

Local Spatial Frequency Theory Psychophysical channels are hypothetical mechanisms inferred from behavioral measures rather than directly observed biological mechanisms. If these channels are real, however, they must be implemented somewhere in the visual system. To face this problem there arises second theory about the function of the cells in striate cortex. There is now substantial evidence [Palm02] that these cells may be performing a *local*

spatial frequency analysis of incoming images. A local, piecewise, spatial frequency analysis can be accomplished through many small patches of sinusoidal gratings that "fade out" with distance from the center of the receptive field. This sort of receptive field structure, called a *Gabor function*, is constructed by multiplying a global sinusoidal grating by a bell-shaped Gaussian envelope, see the multiscale transforms introduction in the Appendix B.

The degree of frequency tuning in cortical cells seems to fall along a continuum; some of them are very sharply tuned and others quite broadly tuned. Cells that are tuned to high spatial frequencies have narrower tuning than do cells that are tuned to low spatial frequencies. There is a similar continuum in the degree of orientation tuning; some cells respond only to gratings that are very close to their "favorite" orientation, whereas others respond almost equally to gratings in any orientation. Moreover, cells that are broadly tuned for spatial frequency are also broadly tuned for orientation, and cells that are narrowly tuned for spatial frequency are also narrowly tuned for orientation.

Although the evidence that simple and complex cells in area V1 may be doing a local spatial frequency analysis of input images is impressive, this conclusion is not universally held. Nevertheless, local spatial frequency theory must be counted a very serious alternative to the line and edge detection theory.

Architecture of Image Processing Hypothesis We have reviewed two theories about the function of the cortical cells. The psychophysical view is that these cells describe images in a piecewise, local Fourier analysis. However, the edge detection theory claims that these same cells are actually the physiological implementation of edge detection mechanisms at different spatial scales. These different views are perhaps not as incompatible as they appear.

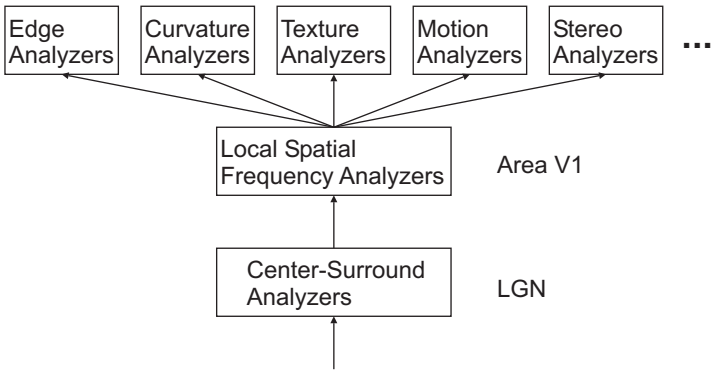


Figure 2.9: A theoretical hypothesis about the architecture of image processing. (After Palmer [Palm02].)

The hypothesis is that local spatial frequency theory and edge detection theory may be appropriate for different levels of the visual system [Palm02], as diagrammed in Figure 2.9: center/surround cells in retina and LGN provide input to local spatial frequency analyzers in area V1 of visual cortex, which then project their output to a variety of different modules that compute edges, surface curvatures, textures, stereopsis, and so on, at later stages. According to this view, edge detectors would then be constructed from the output of local spatial frequency analyzers by coding for the output pattern that is characteristic of luminance edges.

3 Perceptual Evaluation of Image Quality Metrics

"If there were no best among all possible worlds, God would not have created one."

(Gottfried Wilhelm von Leibniz)

Image quality assessment is a challenging task that is traditionally approached by computational models of human visual system. Recently, a new assessment methodology based on structural similarity has been proposed. In this chapter we select two representative models of each group, the Visible Differences Predictor (VDP) and the Structural SIMilarity index (SSIM), for experimental evaluation. We begin with the survey and with the description of models of the two groups (Sections 3.2, 3.3). We then depict the subjective experiments that we have conducted to obtain subjective mean opinion scores (Sections 3.4, 3.5). Inputs to these tests included uniformly compressed images and images compressed non-uniformly with Regions Of Interest (ROI). Then, we assess and discuss the performances of the selected models, and the similarities and differences between them (Section 3.6). We end with a summary of the important advantages of each approach (Section 3.7).

3.1 Introduction

The goal of an objective image quality assessment is to develop quantitative metrics that can automatically predict perceived image quality [Wang02b]. Image quality assessment and comparison metrics play an important role in a broad range of applications. They can be utilized to monitor image quality, they can be employed to benchmark image processing algorithms, and they can be embedded into the rendering algorithms to optimize their performances and parameter settings. It is well known [Teo94], that classical comparison metrics like Root Mean Square Error (RMSE) are not sufficient when applied to the comparison of images, because they poorly predict the differences between the images as perceived by the human observer. This fact has led to the development of more advanced perceptual quality assessment techniques.

Objective image quality metrics serve primarily to assessment of the difference between two images, an original image and a distorted image, see Figure 1.1 on page 2 for illustration. They can be classified according to the availability of an original image, with which the distorted image should be compared. Most existing approaches are known as *full-reference*, meaning that a complete reference image is assumed to be known. We deal with full-reference metrics in this chapter. In some applications, however, the reference image is not available, and a *no-reference* or "blind" quality assessment approach is desirable. In a third type of methods, the reference image is only partially available, in the form of a set of extracted features made available as side information to help evaluate the quality of the distorted image. This is referred to as *reduced-reference* quality assessment.

3.2 Traditional Perceptual Image Quality Assessment

Traditional perceptual image quality assessment approaches are based on measuring the errors (signal differences) between the distorted and the reference images, and attempt to quantify the errors in a way that simulates human visual error sensitivity features.

In this section, we will first outline the general framework of image quality metrics. Then we will summarize the main perceptually driven image quality metrics. Finally, we will outline merits and shortcomings of these techniques.

3.2.1 General Framework of Perceptual Quality Metrics

A great variety of models has been proposed in the literature. For many of these models, common computational parts can be identified [Wang03]. These parts are: preprocessing, CSF filtering, channel decomposition, error normalization and masking, and finally the error pooling, see Figure 3.1.

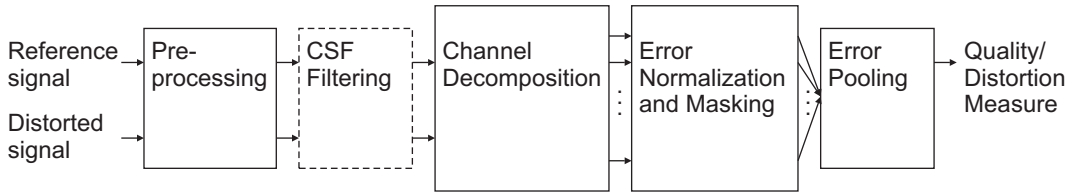


Figure 3.1: Block diagram of typical discrimination/quality metric. (After Wang et al. [Wang03].)

- The *pre-processing* stage may perform alignment, transformations of color spaces, calibration for display devices, point spread function filtering, and light adaptation.
- *CSF* may be implemented before the channel decomposition using linear filters that approximate the frequency responses of the CSF. On the other hand, some metrics implement CSF as weighting factors for channels after the channel decomposition.
- *Channel decomposition* is used to model the frequency selective channels in the HVS. The channels serve to separate the visual stimulus into different spatial and temporal *subbands*. During this phase, quality metrics differ mostly in the chosen filters.
- *Error normalization and masking* is typically implemented within each channel. Most models implement masking in the form of a gain-control mechanism that weights the error signal in a channel by a space-varying *visibility threshold* for that channel. The visibility threshold adjustment at a point is calculated based on the energy of the signal in the neighbourhood of that point, as well as the HVS sensitivity for that channel in the absence of masking effects.
- *Error pooling* is the process of combining the error signals in different channels into a single distortion/quality interpretation. The typical implementation uses *Minkowski*

summation (also called L_p norm) on the two sets of channels to compute the model response r:

$$r = \Big(\sum_l \sum_k |e_{l,k}|^\beta \Big)^{1/\beta},$$

where $e_{l,k}$ is the normalized and masked error of the k-th coefficient in the l-th channel, and β is a constant with a value between 1 and 4.

3.2.2 Traditional Perceptual Image Quality Metrics

3.2.2.1 Pixel-Based Mathematical Metrics

The *mean squared error* (MSE) and the *peak signal-to-noise ratio* (PSNR) are the most popular difference metrics in image and video processing. Neither MSE nor PSNR are perceptual measures, we mention them here for completeness. The MSE is the mean of the squared differences between the gray-level values of pixels in two pictures I and $\bar{I}$:

$$MSE = \frac{1}{XY} \sum_x \sum_y [I(x,y) - \bar{I}(x,y)]^2,$$

for pictures of size $X \times Y$. The average difference per pixel is thus given by the root mean squared error $RMSE = \sqrt{MSE}$.

The PSNR in decibels is defined as:

$$PSNR = 10 \log \frac{m^2}{MSE},$$

where m is the maximum value that a pixel can take (e.g. 255 for 8-bit images). *Color PSNR* is a version of the PSNR that accounts for colors, using perceptually uniform differences. Both MSE and PSNR are well-defined only for luminance information, there is no agreement on the computation of color values.

3.2.2.2 Model after Mannos and Sakrison

The first perception based image quality metric for luminance images was developed by Mannos and Sakrison [Mann74]. Computation of the proposed model begins by normalizing all the luminance values L_{ij} by the mean luminance L_m. The nonlinearity in perception is accounted for by taking the cubed root of each normalised luminance. A Fast Fourier Transform is computed of the resulting values, and the magnitude of the resulting transform at frequencies in the horizontal and vertical directions (u, v), (where u and v are expressed in terms of cycles per visual degree) is denoted $f_{uv}(|| \sqrt[3]{L/L_m} ||)$. The magnitudes f_{uv} are then filtered with the CSF $A_M(u,v) = A_M(r)$, where $r = u^2 + v^2$, to account for spatial frequency sensitivity to produce the array of values g_{uv} :

$$A_M(r(u,v)) = 2.6 \cdot (0.0192 + 0.144\sqrt{r}) \exp[-(0.144\sqrt{r})^{1.1}],$$

$$g_{uv} = f_{uv}\left(\left|\left|\sqrt[3]{\frac{L}{L_m}}\right|\right|\right) \cdot A_M(r(u,v)).$$

Finally, the distance between the two images is computed by finding the Mean Square Error of the values g_{uv} for each of the two images:

$$M(X;Y) = \frac{1}{N}\sum_{all\ u,v}(g_{X,uv} - g_{Y,uv})^2.$$

This technique therefore measures similarity in Fourier amplitude between images. It was shown to correlate quite well with subjective ranking data. Despite its simplicity, this metric was one of the first works in engineering to recognize the importance of applying vision science to image processing.

3.2.2.3 Model after Gervais

Another simple model was adapted from a study of confusion between letters of the alphabet [Gerv84] by Rushmeier et al. [Rush95]. The model includes the effect of phase as well as magnitude in the frequency space representation of the image. The luminances are normalised by dividing by the mean luminance. An FFT is computed producing an array of phases $p_{uv}(||L/L_m||)$ and magnitudes $f_{uv}(||L/L_m||)$. The magnitudes are then filtered with an anisotropic CSF filter function constructed by fitting splines to psychophysical data presented by Campbell et al. [Camp66], producing the filtered values $g_{uv}(||L/L_m||)$. The distance between images is then computed using:

$$M(X;Y) = \frac{1}{N}\sum_{all\ u,v}\left(\Big((\log g_{X,uv}+1) - (\log g_{Y,uv}+1)\Big)\Big(1 + p_{X,uv} - p_{Y,uv}\Big)\right)^2.$$

Since the Gervais model includes phase information (i.e. pixel position), its performance suffers due to subjectively minor registration problems between images. However, in situations where geometric alignment is not a problem, or is of critical importance for some other reason, this model may actually outperform the others.

3.2.2.4 Visible Differences Predictor

The Visible Differences Predictor [Daly93] (VDP) is one of the best-known image distortion metrics. The VDP model interprets early vision behavior, from retinal contrast sensitivity to spatial masking. Figure 3.2 shows the use of the VDP, which consists of three main stages: components for *calibration* of the input images, a *human visual system (HVS) model* and a method for *displaying* the HVS visible *differences*. The input to the algorithm includes two images and parameters for viewing conditions, whereas the output is a map describing the visible differences between them (see Figure 1.1 on the page 2). The output map defines the probability of detecting the differences between the two images as a function of their location in the images. This metric, *probability of detection*, provides a description of the threshold behavior of vision but does not discriminate among different suprathreshold visual errors.

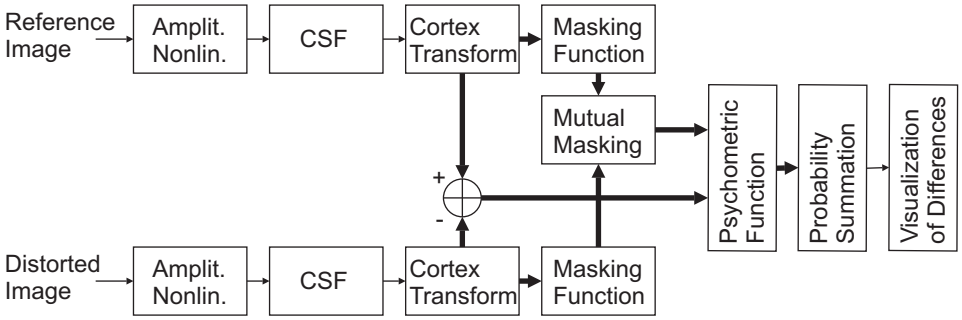


Figure 3.2: Block diagram of the Visible Differences Predictor [Daly93] (bold lines indicate parallel processing).

Calibration First, the input images represented by unitless digital numbers are calibrated. The calibrating input parameters are: the *viewing distance* for which the VDP will make its visual prediction, and the *physical pixel spacings*, which along the viewing distance map the visual frequencies expressed in cycles per degree (c/deg) to frequencies expressed digitally as a fraction of the Nyquist frequency.

Human Visual System Model The human visual system model is the key element of the VDP. It concatenates on the lower-order processing of the visual system, such as the optics, retina, lateral geniculate nucleus, and striate cortex. The HVS model consists of a number of processes that limit visual sensitivity. Three main sensitivity variations are accounted for, namely, as a function of *light level*, *spatial frequency*, and *signal content*. Sensitivity S is defined as the inverse of the contrast C_T required to produce a threshold response, $S = 1/C_T$, where contrast is defined as $C = (L_{max} - L_{mean})/L_{mean}$, where L_{max} and L_{mean} refer to the maximum and mean luminances.

The variations in sensitivity as a function of light level are simulated by **amplitude nonlinearity**. Each input luminance L_{ij} is transformed by simplified version of the retinal response to an "amplitude non-linearity value" b_{ij} defined as: $b_{ij} = L_{ij}/(L_{ij} + 12.6L_{ij}^{0.63})$, where the constants 12.6 and 0.63 apply when luminance is expressed in cd/m^2. For this model the adaptation level for an image pixel is solely determined from that pixel.

The variations as a function of spatial frequency are modeled by the **contrast sensitivity function**, implemented as a filtering process. The Fast Fourier Transform (FFT) is applied to the values b_{ij}. The resulting magnitudes, $f_{uv}(b)$ are filtered by the CSF which is the function of the image size in degrees and light adaptation level L_m. The resulting contrast sensitivity filter $A_D(r(u,v))$ is given by:

$$A_D(r(u,v)) = (0.008/r^{1.5} + 1)^{-0.2} 1.42\sqrt{r}\exp(-0.3\sqrt{r})\sqrt{(1 + 0.06\exp(0.3\sqrt{r})},$$

where $r = u^2 + v^2$.

The variations in sensitivity due to a signal content are reffered to as *masking*. Masking effects are modeled by the **detection mechanism**, which is the most complicated element of the VDP. It consists of four subcomponents: image channeling, spatial masking, psychometric

function, and probability summation. *Image channeling* involves a decomposition similar to the *cortex transform* introduced by Watson [Wats87]. Cortex transform is a multi-resolution pyramid (see e.g. [Gonz02]) that simulates the spatial-frequency and orientation tuning of simple cells in the primary visual cortex (see Section 2.3). During the image channeling stage, the input image is fanned out from one channel to 31 channels or bands as follows. Each channel is associated with one cortex filter which consists of a *radial filter* (*dom*, difference of mesa filter) and an *orientational filter* (*fan* filter). The total number of radial filters is six resulting in five frequency bands and one base band. Each of these bands except for the base band is further fanned out into six channels of different orientation, see Figure 3.3. Thus, five frequency bands times six orientations per bands plus one base band results in 31 channels.

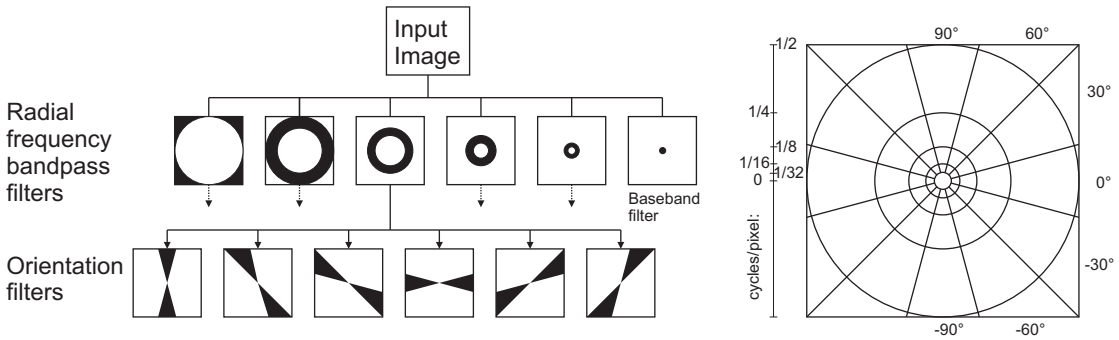


Figure 3.3: Cortex transform [Daly93]. On the left: organization of the filter bank. On the right: decomposition of the image frequency plane into the radial and orientation selectivity channels.

Spatial masking reduces the detectability of a given stimulus through the simultaneous presence of an additional suprathreshold stimulus. The masking depends on several factors, such as mutual masking, learning effects, the nature of masking signal, etc. Due to visual masking, threshold values can be elevated. This is accounted for after the transformation of all channels back to the spatial domain. For every channel and for every pixel, the elevation of the detection threshold is calculated based on the mask contrast for that channel and that pixel. Mutual masking can be considered by taking the minimal threshold elevation value for the corresponding channels and pixels of the two input images.

Psychometric function estimates the probability of detecting the differences for a given channel. The applied psychometric function describes the increase in the probability of detection as the signal contrast increases. Once the detection probabilities have been computed for each band of the filter hierarchy, the probability images are combined into single image by pooling together probability contributions from all bands as a function of position.

Difference Visualization There are two ways for visualizing the VDP output. The first technique is the *free-field difference map*, where the visible difference predictions appear on a uniform field with a gray value near the system mean. The second method, the *in-context difference map*, is the mapping of the output probabilities in color on the reference image. It is assumed that the difference can be perceived for a given pixel when the probability value is greater than 0.75, which is a standard threshold value for discrimination tasks.

3.2.2.5 Perceptual Distortion Measure by Teo and Heeger

Teo and Heeger [Teo94] presented a perceptual distortion measure based on so-called *normalisation model* – the nonlinear model of early phases of human vision. The model fits empirical measurements of the response properties of neurons in the primary visual cortext (see Section 2.1.3), and the psychophysics of spatial pattern detection. In the primary visual cortex, so-called *contrast gain control* mechanism keeps natural responses within the permissible dynamic range while at the same time retaining global pattern information. In the metric, contrast gain control is realized by an excitatory nonlinearity that is inhibited divisively by a pool of responses from other neurons. The channel decomposition process uses quadrature steerable filters with six orientation levels and four spatial resolutions. The distortion measure is computed from the resulting normalized responses by a simple squared-error norm to produce the difference map, similar to one produced by the VDP. Masking is modeled through contrast normalization and response saturation.

Authors in the paper [Teo94] thoroughly demonstrate that the proposed measure is far better than the MSE and illustrate the usefulness of the model in measuring perceptual distortion in real images.

3.2.2.6 Visual Discrimination Model

The Sarnoff Visual Discrimination Model [Lubi95] (VDM) is another image discrimination metric. The overall structure of the model is outlined in Figure 3.4. The VDM operates in the spatial domain. First, the inputs are convolved (refer to [Smit99]) with an approximation of the point spread function of the eye's optics. The signals are then re-sampled to reflect the photoreceptor sampling in the retina. A Laplacian pyramid [Burt83] is used to decompose the images into seven resolutions (each resolution is one-half of the immediately higher one), followed by band-limited contrast calculations. A set of orientation filters implemented through steerable filters of Freeman and Adelson [Free91] is then applied for orientation selectivity in four orientations. The CSF is modeled by normalizing the output of each frequency-selective channel by the base-sensitivity for that channel. Masking is implemented through a sigmoid non-linearity, after which the errors are convolved with disk-shaped kernels at each level. Finally, a distance measure or JND map is computed as the L_p-norm (Minkowski summation) of the masked responses.

The VDM is one of the few models that take into account the eccentricity of the images in the observer's visual field. It was later modified to the Sarnoff JND metric for color video [Wink00]. The complexity of Sarnoff VDM is $O(N)$, because the VDM operates in the spatial domain and avoids expensive FFT and FFT^{-1} transformations which take up to 40% of the execution time in the Daly VDP.

3.2.2.7 Gabor Pyramid Model of the HVS

Taylor et al. [Tayl97] presented a Gabor pyramid-based model of the human visual system (HVS) for image quality assessment. Their model departs from previous approaches in three ways:

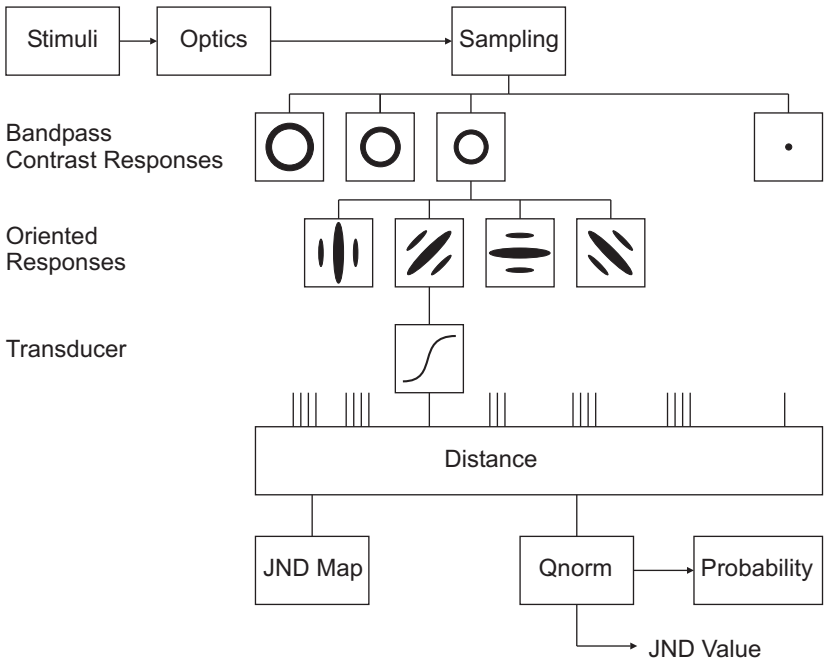


Figure 3.4: Block diagram of the Visual Discrimination Model [Lubi95].

- a physiologically and psychophysically plausible Gabor pyramid is used to model a receptive field decomposition
- psychophysical experiments are involved to directly assess the percept to be modeled
- the discrimination performance is modeled by using discrimination thresholds instead of detection thresholds

A number of physiological studies have confirmed the hypothesis that mammalian visual systems contain neurons whose receptive fields closely resemble Gabor patches [Tayl97]. Because of the physiological and psychological plausibility of Gabor decomposition, the proposed model involves the *Gabor pyramid*.

Similarly to other full-reference metrics, the model accepts two grayscale images as inputs and generates a probability map as output, see Figure 1.1 for illustration. A block diagram of the model is shown in Figure 3.5. A multiresolution decomposition is performed on each image to generate a number of channels, each containing the response of an ensemble of visual receptors. The receptors are modeled by Gabor functions of varying frequency and orientation. The multiresolution pyramid is built by lowpass filtering and decimating the original image. Each output image for a particular pyramid level is called *base image*. The base image for each pyramid level is convolved (refer to [Smit99]) with even and odd symmetric Gabor wavelets at eight orientations. The square root of the sum of the squares of the resulting even-odd image pairs describes the response of an ensemble of neurons tuned to a particular spatial frequency and orientation. These images are called the *channel images*.

The Psychometric Look Up Table (LUT) consists of a family of psychometric functions that have been empirically determined by psychophysical experiments described below. The Psy-

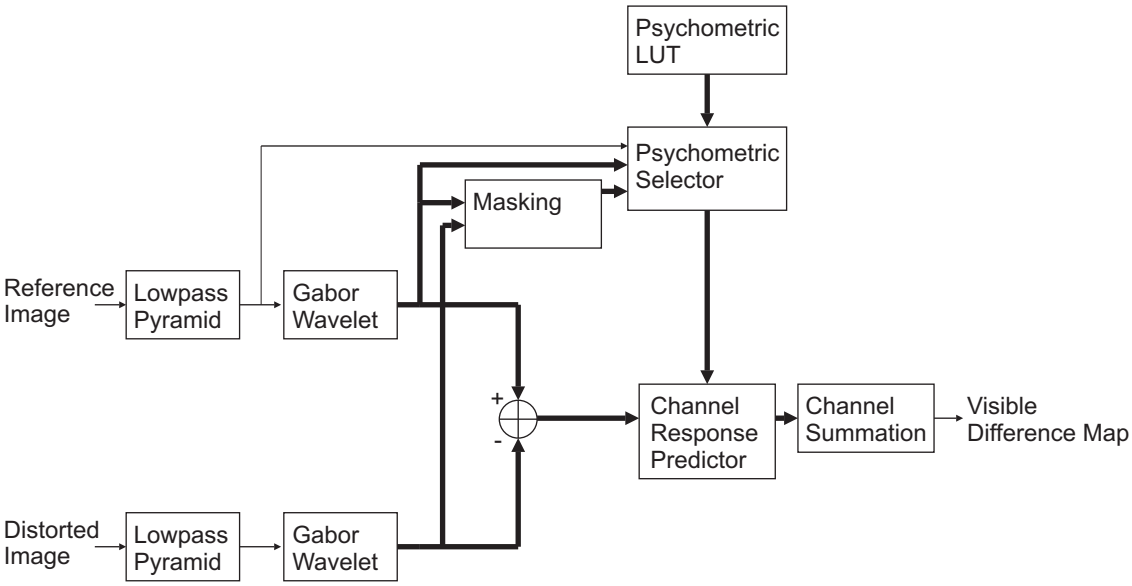


Figure 3.5: Block diagram of the Gabor Pyramid Model [Tayl97] (bold lines indicate parallel processing).

chometric Selector selects the appropriate psychometric function from the family of psychometric functions in the Psychometric LUT. A higher pyramid level base image determines the adaptation level, and the channel image determines the frequency, orientation, and reference contrast levels used to select the appropriate psychometric function. The difference between the contrast images for each channel is then applied to the appropriate psychometric function to produce a separate probability map for each channel. All of the probability maps from the different channels are combined using probability summation.

Two psychophysical experiments were conducted to determine the parameters of the model. The first experiment tested the visual system's sensitivity to Gabor patches as a function of spatial frequency, orientation, and average luminance. The second experiment compared the relation between detection and discrimination thresholds.

3.2.2.8 Stochastic Color Image Difference by Neumann et al.

Neumann et al. [Neum98b] proposed a stochastic approach to image comparison which operates in the original image space (i.e. there is no need for Fourier or wavelet transforms). The method places a limited number of random-sized rectangle pairs into compared images and determines an average color in each rectangle. For every rectangle pair the average color difference in CIELuv color space is computed and weighted according to the rectangle size and the contrast sensitivity function (CSF). The overall output of the method is obtained by combining the CSF weighted differences for all tested rectangles [Prik01]. Positions and orientation of rectangles are chosen quasi-randomly (Halton series), which makes the metrics deterministic (i.e. repeating the computation with a set of fixed input parameters leads to the equal comparison results).

3.2.2.9 Wavelet Visible Difference Predictor

Bradley's [Brad99] wavelet visible difference predictor (WVDP) is largely based on previously mentioned (see Section 3.2.2.4) visible differences predictor, but has a number of modifications that make it more amenable to potential integration into a wavelet based image comparison scheme. These modifications include the use of the *separable wavelet transform* instead of the cortex transform, the application of the *wavelet contrast sensitivity function,* and a simplified definition of sub-band contrast that allows prediction of noise visibility directly from wavelet coefficients, see Figure 3.6.

Another wavelet based metric has been proposed by Lai and Kuo [Lai00]. Their metric is based on the Haar Wavelet and the masking model can account for channel interactions as well as suprathreshold effects.

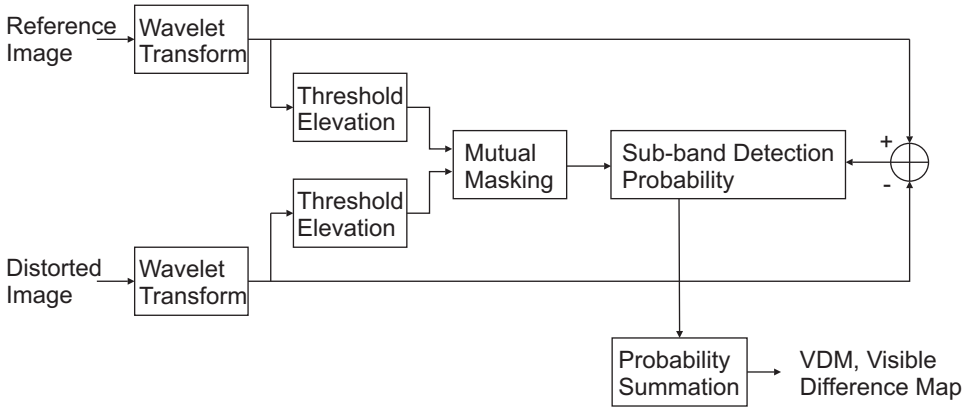


Figure 3.6: Block diagram of the Wavelet Visible Differences Predictor [Brad99].

3.2.2.10 S-CIELab and ST-CIELab

Tong et al. [Tong99] proposed a single-channel quality metric called *ST-CIELab* (spatio-temporal CIELab). ST-CIELab is an extension of the spatial CIELab (S-CIELab) image quality metric [Zhan96]. Both are backward compatible to the CIELab standard, i.e. they reduce to CIELab for uniform color fields. The ST-CIELab metric is based on a spatial, temporal, and chromatic model of human contrast sensitivity. The contrast sensitivity is modeled in an *opponent color space*, where the color information is encoded as white-black, red-green and blue-yellow color difference signals. After the CSF modeling, the data are transformed to CIELab space, whose difference formula is used for pooling.

3.2.2.11 Multistage Perceptual Quality Assessment Model

Multistage perceptual quality assessment model [Oh01] (MPQA) was proposed to compare original and lossy compressed digital angiogram images. As shown in Figure 3.7, the MPQA model includes amplitude nonlinearity, octave bandwidth spatial frequency decompositions into six orientations using Watson's cortex transformation [Wats87], and contrast masking

based on CSF modeling and region classification from the decomposed images. A perceptual distortion visibility map (PDVM) is produced via a distance computation and summation of efforts across different spatial frequency bands. A perceptual quality rating (PQR) is then calculated from the PDVM converting fidelity to quality, and transformed into a one to five scale, PQR_{1-5}.

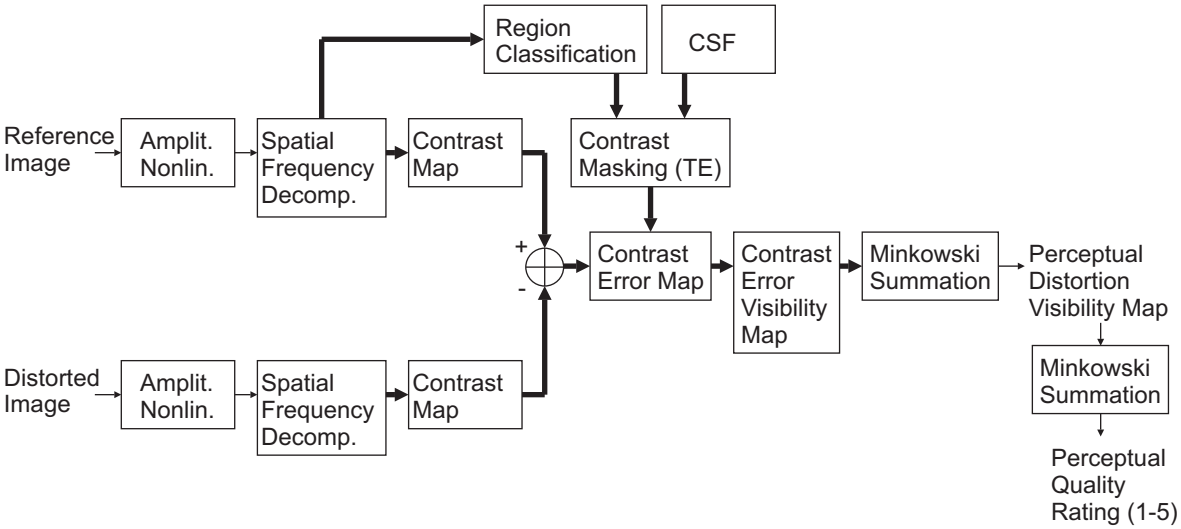


Figure 3.7: Block diagram of the MPQA model [Oh01] (bold lines indicate parallel processing).

As one may notice, the MPQA model is based on previously mentioned quality assessment models (mainly on the VDP, see Section 3.2.2.4), however it differs in the inclusion of contrast masking as a function of *background uncertainty*. The human eye can tolerate larger errors in high uncertainty stimuli (e.g. in textured areas) than in low uncertainty stimuli of the same contrast (e.g. along edges). The spatially decomposed images are therefore classified into *flat*, *edge*, and *texture* regions to consider the relationship between stimulus and background uncertainty. Flat regions are the areas with lower contrast than the base threshold contrast given by CSF. Edge areas are detected using a Sobel edge detector. Remaining regions are classified as texture regions. The threshold values are then increased according to the area type.

3.2.3 Summary

As we have seen, significant number of image quality assessment metrics have been proposed in the past. However, particularly in the field of image comparison and image and video quality assessment, there is still a lot of work to do.

- Most of the above mentioned approaches incorporate only a few of many factors that influence human perception, typically just the luminance adaptation and the contrast sensitivity. Nevertheless, the perception of *color* by human visual system is a very important task, for example it helps us to define the shapes in a perceived scene. Since many of mentioned approaches have only been applied to binary or grayscale images,

little consideration has been given to the analysis of full-color images. Even in the field of image segmentation, the development of segmentation algorithms based on color is still incipient [Font01].

- Almost all of above mentioned models are severely sensitive to shifts, dilatations, rotations, etc., inbetween the pair of input images. However, this does not correspond to the human practice, where such differences are not held so strictly.
- As most of perceptual image quality metrics are relative (full-reference) – they take the reference image and the examined image as inputs, there is a lack of *absolute* (no-reference) perceptual picture quality metrics.
- Contemporary perceptual image metrics typically consider the image as a whole. Therefore, there is a lot of work in comparison of images with *Regions Of Interest* (ROI), e.g. the JPEG 2000 [Taub02] ROI compressed images.
- What does the RMS error mean for 2D comparison, the METRO measure [Cign98] is in 3D. However, not as much attention as to 2D perceptual comparison was given to perceptual comparison of 3D models. Appropriate 3D error metric should involve HVS properties.
- We suppose that more work could be done to extend or replace existing perceptual comparison techniques, to produce a more perceptually robust solution even for *not* strictly photorealistic images.
- Since the amount of contemporary quality metrics is considerable, evaluation and comparison of them become an important issue. However, only a few *comparative studies* exist that have investigated the prediction accuracy of metrics in relation to others.
- So far, no *general-purpose metric* has been found that is able to replace subjective testing.

3.3 Structural Similarity Based Image Quality Assessment

Differently from the traditional error sensitivity based approach, structural similarity based image quality assessment has been proposed [Wang02a, Wang04]. Structural similarity approach is based on the philosophy that the main function of the human visual system is to extract structural information from the viewing field, and that the human visual system is highly adapted for this purpose. Therefore, a measurement of structural information loss can provide a good approximation to the perceived image distortion.

3.3.1 The Structural SIMilarity Index

The Structural SIMilarity Index (SSIM) [Wang04] is a specific example of a *structural similarity* quality measure. The structural information in an image is defined as those attributes that represent the structure of objects in the scene, independent of the average luminance and contrast. The diagram of the quality assessment using the SSIM is shown in Figure 3.8. (Note: the SSIM is a generalization of the Universal Quality Index [Wang02a]).

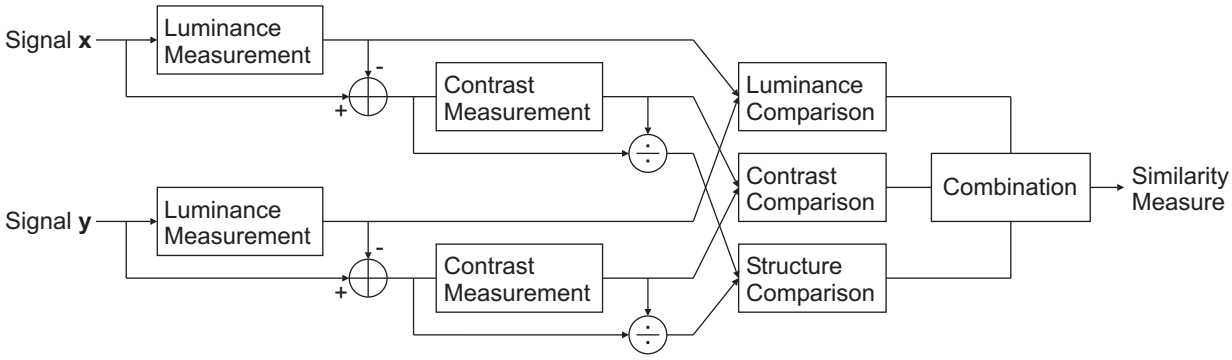


Figure 3.8: Diagram of the structural similarity (SSIM) quality measure [Wang04].

The SSIM separates the task of similarity measurement into three comparisons. First, the *luminance* of each signal x and y is compared. The luminance comparison function $l(x, y)$ is a function of μ_x and μ_y:

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

where the $\mu_x = 1/N \sum_{i=1}^{N} x_i$ (the mean intensity) is the estimate of luminance.

Second, the mean intensity is removed from the signal and the *contrast comparison* function is evaluated as follows:

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

where the $\sigma_x = (1/(N-1) \sum_{i=1}^{N}(x_i - \mu_x)^2)^{1/2}$ (the square root of variance) is the estimate of the signal contrast.

Third, the signal is normalized by its own standard deviation, so that the two signals being compared have unit standard deviation. The *structure comparison* is defined as follows:

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

using the correlation (inner product) σ_{xy} between the normalized signals computed as $\sigma_{xy} = 1/(N-1) \sum_{i=1}^{N}(x_i - \mu_x)(y_i - \mu_y)$. Where C_1, C_2, and C_3 are constants included to avoid instabilities.

Finally, the three components are combined to yield an overall similarity measure (structure similarity index):

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

where $\alpha > 0$, $\beta > 0$, and $\gamma > 0$ are parameters used to adjust the relative importance of the three components. This definition satisfies the conditions of symmetry: $SSIM(x, y) = SSIM(y, x)$, boundedness: $SSIM(x, y) \leq 1$, and unique maximum: $SSIM(x, y) = 1$ iff $x = y$. The Universal Quality Index (UQI) corresponds to the special case that $C_1 = C_2 = 0$, which produces unstable results when either $(\mu_x^2 + \mu_y^2)$ or $(\sigma_x^2 + \sigma_y^2)$ is very close to zero.

For *image quality assessment*, it is useful to apply the SSIM index locally. Localized quality measurement can provide a spatially varying quality map of the image, which delivers more information about the quality degradation of the image. The local statistics μ_x, σ_x, and σ_{xy} are computed within a local window, which moves pixel-by-pixel over the entire image. When a single overall quality measure of the entire image is required, we use a mean SSIM (MSSIM) index to evaluate the overall image quality:

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

where X and Y are the input images, x_j and y_j are the image contents at the j-th local window, and M is the number of samples in the quality map.

3.3.2 An SVD Based Gray-Scale Image Quality

Shnayderman et al. [Shna04, Shna06] published a gray-scale metrics (called MSVD) based on a singular value decomposition. Authors show that the proposed simple metric is able to measure the distortion within a distortion type at different distortion levels and also across different distortion types.

The Singular Value Decomposition (SVD) is a factorization of a rectangular matrix. Every real matrix A can be decomposed into a product of three matrices $A = U \cdot S \cdot V^T$, where U and V are orthogonal matrices, and S is a diagonal matrix $S = diag(s_1, s_2, ...)$. The entries of S are called the singular values of A, the columns of U and V are called left and right singular vectors of A, respectively [Shna04]. If the SVD is applied to the full images, we obtain a *global measure*, whereas if a smaller block (e.g. 8×8 pixels) is used, we compute the *local error*.

The *local error* D_i (distortion map) is computed as the distance between the singular values of the original image block and the singular values of the distorted image block:

$$D_i = \sqrt{\sum_{i=1}^{n}(s_i - \hat{s}_i)^2},$$

where n is the block size, and s_i, $\hat{s}_i$ are singular values of the original and distorted block, respectively.

The *global measure* is derived from the local measure as follows:

$$MSVD = \frac{\sum_{i=1}^{M} |D_i - D_{\text{med}}|}{M},$$

where $M = (k/n) \cdot (k/n)$ is the number of samples in the distortion map, k is the image size, n is the block size, and D_{med} represents the median of the sequence of D_i.

3.3.3 Gradient-Based Structural Similarity

By studying the SSIM (refer to Section 3.3.1), Chen et al. [Chen06] recently realized that it does not assess differences for blurred images well. To rectify this, authors proposed an improvement called Gradient-based Structural SIMilarity (GSSIM).

According to the SSIM, the GSSIM performs three comparisons (i.e. luminance, contrast and structure comparisons). However, differently from the SSIM, the contrast comparisons and the structure comparisons are based on gradient operations. Specifically, authors use the Sobel operator [Gonz02] to construct the *gradient map* as follows. For each pixel $p_{i,j}$, a gradient vector is computed: $\vec{E}_{i,j} = (dx_{i,j}, dy_{i,j})$, where dx and dy are obtained by the Sobel operator using a mask of the size 3×3 pixels. Magnitude of this vector is approximated by $Mag_{i,j} = |dx_{i,j}| + |dy_{i,j}|$ and magnitudes then form the image's gradient map.

The gradient-based *contrast comparison* $c_g(x, y)$ and the *structure comparison* $s_g(x, y)$ are consecutively evaluated as follows:

$$c_g(x, y) = \frac{2\sigma_{x'}\sigma_{y'} + C_2}{\sigma_{x'}^2 + \sigma_{y'}^2 + C_2},$$

$$s_g(x, y) = \frac{\sigma_{x'y'} + C_3}{\sigma_{x'}\sigma_{y'} + C_3},$$

where the x' and y' represent blocks of the gradient map, the $\sigma_{x'}$ and $\sigma_{y'}$ are the standard deviations of vectors x' and y' respectively, $\sigma_{x'y'}$ is the covariance of x' and y', and C_2, C_3 are small constants.

The luminance, contrast and structure components are combined in a same way as in the SSIM case (see Section 3.3.1) to yield an overall similarity measure. Experimental results have shown that for blurred images the GSSIM is more consistent with HVS than the SSIM.

3.4 Selection of Models for Comparison

Several studies on performance of traditional perceptual image quality models have been published in history [Li98, Jack97, Erik98, Wats00]. However, we are not aware of any independent evaluation of traditional and structural similarity approaches, apart from the article where the SSIM model was introduced [Wang04].

For comparison of the traditional error-sensitivity and structural similarity based approaches we have chosen a representative from each group. The Visible Differences Predictor (VDP), see Section 3.2.2.4, is a typical example of an image quality metric based on *error sensitivity,* whereas the Structural SIMilarity index (SSIM), see Section 3.3.1, is a specific example of a *structural similarity* quality measure. We have included the responses of the Universal Quality Index (UQI) into the evaluation as well (the UQI is a special case of the SSIM), but the obtained results of the UQI were generally worse than the results of SSIM, and therefore we decided to report just the SSIM results.

The VDP and the SSIM are full-reference metrics, i.e. the input to the selected metrics consists of two images and (in the VDP case) also of the parameters for viewing conditions, whereas the output is a map describing the visible differences between them. The output map defines

the probability of detecting the differences between the two images as a function of their location in the images.

3.5 Measurements

First, we performed a subjective experiment to obtain subjective opinion scores, i.e. the ground truth. Then, we executed the objective testing using the VDP and the SSIM using the same group of inputs. The results of the subjective and objective experiments were then confronted to evaluate the performance of the inquired models.

3.5.1 Subjective Testing

As visual stimuli we have used 32 grayscale photos of an urban construction site, see Table 3.1. We have conducted two experiments: in the first experiment all the input images were compressed uniformly using JPEG 2000 (the quality of compression was intentionally varied in the interval of 30–90, 100 meaning no compression). In the second experiment, the 32 original input images were also compressed using JPEG 2000, but with ROI (regions of interest), i.e. images contained manually-specified regions of interest compressed with better quality than the other areas of the image (the ROI was typically not compressed, while the rest of the image was compressed extensively). Note that the exact compression ratios are not significant for the subsequent evaluation.

In both experiments, the original (not compressed) image was presented alongside the compressed image on a calibrated CRT monitor[1]. Subjects sat approximately 60 cm in front of the monitor by a dimmed office illumination of approximately 250 lux. The total number of 32 subjects (aged between 18 and 57) were asked to express the difference between the original and compressed images by ratings (on the scale 1–10, where 10 represents the indistinguishable difference, i.e. the best result, while 1 denotes the worst result). Subjects had normal or corrected-to normal vision and were non-experts in the field of image comparison. Prior to the evaluation, each subject was verbally introduced to the experiment and he/she had enough time to adapt to viewing conditions. The procedure took approximately 30 minutes for each subject.

3.5.2 Objective Testing

The pairs of images used in the subjective experiments were used also as the input stimuli for both the VDP and the SSIM models. The parameters of the VDP model were set properly to correspond with the subjective experiment observing configuration (as described in previous section).

Both the VDP and SSIM models are relative since they do not describe an absolute value of image quality, but instead they address the problem of differences between two images. The output is a map describing the visible differences between them, see Figures 3.9, 3.10 as examples. The output map defines the probability of detecting the differences between the two images as a function of their location in the images. An advantage is that we can see

[1] LG Flatron 795FT Plus, 17-inch CRT monitor, characterized by ICC profile (D65).

Image	MOS	VDP	SSIM	Image	MOS	VDP	SSIM
	**8.08** *5.77*	**61.2** *62.1*	**0.92** *0.89*		**9.13** *5.17*	**54.9** *56.7*	**0.94** *0.90*
	**8.28** *5.06*	**69.1** *70.4*	**0.92** *0.86*		**7.88** *4.00*	**79.6** *80.9*	**0.85** *0.69*
	**9.40** *8.65*	**48.1** *48.2*	**0.96** *0.96*		**8.67** *5.68*	**61.1** *61.7*	**0.94** *0.89*
	**8.67** *5.32*	**89.5** *91.0*	**0.83** *0.69*		**7.68** *4.93*	**62.2** *64.1*	**0.90** *0.83*
	**7.79** *6.03*	**46.9** *47.4*	**0.93** *0.90*		**7.13** *2.61*	**73.8** *75.6*	**0.86** *0.66*
	**8.32** *2.97*	**53.2** *56.1*	**0.91** *0.78*		**9.21** *6.42*	**70.4** *73.7*	**0.84** *0.66*
	**7.08** *4.74*	**68.3** *69.1*	**0.87** *0.82*		**9.22** *5.68*	**77.2** *78.4*	**0.87** *0.75*
	**7.96** *6.13*	**69.3** *69.9*	**0.82** *0.75*		**9.38** *7.53*	**64.4** *64.6*	**0.93** *0.85*

Table 3.1: Numerical results of the evaluation of image quality metrics. MOS denote subjective mean opinion scores, VDP and SSIM denote VDP and SSIM results, respectively. Values are shown for both experiments – for uniformly (in bold) and for ROI (in italics) compressed images. Only original input images are shown, refer to thesis webpages for distorted images.

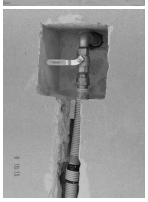
Image	MOS	VDP	SSIM
	**9.25** *7.10*	**63.0** *63.6*	**0.93** *0.89*
	**9.29** *6.83*	**59.1** *59.1*	**0.92** *0.80*
	**9.13** *4.83*	**70.1** *72.6*	**0.92** *0.82*
	**9.04** *7.13*	**73.5** *75.2*	**0.90** *0.83*
	**8.63** *3.90*	**74.5** *79.4*	**0.90** *0.75*
	**8.75** *6.37*	**66.1** *68.6*	**0.92** *0.83*
	**8.88** *4.30*	**75.8** *78.3*	**0.92** *0.81*
	**9.04** *7.90*	**42.7** *43.4*	**0.82** *0.77*

Image	MOS	VDP	SSIM
	**8.29** *3.48*	**57.3** *59.9*	**0.92** *0.81*
	**8.04** *6.41*	**72.1** *72.7*	**0.87** *0.81*
	**9.22** *4.28*	**77.4** *80.6*	**0.91** *0.76*
	**8.96** *6.69*	**76.6** *77.8*	**0.92** *0.83*
	**8.54** *7.41*	**70.2** *72.2*	**0.90** *0.85*
	**7.87** *7.31*	**66.9** *68.2*	**0.90** *0.86*
	**8.00** *6.14*	**68.2** *69.7*	**0.92** *0.87*
	**8.71** *4.48*	**82.4** *84.1*	**0.93** *0.78*

Table 3.1: Numerical results of the evaluation of image quality metrics (cont.). MOS denote mean opinion scores, VDP and SSIM denote VDP and SSIM results, respectively. Values are shown for both experiments – for uniformly (in bold) and for ROI (in italics) compressed images. Only original input images are shown, refer to thesis webpages for distorted images.

the nature of the difference and we can use this information for further improvement of the design.

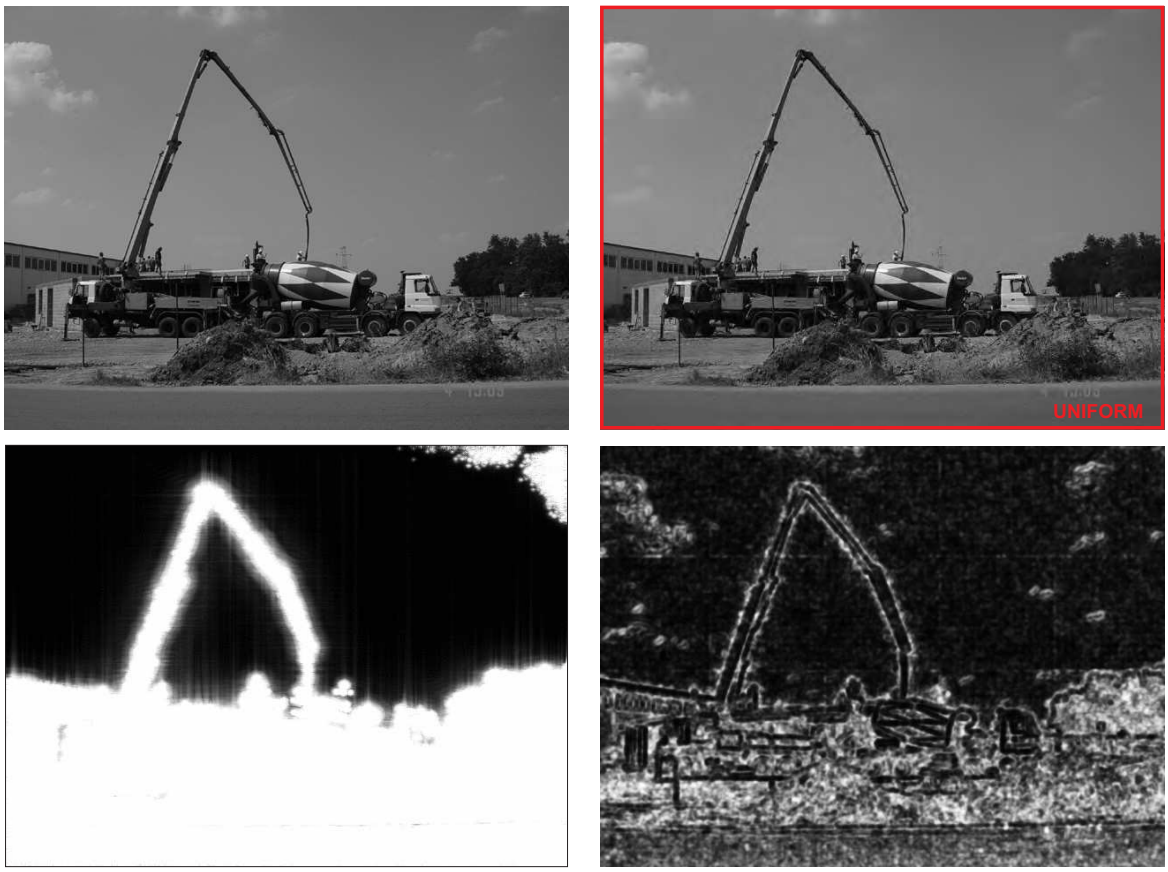


Figure 3.9: Example of image quality assessment using the VDP and the SSIM – uniformly compressed image. Original image (top left), ROI compressed image (top right), VDP detection probability map (bottom left), SSIM detection probability map (bottom right). Probability maps: black pixel means zero probability of difference detection, white pixel means 100% probability.

However, since we also need a single overall quality measure, we use a mean SSIM index (MSSIM) in the case of the SSIM model. MSSIM index values are in range [0, 1], where 0 means the worst quality, while 1 denotes the indistinguishable difference between the input images. For the VDP model, we employ the following approach [Mysz98]: the numerical value of the difference between images ($D_{0.75}$) is the percentage of pixels for which the probability of difference detection is greater than 0.75. It is assumed, that the difference can be perceived for a given pixel when the probability value is greater than 0.75 (75%), which is the standard threshold value for discrimination tasks. This output value therefore ranges in [0, 100], where 0 means the best result (no pixel with probability of difference detection greater than 0.75), while 100 means that all the pixel differences are above the difference detection threshold (the worst result).

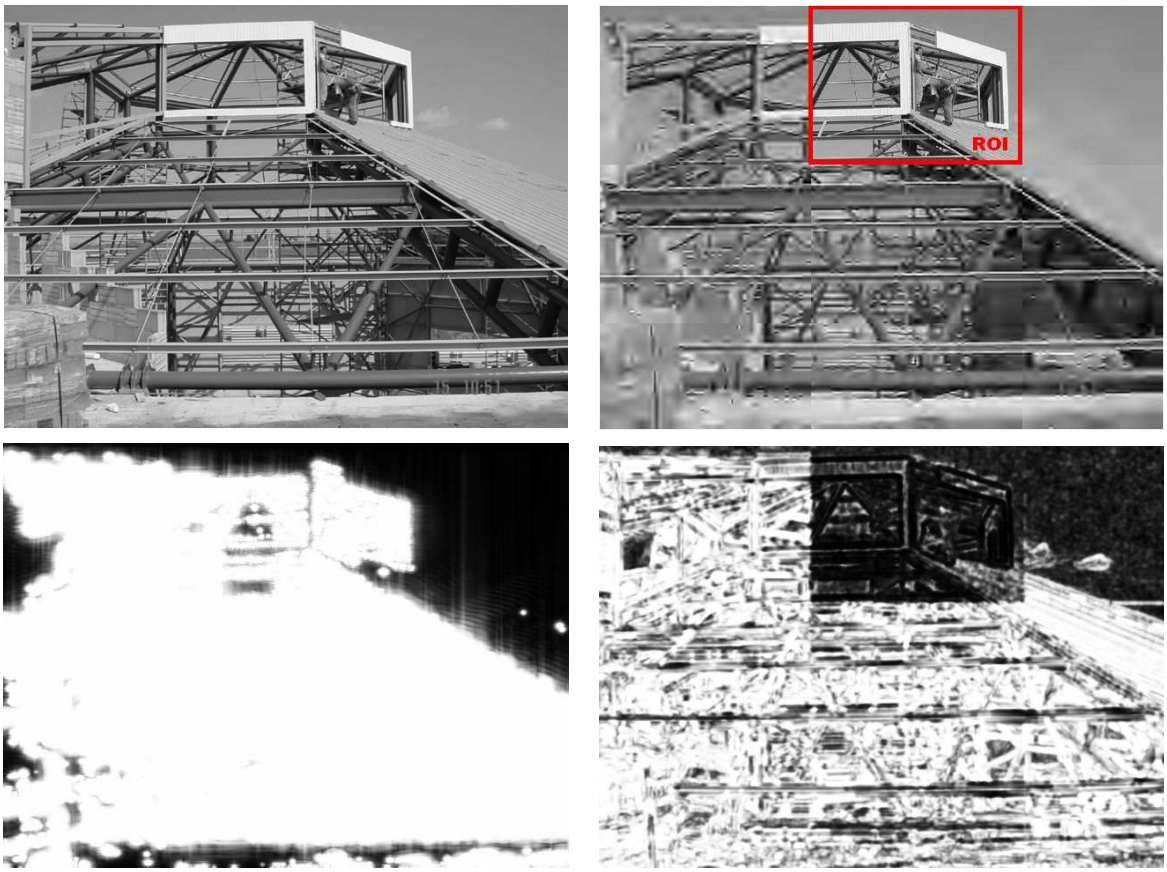


Figure 3.10: Example of image quality assessment using the VDP and the SSIM – ROI compressed image. Original image (top left), ROI compressed image (top right), VDP detection probability map (bottom left), SSIM detection probability map (bottom right). Probability maps: black pixel means zero probability of difference detection, white pixel means 100% probability.

3.5.3 Test Results

Prior to the analysis of the data, we checked the subjective scores for completeness, and for potential outliers. However, we were not forced to omit any subjective score during this phase. Mean opinion scores (MOSs) were then computed for each image pair from the normalized raw scores of each subject. The normalization was performed as follows: $x_i := (x_i - \mu_x)/\sigma_x$ where x_i is a score and μ_x and σ_x are the mean and the standard deviation for a subject respectively.

As a visual illustration of the relationship between subjective data and model predictions, scatter plots of MOS versus the VDP and the SSIM predictions are shown in Figure 3.11. Each point in a plot represents one test image, with its vertical and horizontal coordinates representing its subjective MOS and the model prediction, respectively. As we can see, the SSIM results exhibit better consistency with the subjective data than the results of the VDP.

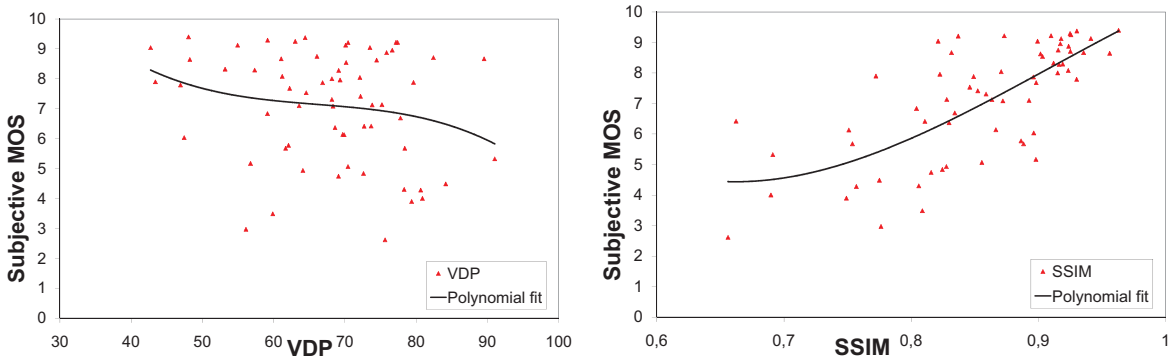


Figure 3.11: Quality predictions compared to subjective mean opinion scores (MOS) for the Visible Differences Predictor (left) and for the Structural SIMilarity index (right).

One of the most important factors for full-reference image quality metrics is the *correlation* of their results to the perceived image quality judged by humans. For the numerical evaluation we use both the standard (Pearson's) and the non-parametric (Spearman's) correlation coefficients, see Appendix A.2. We use the 3rd order polynomial fit function prior to computation of the correlation coefficients, because the mapping of the objective model outputs to the subjective MOS is generally non-linear[2]. Then, we use the non-linear regression functions to transform the set of model outputs to a set of predicted MOS values. Finally, we compute correlation coefficients between these values and the subjective MOS, see Table 3.2 and Figure 3.12. These results are summarized in Table 3.1 and discussed in the following section.

3.6 Discussion

In this section, we discuss the performances of the models. First, we discuss the performances for uniformly compressed input images, then we consider the results for ROI-compressed images. Finally, we point out the advantages of each model.

3.6.1 Quality Assessment Performances

Calculated values of the Pearson's (r) and Spearman's (r_s) correlation coefficients are presented in Table 3.2. Note that Pearson's r is the measure of *prediction accuracy* of a model, while Spearman's r_s assumes only the monotonic relationship between the MOS and model predictions and it is therefore the measure of *prediction monotonicity* of a model, see also Appendix A for explanation of the statistical methods. Pearson's correlation coefficients are visualized in Figure 3.12 along with 95% confidence intervals. The overall performances (i.e. the values in the first two columns of the Table 3.2) show that the correlation to the MOS is much better for the SSIM model than for the VDP. The results imply that there is a statistically significant difference between the VDP and SSIM model, in favor of the

[2]Note that similar approach was adopted by Video Quality Experts Group for evaluation of video quality assessment models [VQEG00].

SSIM. The absolute values of the SSIM r's are slightly lower than the previously published results [Wang04]. This discrepancy is probably caused by the selection of input stimuli.

As one may observe in the third and the fourth column of Table 3.2 and in Figure 3.12 (right), the performances of models in the ROI task were rather poor and there was not a statistically significant difference between the models. This has led us to experiments with the ROI functions, as described in the following section.

Model	r	r_s	r (ROI)	r_s (ROI)
VDP	0.22	0.20	0.44	0.39
SSIM	0.72	0.74	0.51	0.45

Table 3.2: Quality assessment performances of the VDP, and the SSIM models for uniformly and ROI compressed images. (r denotes Pearson's, and r_s means Spearman's correlation coefficients.)

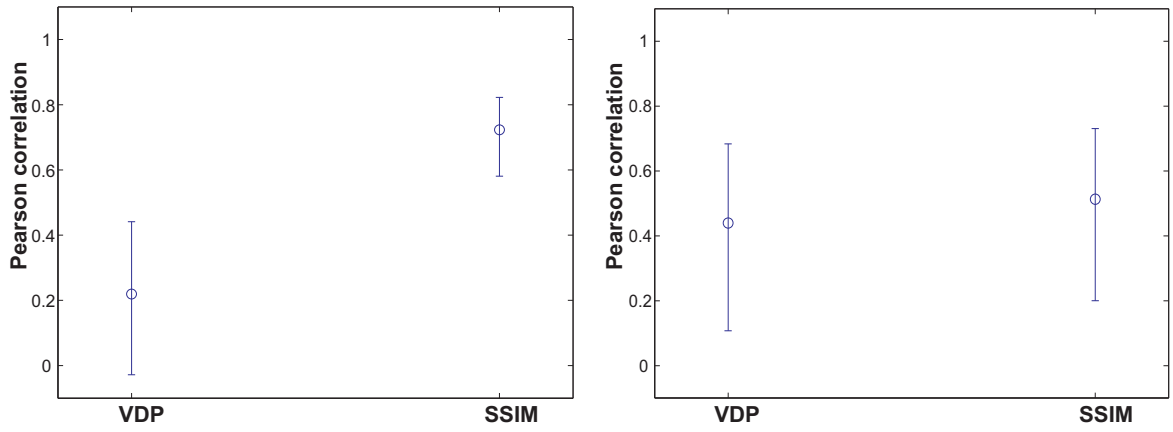


Figure 3.12: Quality assessment performances (Pearson's correlations with 95% confidence intervals) of the VDP, and the SSIM models. Left: overall performance, Right: performance in ROI compression experiment.

3.6.2 ROI Quality Assessment

Half of the input stimuli in our experiments were images compressed with manually-defined ROIs. Since we had an explicit information about the ROI structure for each of the input images we were able to incorporate it into the computation of the VDP and the SSIM responses. Based on the information about the ROI structure (spatial arrangement and compression ratios) we have constructed the "ROI functions" that we have used to scale the output probability maps. We have experimented with various ROI functions including the absolute values of compressions, smoothed and reduced range function, approximation by Gaussians, and the inversion of the absolute values, as illustrated in Figure 3.13.

For all of the tested ROI functions the values of correlation coefficients between the scaled results and MOSs were less than in the non-ROI-aware case, both for the VDP and for the SSIM. We suppose that the ROI information must be involved directly during the computation

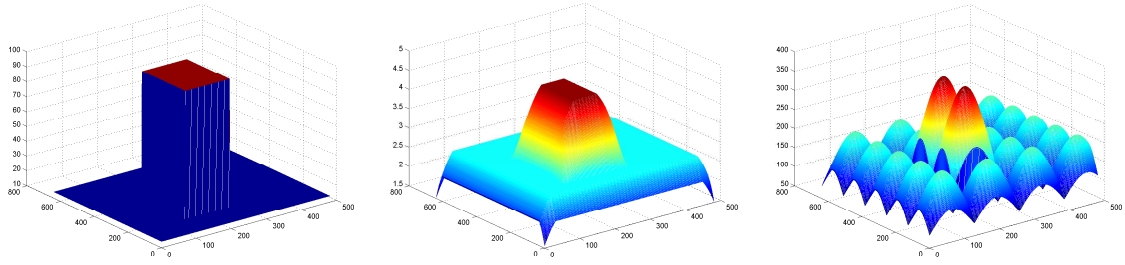


Figure 3.13: ROI functions. Absolute values (left), smoothed and reduced range (middle), approximation by Gaussians (right).

of an actual model taking into consideration such issues as the foveation, etc. This offers a wide area for future research and thorough psychophysical testing.

On the other hand, the SSIM model *detects* the regions of interest in the image quite well, as we can see for example on the bottom right image in Figure 3.10. This feature of the SSIM would be used to assess the locations of regions of interest automatically. However, we suppose that visual attention-aware models [Osbe98] would perform even better than the SSIM in this task.

3.6.3 Advantages of Evaluated Models

The *VDP model* is based on generally-held human visual system assumptions and it is able to handle various phenomena such as the visual masking. The consistency of the VDP model with the HVS has been previously verified [Mart98, Mysz98] and the model was successfully applied to various issues not just in the area of image quality assessment, but also in computer graphics algorithms [Mysz98].

The correlation coefficients between our experimental subjective data set and the *SSIM model* responses exhibit that the SSIM model shows better consistency with subjective data than the VDP model. As we have seen, the SSIM model is able to detect regions of interest in the image. This feature is promising for future research on ROI issues. The SSIM model is simple to implement in contrast to the VDP model, and the MatLab code is publicly available. The computation of the SSIM does not require time consuming Fourier transformations (as the VDP does) and it is certainly faster than the computation of the VDP model.

3.7 Summary

Image quality assessment models are relevant for various computer graphics applications. In this chapter, we have presented an independent comparison of two image quality assessment approaches. We have evaluated the VDP model and the SSIM model as the representatives of the traditional perceptual approach and the structural similarity based approach, respectively. We have described the subjective experiments that we have conducted to obtain mean opinion scores both for uniformly and ROI compressed images.

The evaluation of subjective results and predictions of the models shows that the structural based approach outperforms the traditional approach for involved input stimuli. As the implementation of the SSIM model is more straightforward than the implementation of the VDP, we conclude that the SSIM model is a significant alternative to the thoroughly verified VDP model. The SSIM model is able to detect the ROIs in the image. However, both models perform poorly in the ROI image assessment task leaving a lot of space for further improvements.

4 Perceptual Evaluation of Tone Mapping Methods

"Science never solves a problem without creating ten more."

(George Bernard Shaw)

The problem of reproducing High Dynamic Range (HDR) images on devices with restricted dynamic range (see Figure 1.2 on page 3 for illustration) has gained a lot of interest in computer graphics and image processing communities. There exist various approaches to this issue, which span several research areas including computer graphics, image processing, color vision, physiological aspects, etc. These approaches assume a thorough knowledge of both the objective and subjective attributes of an image. However, no comprehensive overview and analysis of such attributes has been published so far.

In this chapter, we present an overview about the effects of basic image attributes in HDR tone mapping. Furthermore, we propose a scheme of relationships between these attributes, leading to the definition of an overall image quality measure. We present results of subjective psychophysical experiments that we have performed to prove the proposed relationship scheme. Moreover, we also present an evaluation of existing tone mapping methods (operators) with regard to these attributes. Finally, the execution of with-reference and without a real reference perceptual experiments gave us the opportunity to relate the obtained subjective results.

Our effort is not just useful to get into the tone mapping field or when implementing a tone mapping method, but it also sets the stage for well-founded quality comparisons between tone mapping methods. By providing good definitions of the different attributes, user-driven or fully automatic comparisons are made possible.

4.1 Introduction

The dynamic range of visual stimuli in the real world is extremely large. A HDR image can be generated either synthetically or acquired from the real world, but the conventional media used to present these images can only display a limited range of luminous intensity. This problem, i.e. displaying high contrast images on output devices with limited contrast, is the task of *high dynamic range imaging*, and it is approached by HDR tone mapping, see Figure 1.2. A number of different tone mapping methods (operators) have been proposed in history [Devl02, Rein05b]. However, also due to their sheer number, the advantages and disadvantages of these methods are not immanently clear, and therefore a thorough and systematic comparison is highly desirable.

The field of tone mapping (TM) assumes extensive knowledge of findings from various scientific areas. In order to conduct a comparison of TM methods, it is necessary to settle upon a set of *image attributes* by which the images produced by the methods should be judged.

These attributes are not independent, and their interrelationships and the influence on the overall image quality need to be carefully analyzed. This is useful not just for comparing existing HDR approaches, but for evaluating *future ones* as well. The human visual system (HVS) is extremely complex and, besides highly focused laboratory studies, there is a lack of comprehensive user experiments we could build on.

In this chapter, we give a comprehensive list of most of the important attributes involved in the evaluation of a TM method, and we show which relationships exist between the basic attributes by means of two different subjective testing methods. Namely, we investigate the perceived quality of the images produced by particular TM methods with and without the possibility of direct comparison to the original real-world scenes. The evaluation of the attributes and their relationships leads to the definition of an *overall image quality*. This metric can be used to judge how well a given TM method is able to produce naturally looking images. Furthermore, we present the most comprehensive comparison to date in terms of the number of TM methods considered, including 14 different methods.

The chapter is organized as follows. Section 4.2 overviews the previous work on comparison of TM methods and other related work. In Section 4.3, we introduce and describe the term "overall image quality". In Section 4.4, we give a survey of the most important image attributes for tone mapping, and we describe how different methods reproduce these attributes. In Section 4.5 we propose a new scheme of relationships between the image attributes. In Section 4.6 we describe the two applied experimental methods based on human observations, and finally in Section 4.7, we show and discuss the results of these experiments.

4.2 Previous Work

The history of evaluation of TM methods is short. The following works (the only ones, to our best knowledge) were published only in the last few years. This is due to the recent increase in published TM methods on one hand, and due to the very high time, implementation, human, and other demands involved in such an evaluation on the other hand. While this section surveys the previous work, we relate our results to these works in Section 4.7.5.

4.2.1 Experimental Evaluations of Tone Mapping Methods

Drago et al. [Drag03b] performed a perceptual evaluation of six TM methods with regard to similarity and preference. In their study, observers were asked to rate a difference for all pairwise comparisons of a set of four HDR images tone mapped with six TM methods (24 images in total) shown on the screen. A multidimensional perceptual scaling of the subjective data from 11 observers revealed the two most salient stimulus space dimensions. The authors unfolded these dimensions as naturalness and detail and also identified the ideal preference point in the stimulus space. These findings were then used for a final ranking of the six TM methods.

In 2005, Yoshida et al. [Yosh05] compared seven TM methods using two real-world architectural interior scenes. The 14 observers were asked to rate basic image attributes (contrast, brightness, details) as well as the naturalness of the images. The results of this perceptual study exhibited differences between global and local TM methods. Global methods per-

formed better than local methods in the reproduction of brightness and contrast, however, local methods exhibited better reproduction of details in bright regions of images.

Kuang et al. [Kuan04] tested eight TM algorithms using ten HDR images. The authors implemented two paired comparison psychophysical experiments assessing the color and grayscale tone mapping performance respectively. In these tests, 30 observers were asked to choose the preferred image for each possible pair. The results showed the consistency of tone mapping performance for gray scale and color images. In the continuation of this research [Kuan05], Kuang et al. removed two TM methods and added two new images to the group of input stimuli. The authors examined the overall image preference (using paired comparison performed on an LCD desktop monitor) and preferences for six image attributes (using a rating scale) – highlight details, shadow details, overall contrast, sharpness, colorfulness, artifacts. The results show that shadow details, overall contrast, sharpness, and colorfulness have high correlations with the overall preference. More recently and parallel to our work, Kuang et al. [Kuan06b] used three indoor scenes and 19 subjects to evaluate 7 TM algorithms. Using two paired comparisons, the authors evaluated image contrast, colorfulness and overall accuracy. The results showed that bilateral filtering [Dura02] generated more accurate results than other algorithms. Results of the three experiments performed by Kuang and colleagues are summarized in [Kuan07].

Ashikhmin and Goyal [Ashi06], parallel to our work, demonstrated that using real environments is crucial in judging performance of TM methods. The authors compared five TM methods using four real-world indoor environments plus two additional HDR images. 15 subjects were involved in three ranking experiments: first two tests (preference and fidelity) were performed without ground truth while the third (fidelity) was conducted with reference (real scene). The results indicate that there is statistically no difference between preference and fidelity when there is no reference (i.e. equivalence of liking and naturalness criteria). However, the results show a difference in subject's responses for the fidelity test with reference and without reference.

4.2.2 Evaluations using HDR Displays

Ledda et al. [Ledd05] ran an evaluation of six TM methods by comparing to the reference scenes displayed on an HDR display. This HDR display allowed authors to involve many (23) input scenes. Subjects were presented three images at once (the reference and two tone mapped images) and had to choose the image closest to the reference. Statistical methods were used to process subjective data and the six examined methods were evaluated with respect to the overall quality and to the reproduction of features and details.

In the field of HDR displays, Yoshida et al. [Yosh06] analyzed the reproduction of HDR images on displays of varying dynamic range. The authors ran two perceptual experiments to measure subjective preferences and the perception of fidelity of real scenes. 24 participants, 25 HDR images and 3 real-world scenes were involved in the experiments. An outcome of this work is the analysis how users adjust parameters of a generic global TM method to achieve the best looking images and the images that are closest to the real-world scenes. Akyüz et al. [Akyu07] investigated how LDR images are best displayed on current HDR monitors. In two subjective experiments, authors exhibited 10 HDR images to 22 and 16 subjects, respectively. The results show that HDR displays outperform LDR ones and that LDR data

do not require sophisticated treatment to produce a HDR experience. More surprisingly, results show that tone mapped HDR images are statistically no better than the best single LDR exposure.

4.2.3 Other Related Studies

Some exciting contributions were published in the domain of image quality measurement of ordinary LDR images (see the book by Janssen [Jans01] for an overview on this topic). Rogowitz et al. [Rogo98] conducted two psychophysical scaling experiments for the evaluation of image similarity. The subjective results were compared to two algorithmic image similarity metrics and analyzed using multidimensional scaling. The analysis showed that humans use many dimensions in their evaluations of image similarity, including overall color appearance, semantic information, etc.

We find related work also in the field of psychophysical color research and photography, e.g. Fedorovskaya et al. [Fedo97] varied chroma of 4 input images to determine its effect on perceived image quality, colorfulness and naturalness. Results indicate that the enhancement of colorfulness leads to higher perceptual quality of an image. Savakis et al. [Sava00] performed an experiment on image appeal in consumer photography. While image quality is generally an objective measure, image appeal is rather subjective. During the experiment, authors showed 30 groups of prints to 11 people. The task of each subject was to select such a picture from each group that would receive the most attention in a photo album. Moreover, subjects had to comment the positive and negative attributes they used for the selection of the picture. The results show that the most important attributes for image appeal fall into the groups of composition/subject and people/expression, leaving objective attributes less significant.

Jobson et al. [Jobs02] investigated contrast and lightness in visually optimized LDR images. The authors approach the lightness as the image mean and the contrast as the mean of regional standard deviations. Inspecting these measures, the authors experimentally show that visually optimized LDR images are clustered about a single mean value and have high standard deviations, i.e. both the lightness and contrast are improved with the latter being more affected.

In the forthcoming paper, Mantiuk and Seidel [Mant08] show an application of their generic (black-box) TM operator to analysis of TM methods. Authors fit the generic operator to 12 TM methods to visualize their characteristics using fitted parameters of the generic operator. Moreover, authors apply the generic operator to HDR image compression. It is interesting to observe that global TM methods result in less distorted reconstruction than local ones, even though one would favor local methods to preserve more information.

4.2.4 Our Approach

Differently from the mentioned approaches, we adopt both a direct rating (with reference) comparison of the tone mapped images to the real scenes, and a subjective ranking of tone mapped images without real references. This enables us to confront the results from these two subjective experiments. Moreover, we present a methodology for evaluating TM methods using generally known image attributes. With 14 methods in total, and three typical real-

world HDR scenes, the subjective studies carried out to confirm this methodology also contain one of the most comprehensive comparison of TM methods.

In advance to the full experiment, we conducted an initial pilot study [Cadi05] to examine the experimental setup. It was observed that the overall image quality is not determined by a single attribute, but rather a composition of them. Next, we assessed [Cadi06] the results concerning the indoor scenes. Encouraged by these findings, we conducted a full experiment (we extended the input stimuli group by two another, different outdoor scenes), the results of which, including a thorough discussion, new statistical methodology etc. are presented in this chapter.

4.3 Overall Image Quality

In this section, we motivate and describe a measure which is useful for determining the performance of a particular TM method.

The first question is whether it is possible at all to find an optimal or "exact" method to tone map an arbitrary HDR input image, based on human vision. Unfortunately, the answer seems to be negative. Take for example a beach scene, where the absolute illuminance is often above 50,000 lux. A captured photograph of that scene, viewed under normal room illumination (about 200 lux), can never reproduce the same amount of colorfulness, because this is a *psycho-physiological* effect that depends on the absolute illuminance (vivid colors start to be perceived above 2000 lux). Therefore, a natural reproduction is only possible to a limited degree.

Another important question is the intent of the reproduction. The classical **perceptual** approach tries to simulate the human vision process and design the TM method accordingly. For example, a scene viewed at night would be represented blurred and nearly monochromatic due to scotopic vision, see Figure 4.1. However, if it is important to understand some fine details or the structure of the visible lines in the result, i.e. the content of the image, the same scene would be represented with full detail, which would be called the **cognitive** approach. If the goal is only the pleasant appearance of the image, we speak about an **aesthetical** approach. Any given TM method will realize a mixture of these three approaches, with a different weighting given to each.

In this chapter, we concentrate on the perceptual approach, and aim to characterize the *overall image quality* resulting from a TM technique in a perceptual sense. In addition, we have chosen a number of important image attributes which are typically used to characterize tone mapped images, and study how well TM methods reproduce these attributes: brightness, contrast, color, detail, and artifacts. The chosen attributes are mostly perceptual, but contain cognitive and aesthetics aspects as well. Beyond these attributes, which are related to color and spatial vision, there are some other important aspects and some "special effects" which can improve or modify the final appearance. Since some of the attributes are not mutually independent (as we will explain later), we propose a scheme of relationships between them (Figure 4.7). The goal of this work is to investigate the influence these attributes have on overall image quality, based on a subjective study.

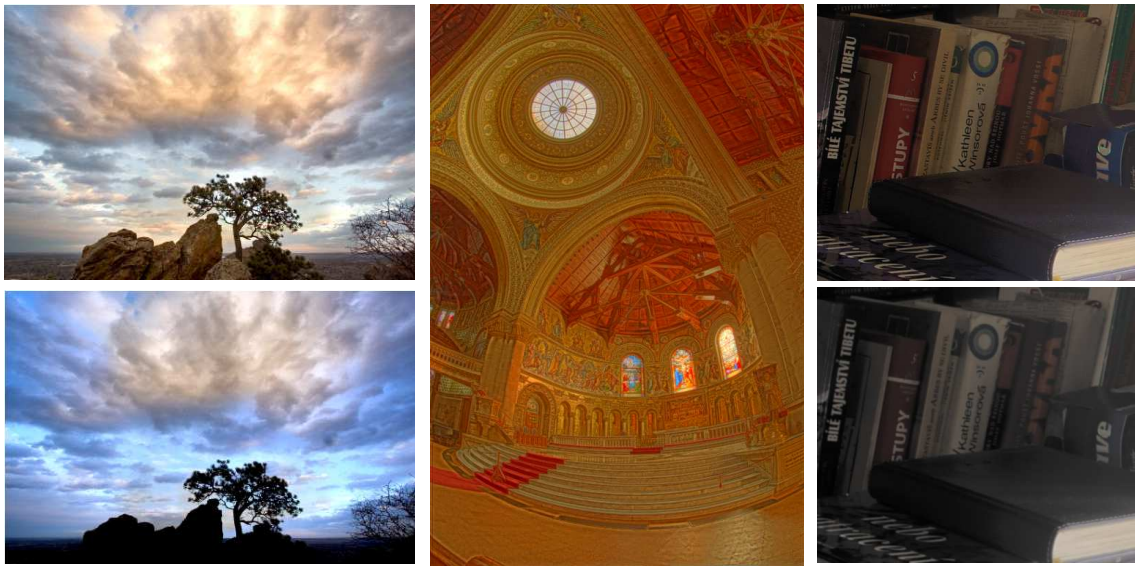


Figure 4.1: Three possible approaches to tone mapping. (From left:) two examples of many possible aesthetical renditions of input HDR image (adopted from [Lisc06]), information-rich (cognitive) rendition of Memorial Church HDR image (computed by [Tumb99a]), perceptual simulation (bottom right) of mesopic vision in the dark part of the original scene (top right).

4.4 Image Attributes

In this section, we briefly survey particular image attributes for tone mapping, and we list some typical TM methods that attempt to reproduce them correctly. As this part has the character of a survey, an informed reader can skip directly to the experiments described in Section 4.6.

4.4.1 Brightness

Brightness is a quantity that measures the subjective sensation produced by the absolute amount of luminance [Adel99]. More specifically, brightness is the attribute of a visual sensation according to which an area appears to emit more or less light [Wysz82]. The magnitude of brightness can be estimated for unrelated visual stimuli (since it is an absolute unit) as well as for related visual stimuli. *Lightness* is defined as the attribute of a visual sensation according to which the area in which the visual stimulus is presented appears to emit more or less light in proportion to that emitted by a similarly illuminated area perceived as a "white" stimulus [Wysz82]. Thus, lightness has meaning only for related visual stimuli. As lightness is judged with reference to the brightness of the "white" stimulus, it may be considered a special form of brightness measure that could be referred to as relative brightness [Wysz82]. In this study, we concern ourselves with the quality of reproduction of an overall brightness of the inquired HDR scene.

Stevens and Stevens, see [Fair05], proposed an expression for the apparent brightness, but although the expression gives a convenient relationship between luminance and brightness for simple targets, the overall brightness of an image is more complex. A method by Tumblin and Rushmeier [Tumb93] attempts to preserve the overall impression of brightness using a mapping function that is based on the model by Stevens and Stevens [Fair05]. This mapping function matches the brightness of a real world luminance to the brightness of a display luminance. Recently, Krawczyk et al. [Kraw06] proposed a method which aims for an accurate estimation of lightness in real-world scenes by means of the so-called anchoring theory of lightness perception. The method is based on an automatic decomposition of the HDR image into frameworks (consistent areas). Lightness of a framework is then estimated by the anchoring to the luminance level that is perceived as white, and finally, the global lightness is computed.

4.4.2 Contrast

Image contrast is defined in different ways, but it is usually related to variations in image luminance. There exist various basic formulae for computation of contrast, see the thesis by Winkler [Wink00] for an overview. Matkovic et al. [Matk05] proposed a complex computational global *contrast measure* called Global Contrast Factor that uses contrasts at various resolution levels in order to compute overall contrast. In this study, we think about overall contrast in a similar way.

Ward's [Ward94] initial TM method focuses on the preservation of *perceived contrast*. This method transforms input luminance to output luminance using a scaling factor. The computation of the factor is based on Blackwell's [CIE81] psychophysical contrast sensitivity model. Because Ward's method scales image intensities by a constant, it does not change scene contrasts for display. Almost the same principle of contrast preservation is exploited also in other methods [Ferw96, Lars97].

Advanced local tone mapping methods (e.g. the method by Reinhard et al. [Rein02] or by Ashikhmin [Ashi02]) are based on a multi-resolution decomposition of the image and approximate contrast in a way similar to Peli [Peli90], see Figure 4.2. Mantiuk et al. [Mant06] proposed a framework for perceptual contrast processing of HDR images. The authors define contrast as a difference between a pixel and one of its neighbors at a particular level of a Gaussian pyramid. This approach resembles the gradient-domain method by Fattal et al. [Fatt02].

4.4.3 Reproduction of Colors

The sensation of color is an important aspect of the human visual system (HVS), and a correct reproduction of colors can increase the apparent realism of an output image. One important feature of the HVS is the capacity to see the level of colors in a bright environment. This ability, measured as color sensitivity, is reduced in dark environments, as the light sensitive rods take over for the color-sensitive cone system, see Figure 4.3. As the luminance level is raised, the cone system becomes active and colors begin to be seen. Furthermore, the HVS has the capability of *chromatic adaptation*. Humans are able to adjust to varying

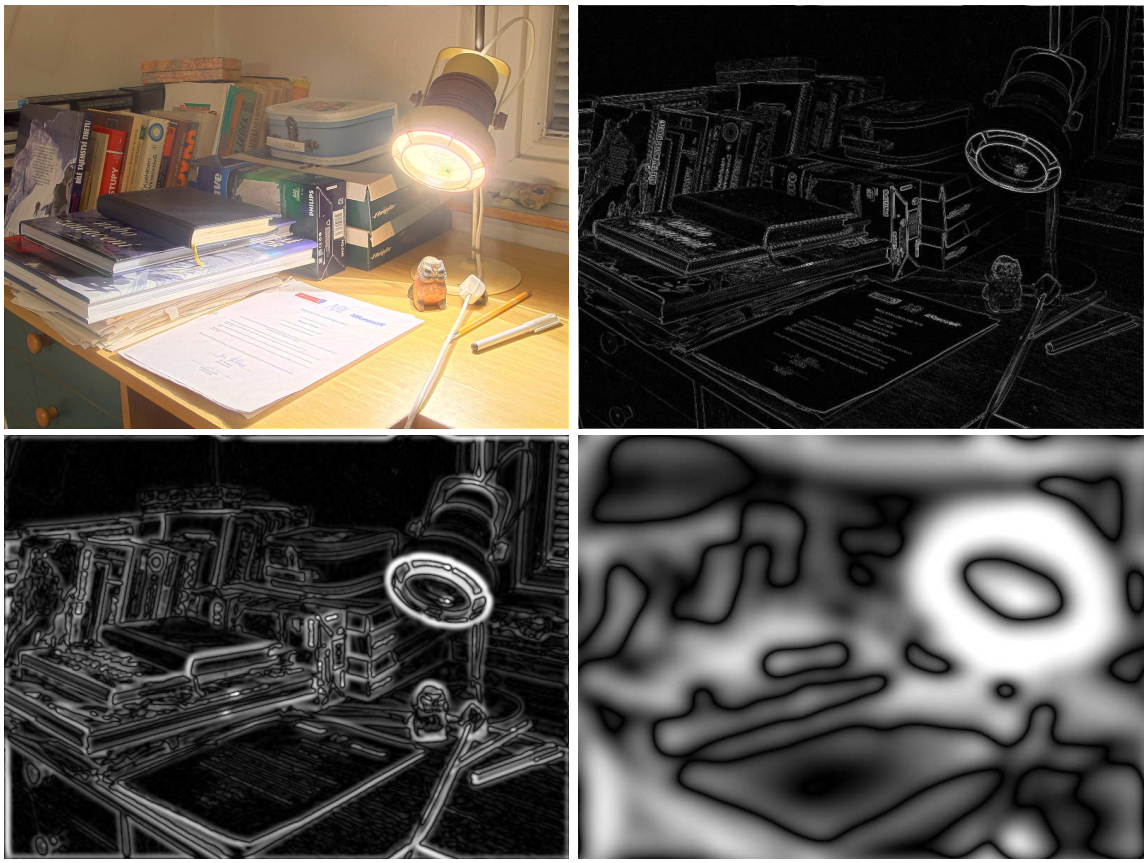


Figure 4.2: Peli's local band-limited contrast on three different spatial resolutions (top-left: original image).

colors of illumination in order to approximately preserve the appearance of object colors. See Fairchild's book [Fair05] for more information on color appearance modeling.

The TM method by Ferwerda et al. [Ferw96] captures changes in threshold color appearance by using separate TVI (threshold versus intensity) functions for rods and cones and interpolation for the mesopic luminance range. Ward et al. [Lars97] used a very similar approach. Pattanaik et al. [Patt98] proposed a comprehensive multi-scale model that accounts for changes both in threshold color discriminability and suprathreshold colorfulness. Using opponent color processing, the model is able to handle changes in chromatic and luminance-level adaptation as well. In their work, Reinhard and Devlin [Rein05a] adapted a computational model of photoreceptor behavior that incorporates a chromatic transform that allows the white point to be shifted.

4.4.4 Reproduction of Details

The reproduction of details is an issue mainly in very dark and very bright areas, because truncation of values occurs most frequently in these areas as a result of the dynamic range limitations of the output device. The simplest methods (e.g. linear scaling or clamping) will



Figure 4.3: Simulation of color sensitivity. Left: original image – no color sensitivity simulation, Right: simulation of the loss of color sensitivity in the dark.

usually reduce or destroy important details and textures (see Figure 4.4). On the other hand, the effort to reproduce details well is a potential cause of *artifacts*.



Figure 4.4: Reproduction of details in a very bright area. Left: global TM method exhibits the loss of details. Right: details preservation owing to mapping by a local method.

Several TM methods focus especially on the reproduction of details. Tumblin and Turk's LCIS method [Tumb99a] produces a high detail, low contrast image by compressing only the large features and adding back all small details. The idea of compressing just the large features and then adding subtle non-compressed details is also used in the methods based on the bilateral [Dura02] and trilateral filter [Chou03].

A different approach was presented by Ward [Lars97]. Ward's method based on histogram adjustment aims to preserve *visibility*, where visibility is said to be preserved if we can see an object on the display if and only if we can see it in the real scene. Ward's method does not strive to reproduce all the details available, but exploits the limitations of human vision to

reproduce just the *visible* details. Also, most local TM methods try to preserve detail along with contrast.

4.4.5 Artifacts

As a consequence of tone mapping, **artifacts** may appear in the output image. The artifacts are degrading the *overall quality* of the output image. Some local TM methods [Chiu93, Schl94] exhibit typical *halo artifacts*, see Figure 4.5. These artifacts are caused by contrast reversals, which may happen for small bright features or sharp high-contrast edges, where a bright feature causes strong attenuation of the neighboring pixels, surrounding the feature or high-contrast edge with a noticeable dark band or halo.

Another possible artifact of TM methods stems from the superficial handling of colors. Many TM methods use very simple rules in handling of the colors, e.g. doing the HDR to LDR transformation just for the luminance component with consequential restoration of the color information. Apart from poor values for the color reproduction image attribute, this can also lead to visible *color artifacts* like oversaturation, see Figure 4.5. Closely related to color artifacts are *quantization artifacts*, especially in dark regions, which stem from applying transformations (like gamma correction) to a low-precision representation of color values.

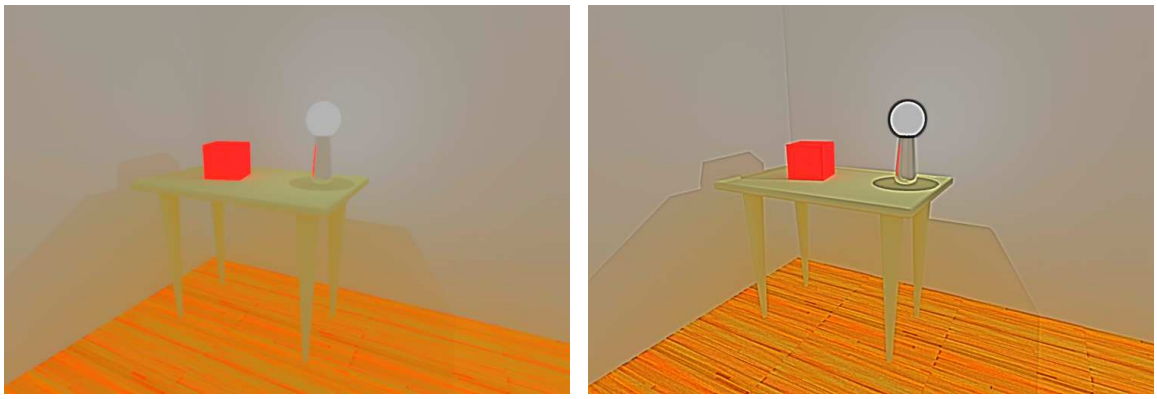


Figure 4.5: Halo artifacts and oversaturation. Left: HDR image after successful tone mapping without halo artifacts. Right: the same image after tone mapping using the local method exhibiting a massive amount of halo artifacts. Both images exhibit oversaturation.

4.4.6 Special Attributes

The following image attributes show up just under special conditions and we do not consider them in our current experiments, in favor of the basic ones. Moreover, we avoided testing of glare and visual acuity simulation, because these effects are usually implemented in the same way as a postprocess after the TM step. However, we present these attributes here to complete the survey of image attributes for tone mapping and it will be an interesting task to include them in future special evaluations.

Visual acuity is the ability of the HVS to resolve spatial detail. The visual acuity decreases in the dark, since cones are not responding to such low light levels. It is interesting that simu-

lating this phenomenon, i.e. reducing the detail in an image, actually enhances the *perceptual quality* of the image.

Owing to the scattering of light in the human cornea, lens, and retina, and due to diffraction in the cell structures on the outer radial areas of the lens, phenomena commonly referred to as **glare effects** [Spen95] are seen around very bright objects, see Figure 4.6. Since the dynamic range of traditional output devices is not sufficient to evoke such phenomena, we must simulate the human response artificially to improve the *perceptual quality* of the image.



Figure 4.6: Bloom (veiling luminance) simulation. Left: the original scene without bloom simulation, Rigt: the same scene with bloom simulation. *Source HDR image courtesy of Greg Ward.*

4.5 Attribute Relationships

In previous sections, we have surveyed the image attributes that are important for tone mapping and influence the overall quality of the output image. These attributes are not independent, and we present a description of their interrelationships in this section.

We propose the scheme shown in Figure 4.7 to illustrate the relationships between the attributes. The **overall image quality**, our measure, is determined by all the attributes. It depends strongly on the overall perceived *brightness*, i.e. highly illuminated scenes should be reproduced bright, while dim scenes should appear dark. Apparent *contrast* should also be reproduced well to make the result natural. The reproduction of *details* or rather the reproduction of *visibility* of objects is certainly essential to make the output image appear natural. Furthermore, since we are typically facing a limited display gamut, the reproduction of *color* is an important factor for perceptual quality as well. The simulation of *visual acuity* loss can significantly improve the perceptual quality of dim or night scenes, while the simulation of *glare* can enhance the perceptual quality of the dark scenes with strong light sources. There is no doubt that the presence of disturbing *artifacts* degrades perceptual quality. But there are also important interrelationships of the attributes:

The perception of **brightness** is affected greatly by the *contrast arrangement* (i.e. by the semantics of an image). Fairchild [Fair05] described the effect of image contrast on the perceived brightness and concluded that the brightness typically increases with *contrast*. It has been shown that brightness increases as a function of chroma (Helmholtz-Kohlrausch effect).

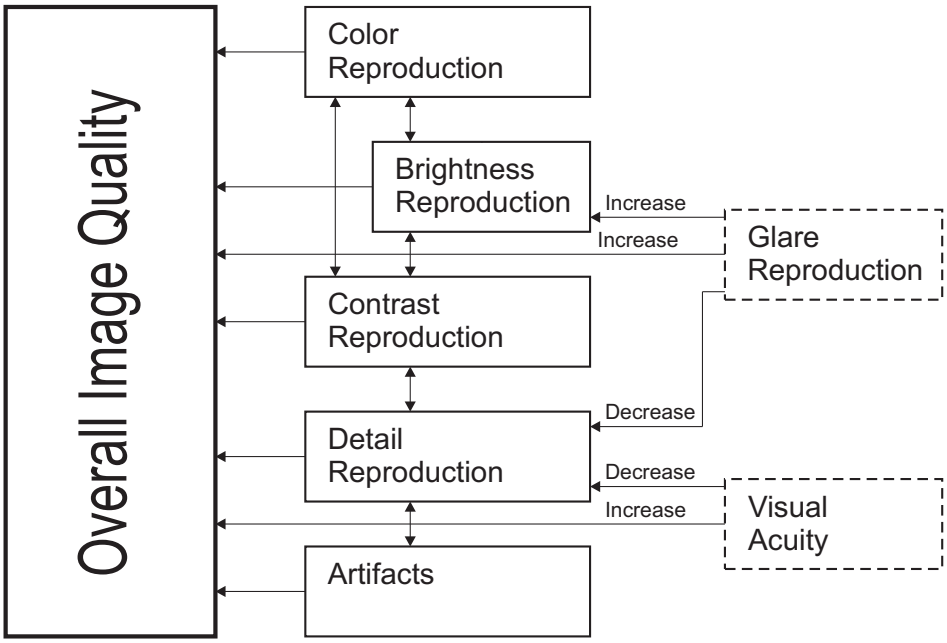


Figure 4.7: The relationships between image attributes. The attributes we did not evaluate in subjective perceptual experiments are in dashed boxes.

Moreover, the simulation of color appearance at scotopic levels of illumination can substantially change the perceived brightness. Finally, the *simulation of glare* plays an important role for the brightness perception. The glare simulation increases the apparent brightness of light sources.

It was shown that **contrast** increases with the *luminance* (Stevens effect, see [Fair05]). Since we can identify the contrast at different spatial resolutions, the perception of contrast is obviously affected by the reproduction of *details*. The experimental results of Calabria and Fairchild [Cala03] confirmed that the perceived contrast depends also on image *lightness, chroma* and *sharpness*.

Colors are related to brightness, because the colorfulness increases with the luminance level (i.e. the Hunt effect [Fair05]).

The reproduction of **details** is strongly affected by the simulation of the *visual acuity*. Since there is available data that represent the visual acuity (e.g. Shaler's curve [Ferw96]), these data place limits on the reproduction of fine details, and may also be utilized to verify the perceptual quality of detail reproduction. Furthermore, the *visibility preservation* diminishes the reproduced details using a threshold function (e.g. the threshold versus intensity curve, TVI). The simulated *glare* can obscure otherwise reproducible details near strong light sources.

Using subjective testing results, Spencer et al. [Spen95] verified that the **simulation of glare** can substantially increase the apparent *brightness* of light sources in digital images.

In the scheme of relationships (Figure 4.7), we can identify attributes that represent **limitations** of the HVS: the simulation of glare, the simulation of visual acuity and (in part) the reproduction of color (in the sense of simulation of the scotopic vision). These attributes

enhance the perceptual quality of the output image, but are not desirable when the goal is different, for example when we aim to reproduce as many details as possible.

4.6 Subjective Perceptual Studies

We have conducted two separate and technically different subjective perceptual studies – 1) a rating-based experiment with reference real-world scenes and 2) a ranking-based experiment with no references, see Figure 4.8. These experiments were conducted to encourage the proposed idea of an *overall image quality* measure and to verify the correlations to and between the image attributes shown in Figure 4.7. Moreover, the execution of two principally different studies gave us the opportunity to relate the obtained subjective results. Finally, we used the results of perceptual studies to evaluate the strengths and weaknesses of 14 TM methods.



Figure 4.8: Example of subjective perceptual experiments setups. Left: rating experiment with real references, Right: ranking experiment without references.

Prior to the main experiments we have conducted a pilot study to examine the setup and to verify that subjects were able to rate "soft-copy" images against the real scenes (i.e. rating experiment verification). During this study we have also fine-tuned the parameters of several TM methods, and we have refined instructions given to subjects. Preliminary ideas of the project as well as the results of our pilot study have been presented in [Cadi05].

It is worth noting that apart from the evaluation of the 14 involved TM methods, the results concerning the relations of image attributes and overall perceptual quality of an image are totally independent of any particular TM method or of the values of its parameters (i.e. the 14 tone mapped images represented a collection of natural input visual stimuli in our subjective perceptual studies). We believe that the collection of images we used is much more natural than the usual artificial stimuli used in vision science for narrow perceptual studies, where

	Min	Max	Mean	Dynamic range
night scene	-2.33	2.77	-0.99	5.13
indoor scene	-1.09	4.27	0.82	5.37
outdoor scene	0.63	6.08	2.69	5.45

Table 4.1: Numerical luminance values ($\log_{10}[cd/m^2]$) for the experimental HDR images.

images are very simple derivations of an original LDR image (thresholding, scaling, chroma variations, or so).

4.6.1 Subjective Testing Setup

We arranged three representative HDR real-world scenes for our experiments: a typical real-world indoor HDR scene, see Table 4.3, a typical HDR outdoor scene, see Table 4.4, and a night urban HDR scene, see Table 4.5. We acquired a series of 15 photos of each scene using a digital camera (Canon EOS300D, Sigma DC 18-200) with varying exposure (fixed aperture f/11, varying shutter speeds) from a locked-down tripod. The focal length was around 50mm (crop factor equivalent) for all scenes – which corresponds to the normal FOV of an observer. The HDR radiance maps were recovered from the recorded series using the method of Debevec and Malik [Debe97]. The dynamic ranges of the resulting HDR images of the indoor scene, outdoor scene and night urban scene were about 10^5:10^{-1}cd/m^2, 10^6:10^1cd/m^2, and 10^3:10^{-3}cd/m^2 respectively (numerical values as reported by the `pfsstat` utility[1] are summarized in Table 4.1), luminance histograms are shown in Figure 4.9.

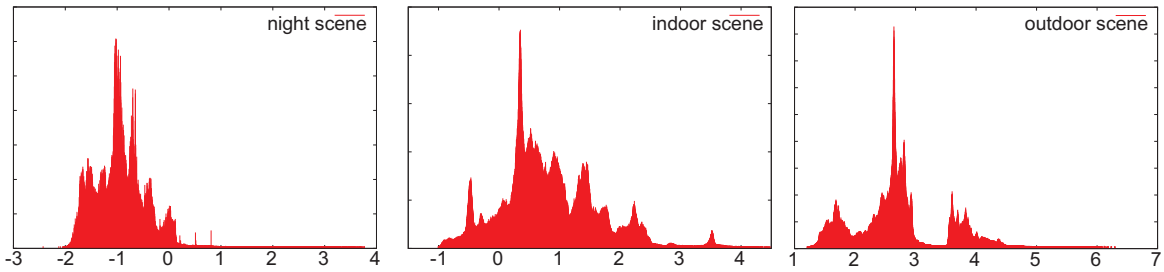


Figure 4.9: Luminance histograms ($\log_{10}[cd/m^2]$) of the experimental HDR images, from left: night scene, indoor scene, outdoor scene.

We transformed these input HDR images using 14 different TM methods, so that we obtained 14 LDR images[2] per scene for investigation. We attempted to include the largest possible amount of methods (see [Devl02, Rein05b] for an overview) into the evaluation, and came up with the 14 techniques (see Table 4.2) to be included into our experiment (abbreviations are used through the entire chapter); for the resulting images see Tables 4.3, 4.4, 4.5. All the evaluated methods were implemented by the author of this thesis with some discussions and help from the original authors of these methods.

[1] Available at `http://www.mpi-inf.mpg.de/resources/pfstools/`

[2] All the tone mapped images as well as the original HDR images are available on the web pages of the project: `http://www.cgg.cvut.cz/~cadikm/tmo`

Abbreviation	Method description	Publication	Global/Local
Ashikhmin02	A Tone Mapping Algorithm for High Contrast Images	[Ashi02]	L
Chiu93	Spatially Nonuniform Scaling Functions for High Contrast Images	[Chiu93]	L
Choudhury03	The Trilateral Filter for High Contrast Images and Meshes	[Chou03]	L
Drago03	Adaptive Logarithmic Mapping For Displaying High Contrast Scenes	[Drag03a]	G
Durand02	Fast Bilateral Filtering for the Display of HDR Images	[Dura02]	L
Fattal02	Gradient Domain High Dynamic Range Compression	[Fatt02]	L
LCIS99	Low Curvature Image Simplifier	[Tumb99a]	L
Pattanaik02	Adaptive Gain Control for HDR Image Display	[Patt02]	L
Reinhard02	Photographic Tone Reproduction for Digital Images	[Rein02]	L
Schlick94	Quantization Techniques for Visualization of HDR Pictures	[Schl94]	L
Tumblin99	Revised Tumblin-Rushmeier Tone Reproduction Operator	[Tumb99b]	G
Ward94	A contrast-based scalefactor for luminance display	[Ward94]	G
Ward97	A Visibility Matching Tone Reproduction Operator for HDR Scenes	[Lars97]	G
Linear Clip	Manual linear clipping		G

Table 4.2: Abbreviations of evaluated tone mapping methods

The sequence of 14 LDR TM images represented the input visual stimuli for each observer, all the testings were performed under controlled ambient luminance level. A total number of 20 subjects aged between 26 and 52 were involved in our experiments. All participating subjects had normal or corrected-to-normal vision and were non-experts in the field of tone mapping. In the two experimental studies, we collected in total $3_{(\text{scenes})} \cdot (10+10)_{(\text{subjects})} \cdot 6_{(\text{attributes})} \cdot 14_{(\text{methods})} = 5040$ values of observation scores.

In the first experiment, based on **rating** (see Figure 4.8 - left), we simultaneously presented an original (**real-world**) **HDR scene** and the appropriate TM images of this scene to human observers. In order to keep the illumination moderately constant, we performed all the testing procedures at the same time of the day as the HDR image was acquired, continually inspecting the illumination conditions using an exposure meter. The TM images were shown separately in random order on a calibrated monitor[3] to a group of 10 subjects. The task of each subject was to express the overall image quality, and the quality of reproduction of basic attributes – overall brightness, overall contrast, reproduction of details, overall reproduction of colors,

[3]FSC P19-2, 19-inch LCD display, with maximum luminance of 280 cd/m^2. We used ICC profiles (D65) for both the monitor and the camera to perform the colorimetric characterization of the devices.

and the lack of disturbing image artifacts for a particular image by *ratings* (on the scale 1–10, where 10 represents the best result, while 1 is the worst) with respect to the actual scene. All subjects were verbally introduced to the experiment and they were instructed to "Rate the images on how close the particular image attribute matches in appearance to the real-world scene" (attribute reproduction results) and to "Rate the images on how close the overall match in appearance is to the real-world scene" (overall image quality results). To avoid any confusion, subjects were personally informed that we were interested in *quality of reproduction* (not the amount or quantity) of inquired image attributes (e.g. "Less detail in the image than in the ground truth is bad, more detail in the image than in the ground truth is bad as well, the closer to the ground truth the better the score should be.") and that they should judge only the particular attribute and avoid any influence of other attributes. Subjects sat at the place of the camera at common viewing distance from the display (approximately 60cm) and they were able to directly observe both the real scene and the display. However, subjects were always instructed to take few seconds to adapt to each. The procedure took approximately 45 minutes for one observer and one scene. We chose the rating scale method in this experiment to stimulate observers to do the direct comparison of the TM image to the real scene.

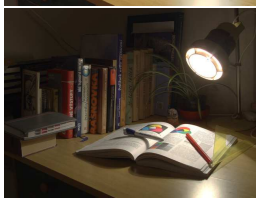
In the second experiment, based on **ranking** (see Figure 4.8 - right), we investigated what happens when subjects have no possibility of directly comparing to the ground truth (or are not affected by a previous experience with the real scene). A group of 10 observers (different ones than in the first experiment), who have **never seen the real HDR scenes** and had therefore virtually no idea about the attributes of original scenes, was selected. The task of each subject was to order (*rank*) *image printouts* resulting from the 14 methods according to the overall image quality, and the quality of reproduction of overall contrast, overall brightness, colors, details, and image artifacts. Similarly to the first experiment, all subjects were verbally introduced to the experiment and they were instructed to "Rank the printouts on how close the particular image attribute matches in appearance to a *hypothetical* real-world scene," the idea being that when a human views an image, she always forms a mental model of the original scene. Thus, description of image attributes was the same as in the first experiment, but observers were instructed to "Imagine how the original real-world scene would look like" and rank the printouts accordingly. The procedure took approximately 35 minutes for one observer and one series of input images. The investigated printouts were high-quality color image printouts on a glossy paper of the same 14 tone mapped images as in the first experiment[4]. Printouts were observed in an office under standard illumination of approximately 550lux.

4.7 Results and Discussion

In order to make the results of two conducted experiments comparable, we converted the *rating* observation scores to the ranking scale by computing the ranks of observations for each person and attribute with adjustment for ties (if any values were tied, we computed their average rank) prior to the following evaluations. For example, a rating observation vector $\mathbb{X}$ is converted to the rank vector $\mathbb{X}'$ as follows:

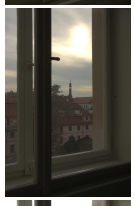
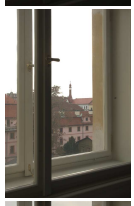
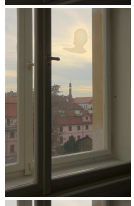
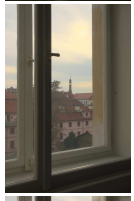
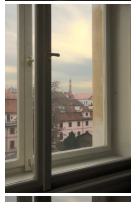
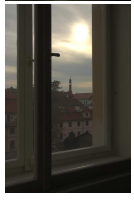
$$\mathbb{X} = (\ 3\ \ 7\ \ 2\ \ 6\ \ 2\ \ 1\ \ 5\ \ 6\ \ 9\ \ 5\ \ 6\ \ 8\ \ 8\ \ 4\ )$$
$$\mathbb{X}' = (\ 4\ \ 11\ \ 2.5\ \ 9\ \ 2.5\ \ 1\ \ 6.5\ \ 9\ \ 14\ \ 6.5\ \ 9\ \ 12.5\ \ 12.5\ \ 5\ )$$

[4] A HP Color Laserjet 3500 was used with the ICC profile to perform colorimetric characterization, in order to achieve a reasonably comparable color representation as in the first experiment.

Method	Image	Brightness	Contrast	Details	Colors	Overall Quality
Linear Clip		**10.6** *2.8*	**7.6** *3.9*	**7.6** *4.7*	**11.3** *3.6*	**8.9** *3.0*
		6.3 *3.1*	**6.6** *3.9*	**3.6** *3.7*	**8.4** *3.7*	**5.7** *2.9*
Ward94		**7.7** *3.0*	**8.1** *4.0*	**5.3** *1.9*	**9.6** *2.4*	**9.7** *3.7*
		7.1 *2.7*	**9.7** *2.2*	**6.8** *2.5*	**9.8** *2.5*	**7.9** *3.4*
Tumblin99		**11.1** *1.6*	**9.5** *3.3*	**7.5** *3.9*	**10.3** *1.9*	**10.8** *3.1*
		9.0 *1.8*	**9.6** *3.5*	**6.9** *1.9*	**10.4** *1.9*	**8.0** *2.5*
Reinhard02		**10.8** *1.9*	**11.6** *2.7*	**10.4** *2.8*	**12.5** *1.4*	**12.2** *1.1*
		11.9 *1.9*	**12.1** *2.6*	**11.8** *1.7*	**12.6** *1.8*	**11.9** *2.3*
Schlick94		**3.8** *2.4*	**7.1** *3.3*	**6.2** *3.8*	**5.6** *2.9*	**9.3** *3.1*
		6.9 *4.2*	**8.7** *4.3*	**6.7** *3.9*	**9.1** *3.9*	**8.2** *4.6*
Ward97		**8.8** *2.5*	**9.8** *3.3*	**8.1** *3.2*	**10.3** *2.3*	**11.5** *1.8*
		10.4 *2.3*	**9.7** *2.1*	**9.4** *1.9*	**10.0** *1.9*	**10.8** *2.2*
Durand02		**8.4** *3.7*	**4.7** *4.4*	**6.9** *4.0*	**4.6** *2.9*	**3.5** *2.7*
		2.9 *1.6*	**2.2** *0.8*	**5.0** *0.9*	**2.4** *0.9*	**2.75** *2.4*

Method	Image	Brightness	Contrast	Colors	Details	Overall Quality
LCIS99		**4.1** *1.5*	**6.2** *2.6*	**5.4** *3.9*	**3.4** *1.2*	**4.6** *1.3*
		11.2 *2.1*	**10.8** *1.9*	**9.8** *3.3*	**9.3** *3.3*	**11.9** *2.4*
Pattanaik02		**11.1** *3.6*	**8.9** *3.6*	**12.4** *2.5*	**8.6** *3.3*	**6.8** *3.3*
		8.2 *4.5*	**8.2** *3.9*	**9.4** *3.5*	**8.1** *3.5*	**7.6** *3.2*
Choudhury03		**5.2** *1.5*	**5.9** *2.3*	**7.0** *2.6*	**5.4** *1.4*	**2.2** *3.6*
		10.2 *2.2*	**8.8** *2.5*	**9.6** *2.3*	**7.7** *2.3*	**10.4** *2.0*
Drago03		**10.9** *1.5*	**9.5** *1.8*	**6.9** *3.9*	**9.0** *3.0*	**8.9** *1.5*
		7.5 *2.2*	**5.8** *1.9*	**5.1** *2.3*	**6.5** *2.3*	**7.2** *1.7*
Ashikhmin02		**8.3** *2.5*	**8.0** *3.7*	**10.2** *2.6*	**8.3** *3.2*	**7.6** *3.3*
		7.3 *2.9*	**6.5** *2.0*	**9.6** *2.6*	**5.0** *2.6*	**7.5** *1.3*
Fattal02		**3.2** *1.0*	**5.4** *3.6*	**7.4** *4.2*	**5.0** *1.8*	**5.8** *2.4*
		3.7 *2.7*	**3.8** *2.6*	**8.2** *2.7*	**2.4** *2.7*	**3.4** *3.4*
Chiu93		**1.1** *0.3*	**2.7** *3.0*	**3.0** *2.5*	**1.1** *0.3*	**1.8** *1.3*
		2.5 *1.9*	**2.7** *2.6*	**3.3** *1.6*	**3.5** *1.6*	**1.9** *0.9*

Table 4.3: Strengths and weaknesses of evaluated TM methods – **indoor scene**. In bold: average ranking scores (1st line) and average rating scores (3rd line); in italics: standard deviations for ranking (2nd line) and for rating scores (4th line). The higher values represent the higher reproduction quality.

Method	Image	Brightness	Contrast	Details	Colors	Overall Quality
Linear Clip		**12.3** *2.2* **10.3** *3.1*	**13.2** *0.9* **10.7** *2.9*	**12.5** *1.9* **9.4** *3.6*	**13.4** *0.5* **10.1** *3.6*	**13.2** *0.9* **11.8** *2.5*
Ward94		**4.3** *2.0* **7.4** *3.1*	**6.2** *2.6* **10.2** *2.3*	**4.1** *3.1* **7.2** *2.3*	**5.7** *1.4* **9.1** *2.3*	**8.2** *1.8* **9.7** *2.2*
Tumblin99		**12.4** *2.3* **10.3** *2.5*	**12.7** *1.2* **8.4** *2.1*	**12.5** *1.9* **7.3** *4.1*	**13.6** *0.5* **9.8** *3.4*	**12.9** *1.0* **11.2** *1.5*
Reinhard02		**9.2** *2.0* **13.1** *1.1*	**10.6** *0.9* **12.0** *1.7*	**9.0** *1.6* **9.7** *4.6*	**10.1** *1.1* **11.7** *1.9*	**7.9** *3.2* **9.7** *2.9*
Schlick94		**9.4** *2.7* **10.6** *2.0*	**10.6** *1.6* **10.3** *2.6*	**10.7** *2.6* **9.7** *2.7*	**10.9** *0.5* **9.8** *2.2*	**11.5** *1.4* **11.1** *2.3*
Ward97		**10.6** *2.1* **9.7** *2.9*	**11.6** *2.3* **9.0** *3.7*	**12.3** *1.1* **8.9** *3.4*	**11.4** *0.8* **8.7** *2.5*	**11.3** *1.8* **8.9** *1.5*
Durand02		**3.6** *2.0* **7.8** *3.4*	**6.0** *2.9* **10.4** *2.5*	**4.8** *2.6* **7.0** *2.0*	**5.6** *1.3* **9.0** *2.4*	**7.7** *1.6* **10.1** *2.7*

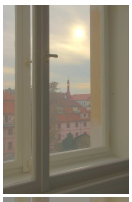
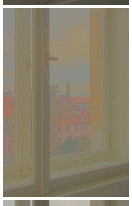
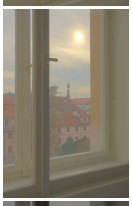
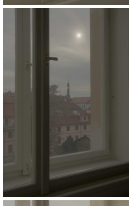
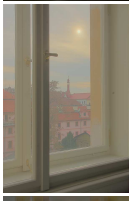
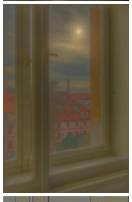
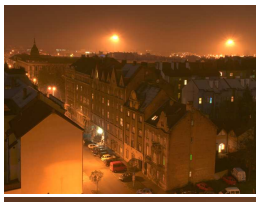
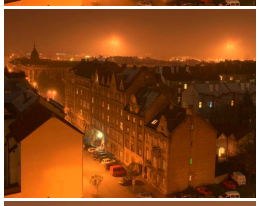
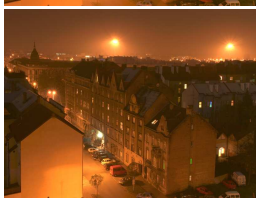
Method	Image	Brightness	Contrast	Colors	Details	Overall Quality
LCIS99		**9.4** *2.2* **7.8** *3.3*	**7.9** *1.6* **7.8** *3.2*	**8.8** *1.5* **8.7** *2.6*	**7.9** *2.0* **8.5** *2.6*	**7.8** *1.7* **7.7** *3.0*
Pattanaik02		**6.1** *4.1* **2.3** *0.8*	**4.1** *1.8* **2.1** *0.8*	**3.4** *2.2* **4.8** *3.4*	**2.1** *0.3* **3.1** *2.1*	**2.1** *0.3* **2.3** *0.7*
Choudhury03		**8.2** *1.7* **6.0** *2.4*	**5.8** *1.3* **6.1** *3.2*	**7.9** *1.4* **7.8** *3.1*	**7.3** *1.7* **7.1** *2.4*	**6.0** *1.5* **7.7** *3.0*
Drago03		**3.6** *2.1* **6.2** *2.3*	**4.1** *2.4* **6.8** *2.3*	**5.1** *2.2* **6.1** *1.7*	**5.4** *2.5* **5.0** *3.6*	**6.9** *2.7* **5.7** *1.5*
Ashikhmin02		**8.7** *2.5* **6.6** *2.4*	**6.9** *1.8* **6.0** *2.1*	**7.7** *2.2* **7.9** *4.1*	**6.8** *1.5* **7.1** *2.3*	**4.9** *1.3* **5.2** *2.5*
Fattal02		**2.8** *1.3* **4.4** *4.1*	**1.8** *1.1* **2.9** *0.9*	**3.1** *1.4* **5.5** *4.7*	**3.7** *1.9* **2.0** *0.7*	**3.5** *1.2* **2.3** *0.9*
Chiu93		**4.4** *3.8* **2.9** *2.2*	**3.5** *3.0* **2.6** *2.6*	**3.1** *2.3* **5.3** *5.1*	**1.1** *0.3* **4.4** *3.8*	**0.3** *1.1* **1.8** *0.5*

Table 4.4: Strengths and weaknesses of evaluated TM methods – **outdoor scene**. In bold: average ranking scores (1st line) and average rating scores (3rd line); in italics: standard deviations for ranking (2nd line) and for rating scores (4th line). The higher values represent the higher reproduction quality.

Method	Image	Brightness	Contrast	Details	Colors	Overall Quality
Linear Clip		**11.3** *3.6*	**12.8** *1.2*	**13.2** *1.3*	**12.2** *2.7*	**12.9** *1.0*
		10.0 *3.0*	**9.8** *2.1*	**8.8** *3.3*	**10.4** *2.9*	**10.2** *3.1*
Ward94		**10.6** *3.0*	**12.1** *1.8*	**12.1** *1.5*	**11.9** *2.2*	**12.5** *1.4*
		10.2 *3.2*	**11.1** *3.2*	**9.6** *3.9*	**11.5** *3.0*	**11.9** *1.7*
Tumblin99		**7.4** *2.3*	**7.6** *1.7*	**8.1** *1.0*	**8.4** *1.5*	**8.8** *1.3*
		8.8 *2.7*	**8.6** *3.1*	**6.8** *2.9*	**8.4** *3.4*	**9.4** *2.7*
Reinhard02		**9.1** *3.6*	**9.0** *2.5*	**9.4** *1.3*	**9.7** *1.4*	**9.3** *1.3*
		6.8 *4.2*	**8.5** *2.4*	**9.8** *3.7*	**10.0** *2.3*	**8.7** *2.6*
Schlick94		**8.8** *2.6*	**9.3** *2.1*	**9.7** *1.3*	**9.5** *1.8*	**10.4** *1.7*
		7.5 *3.1*	**10.9** *1.8*	**9.5** *2.5*	**9.3** *1.9*	**8.8** *1.9*
Ward97		**8.7** *2.8*	**7.0** *1.8*	**7.8** *2.0*	**8.1** *1.4*	**7.7** *1.3*
		3.7 *2.8*	**6.3** *2.0*	**6.5** *2.0*	**6.9** *2.0*	**5.9** *1.9*
Durand02		**11.4** *2.5*	**12.5** *1.4*	**12.8** *0.4*	**11.7** *2.2*	**13.0** *0.6*
		8.9 *3.7*	**10.9** *2.5*	**8.9** *2.9*	**10.9** *2.2*	**11.0** *2.4*

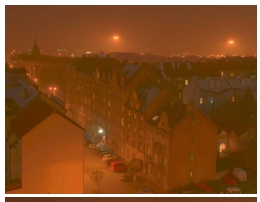
Method	Image	Brightness	Contrast	Colors	Details	Overall Quality
LCIS99		**6.5** *2.3*	**6.2** *1.5*	**5.8** *0.9*	**6.1** *2.7*	**5.7** *0.5*
		7.6 *2.2*	**6.9** *2.9*	**7.5** *2.1*	**6.5** *2.0*	**7.7** *2.6*
Pattanaik02		**9.1** *2.0*	**10.0** *1.0*	**10.5** *1.8*	**9.6** *2.5*	**9.2** *1.6*
		10.4 *3.9*	**12.2** *1.9*	**12.7** *1.3*	**11.7** *2.0*	**12.5** *1.8*
Choudhury03		**7.1** *2.6*	**6.6** *2.3*	**5.3** *0.9*	**5.9** *2.6*	**5.1** *1.0*
		6.7 *1.9*	**5.6** *3.4*	**7.3** *2.9*	**6.5** *2.8*	**6.9** *2.5*
Drago03		**4.9** *4.2*	**4.9** *3.9*	**3.3** *0.6*	**3.8** *2.6*	**3.4** *1.2*
		6.9 *3.9*	**4.9** *2.5*	**5.2** *3.0*	**3.9** *3.2*	**4.4** *1.9*
Ashikhmin02		**5.1** *3.4*	**3.6** *1.1*	**3.6** *1.0*	**4.5** *2.7*	**3.3** *1.0*
		8.2 *3.2*	**4.8** *2.2*	**8.5** *4.4*	**5.2** *2.6*	**4.8** *2.3*
Fattal02		**4.0** *2.5*	**2.4** *0.5*	**2.4** *0.7*	**2.6** *0.5*	**2.7** *0.6*
		5.7 *4.8*	**3.7** *3.2*	**3.0** *3.0*	**2.6** *1.1*	**2.1** *0.4*
Chiu93		**1.0** *0.0*	**1.0** *0.0*	**1.0** *0.0*	**1.0** *0.0*	**1.0** *0.0*
		3.9 *5.0*	**1.1** *0.2*	**1.2** *0.3*	**1.3** *0.6*	**1.1** *0.2*

Table 4.5: Strengths and weaknesses of evaluated TM methods – **night scene**. In bold: average ranking scores (1st line) and average rating scores (3rd line); in italics: standard deviations for ranking (2nd line) and for rating scores (4th line). The higher values represent the higher reproduction quality.

	Q_{rating}	Q_{ranking}
overall quality	2.7984	4.5823
brightness	2.2857	2.7648
contrast	1.2857	0.3984
details	0.1429	3.8353
colors	1.2857	3.6545
artifacts	0.1231	1.3740
critical value $Q_{\text{crit}} = 5.99$		

Table 4.6: Results of two separate Friedman's tests for the effect of input scenes

We analyzed the data using non-parametric statistical tests[5]. Moreover, we also converted these rank order data using the Thurstonian model (condition D) [Thur27, Torg58] to interval scales. Tables 4.3, 4.4, 4.5 show the numerical results separately for each scene, while interval scales are shown along with standard errors in Figure 4.10 (overall average results), in Figure 4.12 (average values for each experiment), and Figure 4.14 (overall image quality ratings for each input scene for each experiment). We describe and discuss obtained results in the following text: Sections 4.7.1 and 4.7.2 statistically prove that neither the experimental setup nor the choice of scenes has a systematic influence on the results. In Section 4.7.3 we discuss the results of examined TM methods. In Section 4.7.5 we quantify the relationship between image attributes proposed in Section 4.5. Finally, in Section 4.7.5 we compare our results to results obtained in previous work.

4.7.1 Effects of Input Scenes and Methods

First, we have to inquire if the *input scene* has a significant systematic effect on the evaluation of the methods and image attributes. We use the Friedman's nonparametric two-way analysis of variance (ANOVA) test [Sieg88] (see also Appendix A.1) for each image attribute independently for ranking and rating datasets. We state the null hypothesis H_0 as follows: there is no significant difference between observation values for the input scenes.

We summarize the results for all image attributes in Table 4.6. If the value of Friedman's statistics Q is higher than the tabulated critical value Q_{crit}, we reject the null hypothesis H_0. For all the cases we use a significance level of $p < 0.05$. As we can observe in the Table 4.6, we cannot reject the null hypothesis for any of the attributes for both experimental setups. This means we were not able to find a statistically significant difference between the three input scenes and we can thus proceed with the evaluation independently of the input scenes.

Next, we have to verify that there are significant differences between the *TM methods* and the evaluation of TM methods thus makes sense. We use Friedman's analysis independently for ranking and rating, with the null hypothesis H_0: there is no significant difference between observation values for 14 evaluated methods.

The results are summarized in Table 4.7. Since all obtained Q values are much higher than Q_{crit}, we reject the null hypothesis for all attributes. This means we found significant differ-

[5]Since we have non-normally distributed observation values (rank orders), we use nonparametric tests throughout this chapter.

	Q_{rating}	Q_{ranking}
overall quality	85.093	110.98
brightness	72.772	83.494
contrast	87.782	92.531
details	56.826	89.617
colors	91.939	111.91
artifacts	75.833	92.768
critical value $Q_{\text{crit}} = 19.16$		

Table 4.7: Results of two separate Friedman's tests for the effect of input methods

ences between the method scores for all attributes and both experiments and we can proceed with the evaluation of TM methods.

4.7.2 Effect of the Experimental Setup

The next question is if there is a statistically significant difference between the data obtained from the two different *experimental setups* (two conducted psychophysical experiments). Recall that in the rating experiment, observers were able to directly rate the quality of image attributes against the real reference (real HDR scene), while in the ranking experiment they had to rank the images according to the quality of image attributes without knowledge of the original scene, see Figure 4.8. The second experiment, even though without reference, was not a simple preference experiment, since observers were instructed to rank images according their mental model of the original real-world scene. We chose two different evaluation methods because unlike in the second experiment, in the first experiment we did not want to show all the 14 images simultaneously with the reference scene. We rather wanted to stimulate the observer to rate a single image against the real reference, thus slightly eliminating the ranking of tested images (this is however never fully possible). The rating scale was chosen so that the scores were in the interval $[1, 10]$.

To examine the differences between the rating and ranking experiments[6] for each attribute we used the Kruskal-Wallis test [Sieg88] (nonparametric version of one-way ANOVA, see Appendix A.1). The critical value for the test ($\nu = 14 \cdot 10 \cdot 3 - 1 = 419$ degrees of freedom) is $\chi^2_{\text{crit}} = 467.73$. All the obtained results of the test were much smaller than the critical value, therefore we did not detect any significant difference between experiments for any attributes using the nonparametric ANOVA.

Since using the Kruskal-Wallis test we did not find any statistically significant differences between the rating and ranking experiments, we also applied another more rigorous test, the profile analysis [Lehm78, Renc02], to the observed data. Profile analysis is a nonparametric test used to verify that changes in a particular stochastic variable have the same tendency for several different objects (rating and ranking experiments in our case). We state the null hypothesis H_0 as follows: the mean values of observation vectors $\mathbb{X}_{\text{rat}_i}$ and $\mathbb{X}_{\text{ran}_i}$, where $\mathbb{X}_{\text{rat}_i}$ and $\mathbb{X}_{\text{ran}_i}$ is a vector of observed values from the rating and ranking experiment respectively,

[6] Recall that the rating is converted to ranking by computing the ranks of observations for each person and attribute with adjustment for ties.

	V^*_{INDOOR}	V^*_{OUTDOOR}	V^*_{NIGHT}
overall quality	-0.2016	0.0000	0.0000
brightness	0.7928	0.0000	1.0296
contrast	0.8021	0.3077	0.1156
details	-0.1077	-0.1287	0.4361
colors	0.4008	-0.7951	-0.1773
artifacts	0.0000	0.2068	0.0000
critical value $V^*_{\text{crit}} = 2.394$			

Table 4.8: Results of Profile analysis

Source of Variation	SS	df	MS	F	p
experimental setup	-0	1	-0	-0	≈ 1
input scene	0	2	0	0	≈ 1
TM method	5936.1	13	456.62	49.96	≈ 0
subject (observer)	0	9	0	0	≈ 1
Residual	7439.4	814	9.13		
Total	13376	839			

Table 4.9: Results of non-parametric MANOVA test (where SS denotes Sum of Squares, df means Degrees of Freedom, MS denotes Mean Square, F is F value, and p is p-value for the null hypothesis, refer to Appendix A.1 for reference).

differ just in shift (we say they have parallel profiles). According to the profile analysis process, we compute the test quantity V_t^* for each variable t and we reject H_0 if V_t^* is higher than the computed critical value V^*_{crit}.

First, we calculated H_0 for the ranking and rating results for the profiles over the scenes for each image attribute separately. The observation vectors were then: $\mathbb{A}_{\text{rat}_i} = (A_{\text{INDOORrat}_i}, A_{\text{OUTDOORrat}_i}, A_{\text{NIGHTrat}_i})$ and $\mathbb{A}_{\text{ran}_i} = (A_{\text{INDOORran}_i}, A_{\text{OUTDOORran}_i}, A_{\text{NIGHTran}_i})$, where $\mathbb{A}$ denotes particular image attribute, and $A_{\text{DESK}*_i}$, $A_{\text{WINDOW}*_i}$, and $A_{\text{NIGHT}*_i}$ are the observation values for the Desk, Window and Night scene respectively. The obtained profile analysis results are summarized in Table 4.8. These results show that we cannot reject H_0 for any attribute, this means we did not find a significant differences in profiles for each input scene for the rating and ranking experiments.

Next, we averaged the scores for the input scenes for each attribute for each experimental setup separately and we performed another profile analysis over the following vectors: $\mathbb{X}_{\text{rat}_i} = (OIQ_{\text{rat}_i}, Bri_{\text{rat}_i}, Con_{\text{rat}_i}, Det_{\text{rat}_i}, Col_{\text{rat}_i}, Art_{\text{rat}_i})$ and $\mathbb{X}_{\text{ran}_i} = (OIQ_{\text{ran}_i}, Bri_{\text{ran}_i}, Con_{\text{ran}_i}, Det_{\text{ran}_i}, Col_{\text{ran}_i}, Art_{\text{ran}_i})$, where OIQ_{*_i}, Bri_{*_i}, etc., are averages over input scenes for image attributes overall image quality, brightness, etc., for rating and ranking experiments. The critical value is in this case $V^*_{\text{crit}} = 2.6383$ and the resulting values are $V^*_{\text{OIQ}} = -0.3489$, $V^*_{\text{Bri}} = -0.4791$, $V^*_{\text{Con}} = 0.0565$, $V^*_{\text{Det}} = -0.1409$, $V^*_{\text{Col}} = -0.0404$, $V^*_{\text{Art}} = 0.1727$. Since the V^*_{crit} is higher than the resulting V^* for all image attributes, profile analysis did not find a significant difference in the rating and ranking observation data.

Finally, to account for all the factors (i.e. 'subject (observer)', 'TM method', 'input scene' and 'experimental setup') together in one statistical test, we utilized the recently published permutational multi-factorial MANOVA [Ande03]. This test is a non-parametric analogy of the parametric multi-factorial multivariate ANOVA [Taba07]. Results of permutational MANOVA (summarized in Tab. 4.9) show that the factors 'subject', 'input scene' and 'experimental setup' are statistically not significant, i.e. scenes, subjects and types of experiment do not have a significant effect on the resulting scores. The only significant *main effect* is with the factor 'TM method', which means that there are significant differences in responses of subjects depending on the type of the TM method. This correlates with the results reported above, and again justifies our experimental setup. Moreover, we also inquired *interaction effects* and found a significant effect of 'input scene' × 'TM method' ($F = 11.23$, $p < 0.001$) which means that the scores depend on the combination of scene and input method, i.e. there probably exist methods whose performance differs for particular input scenes.

In this section, we made a lot of effort to find a statistically significant difference between the two experiments, but we have not found one. This is a very interesting and important result, because it suggests that for a *perceptual* comparison of TM methods it is sufficient to use ranking without a reference experimental setup. This type of psychophysical testing is much cheaper in terms of money and time than the setup with original scene and ratings.

4.7.3 Evaluation of HDR Tone Mapping Methods

We should stress here again that all our evaluations are targeted at the perceptual dimension of TM, i.e. the holy grail is to reproduce the visual sensation of the real HDR scene as closely as possible (as opposed to, for example, information preservation). Moreover, since all the evaluated methods were implemented personally by the author of the thesis, the results in this section represent also the "achievability" of the results. We do not claim that slightly better results for a particular method could not be achieved after a thorough artificial parameter tuning. We have tested three different HDR scenes with a variety of characteristics, but other input scenes may potentially lead to slightly different results. We should also stress that our evaluation does not reflect computation time, implementation difficulties and other factors, that are also significant in practical applications of TM methods.

The observed values represent the *quality* of reproduction of a particular image attribute, and not its *amount*. For example the average observation values for the reproduction of details show the quality of reproduction of details, not the amount of details. Subjects were instructed to rank/rate the images accordingly, therefore too many or too few details are both rated worse than the right amount of details.

4.7.3.1 Overall Results

The overall results (see interval scores shown in Figure 4.10) suggest that the *best overall quality* is generally observed in images produced by global TM methods (TM curves). Interestingly, the average best score is achieved by the simplest possible approach, the manual *linear clipping* of luminance values! However, this is not such a surprising result, because also our previous pilot studies [Cadi05] have shown the superiority of global approaches in the perceptual dimension of TM. A possible explanation of this is also suggested by our analy-

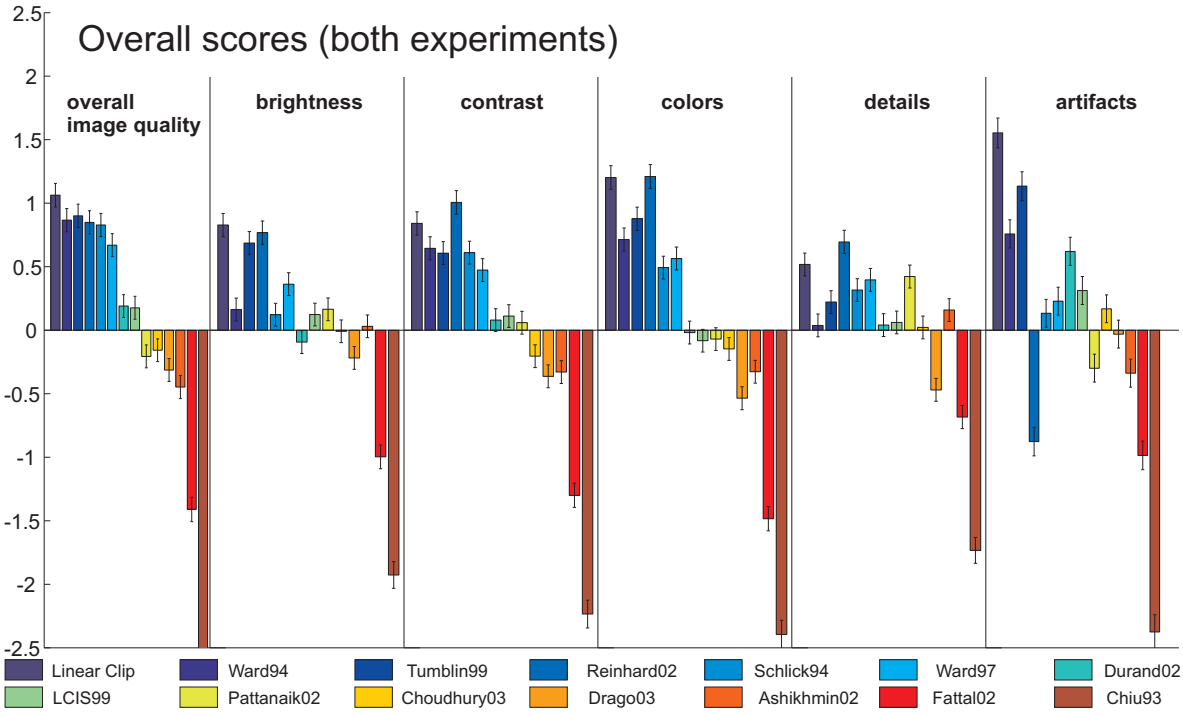


Figure 4.10: Overall accuracy scores for all examined TM methods. Left to right: overall perceptual quality, reproduction of brightness, reproduction of contrast, reproduction of details, reproduction of colors, lack of disturbing artifacts. In each chart the higher value represents the higher reproduction quality.

sis (see Section 4.7.4): the proper reproduction of *overall* image attributes (overall contrast, overall brightness, colors) is essential for the natural perception of the resulting image, more so than *local* attributes. The HVS is evidently highly sensitive to any disruptive factors in the overall image attributes, far more than to the absence of some image details. Recall that the group of six best-rated TM methods contains just one local approach – the method Reinhard02 [Rein02], but an essential part of that method is basically a global TM method with advanced parameter estimation.

The worst rated methods were Fattal02 – the gradient-based approach, which we believe is a good method, but not so for perceptual applications, and an early local approach Chiu93. At the bounds of the quality interval, the best and the worst methods exhibit also the lowest variance, while the middle zone with often uncertain judgments has higher variances. The observers have typically the same opinion about the best/worst question, but difficulties with the evaluation of some similar cases.

The plot of means of the overall image quality attribute (obtained by a non-parametric MANOVA test [Ande03]) with 95% confidence intervals shows the categorization of TM methods more clearly (see Figure 4.11 (left)). As we may observe, there are no statistically significant differences in the overall image quality for the first six methods, which are largely the global tone mapping methods (visualized in blue). The second group (black color) comprises in fact deeply local TM approaches that operate averagely in the perceptual dimen-

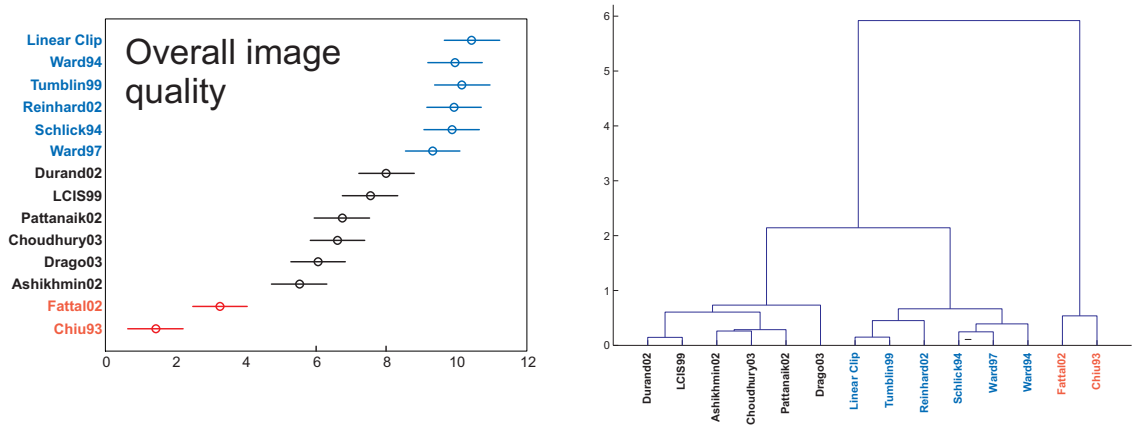


Figure 4.11: Left: average overall image quality with confidence intervals. Circles show OIQ means with 95% confidence intervals (horizontal axis) – the higher value the better quality. Right: average Mahalanobis distances of overall image quality for all methods.

sion of HDR tone reproduction. Finally, in the third group (red color) are perceptually not satisfactory methods. In Figure 4.11 (right) we show the dendrogram of distances of overall image quality between the enquired methods. This graph also shows the described clustering of the methods into three groups.

The evaluation of artifacts (the higher value the better quality, i.e. the less amount of artifacts) shows another interesting result. The approach by Reinhard et al. shows high variance in this attribute, because it produced two relatively good images, but one with very disturbing artifacts, see Table 4.4. Due to the nature of Reinhard's method, the artifacts could not be completely avoided.

4.7.3.2 Comparison of the Two Experiments

In Figure 4.12 we show average results for the two performed experiments separately. These results indicate how well the methods performed in rating (with reference) and ranking (without a reference) experiments. Similarly to overall results, methods Chiu93 [Chiu93] and Fattal02 [Fatt02] performed constantly worst in both experiments. In the rating experiment, Reinard02 [Rein02] exhibits the best scores in all attributes but the artifacts, where it is the third worst rated (alike in the ranking experiment). In the ranking experiment, the linear clipping exhibits constantly the best scores in all attributes.

Generally, the results exhibit similar trends for all the enquired attributes as suggested by statistical analysis in previous sections. The relations of two experiments for each image attribute are visualized in Figure 4.13 along with linear fit and coefficients of determination R^2 (R^2 is a measure of the global fit of the model; $R^2 = 1$ would indicate that the fitted model explained all variability, while $R^2 = 0$ indicates no linear relationship between the results of our two experiments.) The highest agreement between two experiments is for overall contrast, overall image quality, and for the lack of artifacts attribute. The lowest agreement exhibits the detail attribute and we deal with this result in the next section.

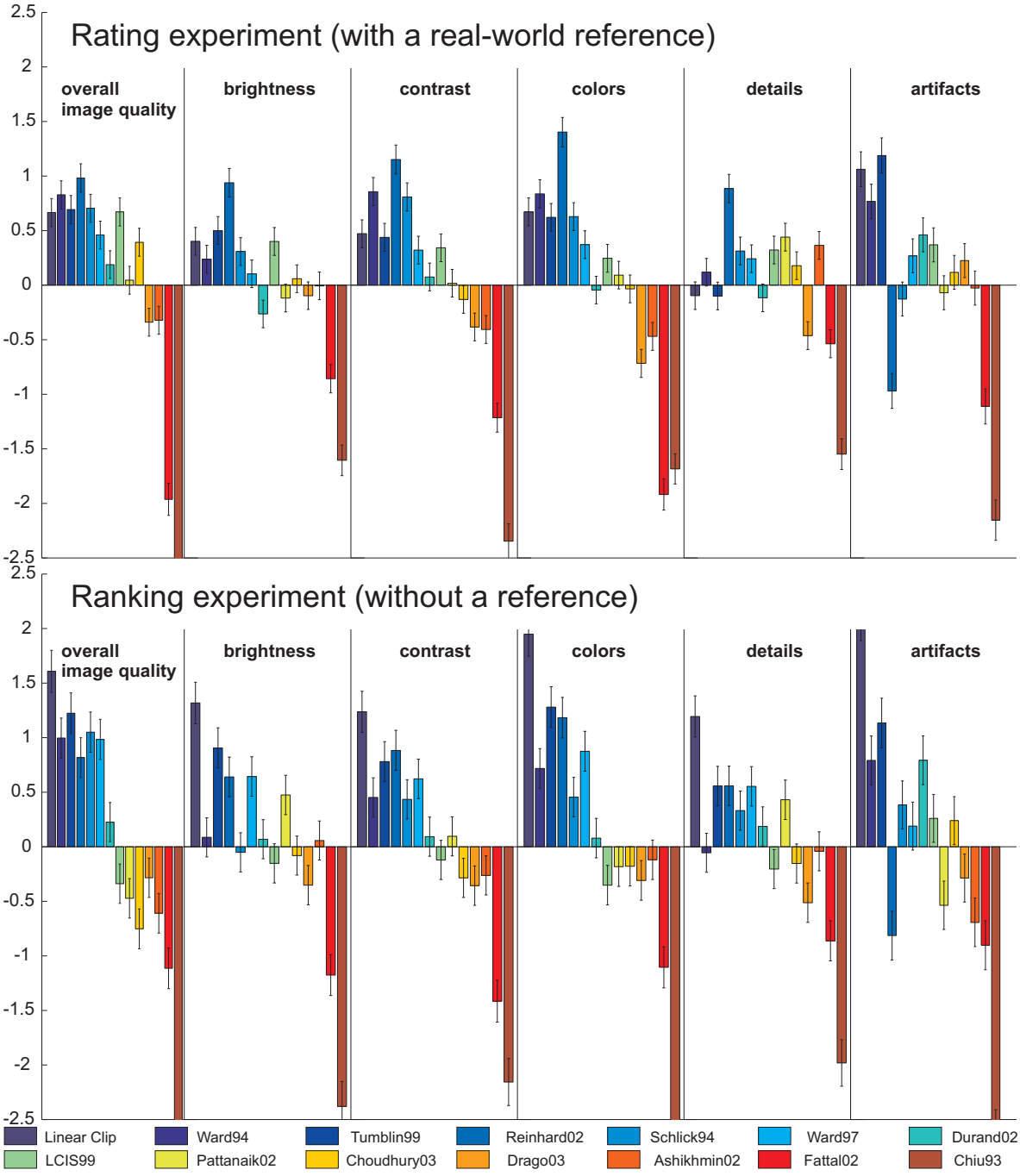


Figure 4.12: Accuracy scores for rating (with a reference) experiment (top) and for ranking (without a reference) experiment (bottom) for all examined TM methods. Left to right: overall perceptual quality, reproduction of brightness, reproduction of contrast, reproduction of details, reproduction of colors, lack of disturbing artifacts. In each chart the higher value represents the higher reproduction quality.

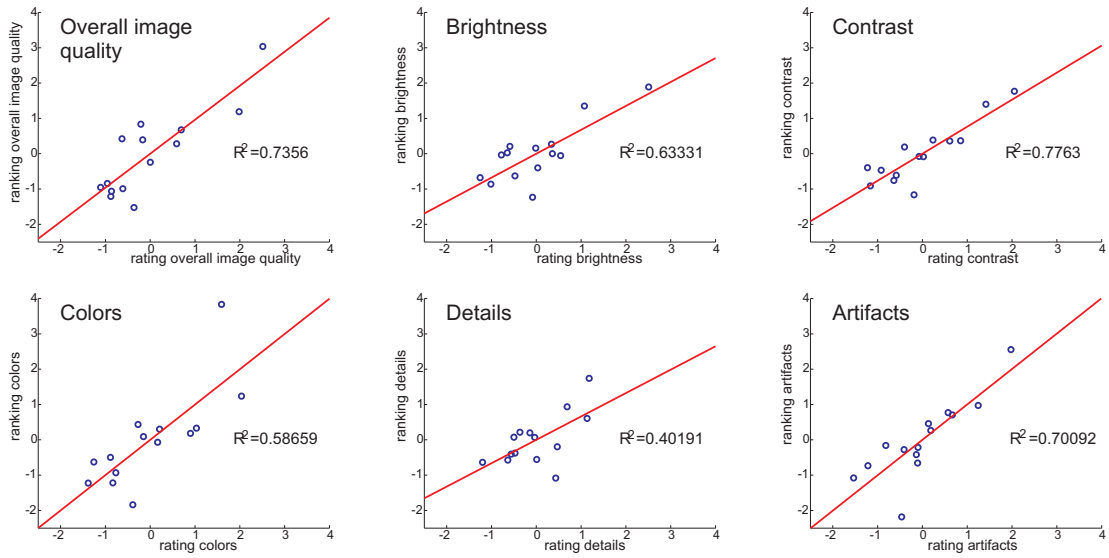


Figure 4.13: Relations of the ranking experiment (vertical axes) and rating experiment (horizontal axes) interval scale results for all image attributes.

4.7.3.3 Comparison of the Results for Input Scenes

Statistical analysis as reported in Section 4.7.2 suggests that even though our input scenes do not have a systematic effect on obtained results, there probably exist methods whose performance differs for particular scenes. To examine the effect of the input scenes on the results, we show the overall image quality scores separately for each scene, see Figure 4.14. We notice rather similar trends in results for the two outdoor scenes (outdoor and night scene), while the indoor scene exhibits slightly different pattern. Since there is a book with tiny writing which dominates the indoor scene, perhaps, there is a higher stress on reproduction of details in this case.

Notice that methods visualized in shades of blue color perform very well for at least two scenes. Chiu93 and Fattal02 on the other hand perform constantly poorly over all scenes in both tests. Pattanaik02 shows interesting consistent behavior – it performs very well for the night scene, averagely for the indoor scene, and poorly for outdoor scene. In case of the indoor scene, LCIS99 and Choudhury03 show the highest discrepancy between rating and ranking experiments. In this case, subjects in the rating experiment perhaps put more stress to the detail attribute in the detriment of other attributes while subjects in the ranking experiment not that much. This is in accordance with results reported in Section 4.7.3.2, where the detail attribute exhibited the lowest agreement.

4.7.4 Overall Image Quality and Relationships of Attributes

Beyond the discussed results, we analyzed the *dependencies of overall image quality* on the quality of reproduction of the five evaluated perceptual image attributes. Our investigations are formulated by means of the experimental results in five-dimensional functions, namely as

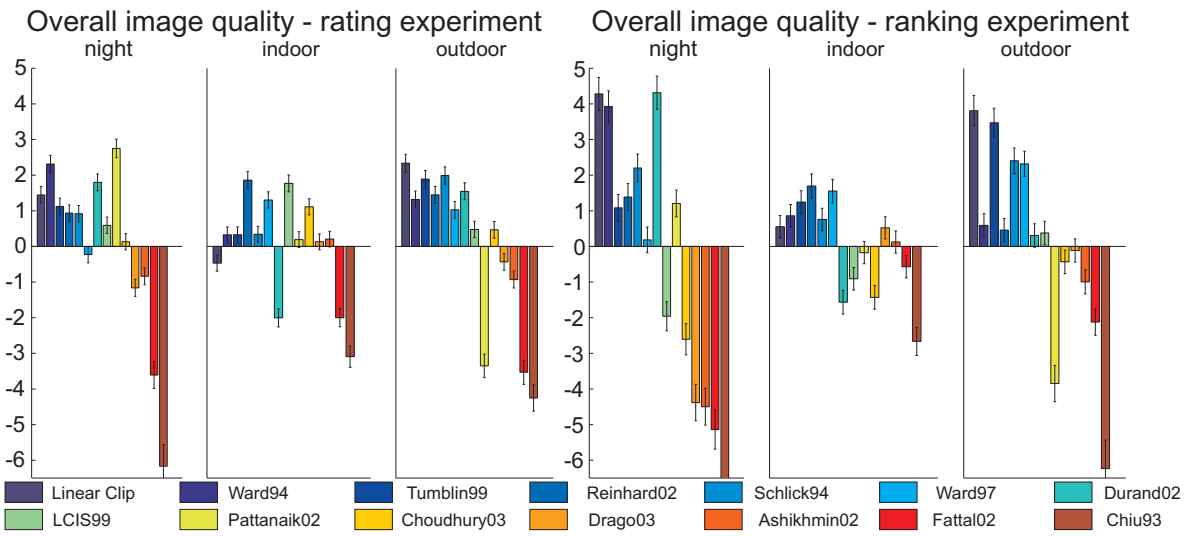


Figure 4.14: Overall image quality scores for each input scene. Rating (with a reference) experiment results (left) and ranking (without a reference) experiment results (right) for all examined TM methods. Left to right: night scene, indoor scene, outdoor scene.

the dependence of the overall image quality on the brightness, the contrast, the color, the detail reproduction, and the artifacts attributes.

We used different methods to fit functions to the attribute observation scores receiving the best approximation to the independently observed overall image quality. Using the simplest approach, *multivariate linear regression*, we obtained the following result:

$$OIQ = 0.07 \cdot Bri + 0.37 \cdot Con + 0.06 \cdot Det + 0.36 \cdot Col + 0.21 \cdot Art, \tag{4.1}$$

where OIQ is an overall image quality function, Bri, Con, Det, and Col, represent the quality of reproduction of brightness, contrast, details, and colors, respectively, all in the interval of $[0, 1]$ (0 meaning the worst reproduction). Art denotes the artifacts attribute in the interval of $[0, 1]$ (1 meaning no artifacts). To state how well the model explains the data, we computed the coefficient of determination: $R^2 = 0.76$. The high value of R^2 shows in our case that the linear regression approach is reasonable (a satisfactory value of R^2 for psychophysical experiments is over 0.7). In the second step, we determined which of the attributes actually contributed to the model. For this, we used the p-values of each attribute:
$p_{\mathrm{Bri}} = 0.8624$, $p_{\mathrm{Con}} < 0.0001$, $p_{\mathrm{Det}} = 0.0390$, $p_{\mathrm{Col}} < 0.0001$, $p_{\mathrm{Art}} < 0.0001$.
The only p-value that is higher than the threshold 0.05 is the brightness attribute, which means that the reproduction of brightness does not significantly influence the model. Furthermore, we can observe in the equation (4.1) that the *overall contrast* has the biggest weight factor and the detail reproduction the smallest one. This result may look surprising, as one would expect details to be more important. However, the global appearance of an image seems to depend much more on the quality of reproduction of other image attributes (contrast, color) and this confirms the good results of global TM methods as described in Section 4.7.3.

The low factor of *brightness reproduction* deserves special attention – it means that the brightness factor does not contribute to the proposed linear model. This could be caused by the fact

	OIQ	Bri	Con	Det	Col
brightness (Bri)	0.58				
contrast (Con)	0.80	0.64			
details (Det)	0.66	0.60	0.66		
colors (Col)	0.80	0.59	0.77	0.67	
artifacts (Art)	0.65	0.43	0.55	0.55	0.56

Table 4.10: Spearman correlations between the qualities of reproduction of image attributes.

that there is not a significant difference in reproduction of this attribute between the methods. However, we have found a significant difference in brightness already, see Section 4.7.1. To have another guideline, we computed the Spearman correlation coefficients between attributes, see the Table 4.10. These results show that there is a significant correlation between the brightness quality and the overall image quality. In the same time (not being in contradiction), the equation (4.1) suggests that the impact of brightness quality spreads into the other attributes, it reveals itself only indirectly. This effect is perhaps the best example that the basic attributes are very coherent or inseparable. Incidentally, the equation (4.1) shows **which attributes we should test** if we want to compare TM methods. There is no significant reason to evaluate the brightness since its effect is included in other attributes. The details quality attribute shows a similar Spearman correlation coefficient and weight factor in formula (4.1) as the brightness. However, because of its very small p-value, it contributes directly to the overall image quality, in contrast to the brightness.

Finally, we used multiple linear regression to examine the image attribute relations (Figure 4.7), with the following results:
$Bri = 0.35 \cdot Con + 0.26 \cdot Det + 0.13 \cdot Col + 0.0004 \cdot Art$,
$R^2 = 0.69$, $p_{Con} < 0.0001$, $p_{Det} < 0.0001$, $p_{\mathrm{Col}} < 0.0001$, $p_{\mathrm{Art}} = 0.99$.
Since the p-value of artifacts is over the 0.05 threshold, this result implies the idea that image artifacts do not contribute significantly to the perception of brightness quality.
$Con = 0.22 \cdot Bri + 0.14 \cdot Det + 0.49 \cdot Col + 0.12 \cdot Art$,
$\mathrm{R}^2 = 0.67$, $p_{\mathrm{Bri}} < 0.0001$, $p_{\mathrm{Det}} < 0.0001$, $p_{\mathrm{Col}} = 0.001$, $p_{\mathrm{Art}} = 0.001$,
$Det = 0.25 \cdot Bri + 0.19 \cdot Con + 0.30 \cdot Col + 0.23 \cdot Art$,
$\mathrm{R}^2 = 0.56$, $p_{\mathrm{Bri}} < 0.0001$, $p_{\mathrm{Con}} < 0.0001$, $p_{\mathrm{Col}} < 0.0001$, $p_{\mathrm{Art}} < 0.0001$,
$Col = 0.10 \cdot Bri + 0.50 \cdot Con + 0.23 \cdot Det + 0.12 \cdot Art$,
$\mathrm{R}^2 = 0.66$, $p_{\mathrm{Bri}} < 0.0001$, $p_{\mathrm{Con}} < 0.0001$, $p_{\mathrm{Det}} < 0.0001$, $p_{\mathrm{Art}} < 0.0001$,
$Art = 0.08 \cdot Bri + 0.23 \cdot Con + 0.34 \cdot Det + 0.27 \cdot Col$,
$\mathrm{R}^2 = 0.39$, $p_{\mathrm{Bri}} = 0.99$, $p_{\mathrm{Con}} < 0.0001$, $p_{\mathrm{Det}} < 0.0001$, $p_{\mathrm{Col}} < 0.0001$.

Due to rather small values of the coefficient of determination R^2 we can not make a deeper observation from the above equations. However, they show evidence of the relations between the attributes and their approximate weight factors. Moreover, it is evident that the basic attributes are very hard to separate. As we predicted in Section 4.5, there are cross effects, or more complex basic factors, which are not directly observable. However, for the amount of observation data we have, the linear regression approach is very reasonable and satisfactory, since we would need extremely large psychophysical experiments (with hundreds of subjects) for non-linear fits with cross effects of image attributes.

4.7.5 Comparison to Other Studies

In this section, we discuss and relate our results to other studies. A complete direct comparison is not possible, because we have evaluated more methods than the previous studies, and the aims of particular studies were slightly different. We should emphasize that our study was targeted to the natural reproduction of real scenes. Since our experimental input data are bound to natural scenes, the global TM methods (and local methods with a proper global part) were generally ranked better than the 'detail-hunting' and non human vision-aware approaches[7]. Our results show that the quality of reproduction of overall brightness, overall contrast and colors is much more important than the reproduction of details when naturalness is ranked in real scenes.

Still, the *good performance of global methods* is perhaps the most surprising result of our study. However, this is in a good accord with a recent psychophysical evaluation performed by Akyüz et al. [Akyu07], who show that outputs of sophisticated TM methods are statistically no better than the best single LDR exposure. Results of Yoshida et al. [Yosh05] also show distinctions between global and local methods, more specifically global methods performed better in the reproduction of brightness and contrast, while local methods exhibited better reproduction of details in bright regions of images. Even though Yoshida et al. claim that local methods perform better, we do not interpret their results so for the perceptual dimension, since (as one may see) in their results for naturalness (i.e. overall image quality) the first and the second best rated (out of seven) methods are global TM curves (Ward97 and Drago03). In the results of Ledda et al. [Ledd05] two investigated global methods performed averagely, in favor of the iCAM [Fair04] and Reinhard02 methods, but note that these methods are very strong in their global parts[8]. Looking at the results in the naturalness dimension reported by Drago et al. [Drag03b], we do not see the distinction between global and local methods, since Tumblin99 performs the best, but Ward97 is interestingly rated the worst. However, we should recall that observers did not have any reference in this experiment. Contrary to our results, Kuang et al. [Kuan07] report that local methods outperform global methods. However, basically the only global method that appears in their experiments is Ward97 with quite compelling results. To sum up: our results imply and we strongly believe that for a good performance in a perception targeted TM task, the TM method needs to have a significant global TM part. Then, the result may be sometimes enhanced using a local part that does not vanish in the global trend, e.g. [Cadi07].

The question of *correlation between the accuracy and preference experiments* is also very interesting. Ashikhmin and Goyal [Ashi06] demonstrate that using real environments is crucial in judging performance of TM methods and clearly show that there is a difference in subject's responses for a fidelity test with reference and without reference. Contrary to that, Kuang et al. [Kuan07] report a very strong correlation between the accuracy and preference experiments and state that one can use preference experiments in place of accuracy experiments with a real-world reference. Our results are perhaps closer to Kuang et al., since we did not detect

[7] Our results show that *statistically*, global techniques frequently outperform local TM approaches, even though local methods are generally claimed to perform better. Evidently this does not hold for all scenes, as can also be seen in our results. However, this is also a trend which matches our subjective personal experience.

[8] iCAM is generally a local method, but the adaptation values (for both luminance and colors) are calculated using a heavily blurred source image (very wide Gaussian), so that the method has a very strong global part and the method behaves to a big extent close to a global one.

statistically significant differences between the two performed experiments. However, our results do not exhibit as strong correlation as that of Kuang for overall image quality, and specifically not for overall brightness and reproduction of the details attributes.

Comparing *particular method performances* is quite tricky, since the results of TM methods may depend on implementation, and used parameters. Our results are in a good agreement with the evaluation performed by Drago et al. [Drag03b], where the Reinhard02 method was ranked the best and Schlick94 method was also ranked quite well. The difference is in Ward97 (histogram-based approach), where authors deliberately omitted the human-based ceiling function (we did not) and therefore the method favors the reproduction of details at the expense of naturalness. The consequences of Kuang et al. [Kuan04, Kuan06b] are also similar to ours: Fattal02 was considered not very natural while Reinhard02 (photographic mapping) was nearly the best ranked; we did not test iCAM. The only difference is with Durand02 (bilateral filtering method), which was ranked the best in Kuang's study (in our overall ranking Durand02 performed averagely). We believe this is caused by the implementation of the bilateral filter, since Kuang et al. use their specific modification of the original algorithm. In accordance with the original method description [Dura02], we have compressed the base layer using a scale factor in the log domain. More plausible global compression would result in a positively better outcome, but we aimed to compare purely the original approaches. This supposition is also supported by the conclusions of Ledda et al. [Ledd05], where the bilateral filtering approach performed the worst while other overlapping methods show perfect agreement as well (in the overall similarity test). Similarly to our results, in Yoshida et al. [Yosh05], the best-natural rated method was the Ward97, which is in accord with our results. The other results could not be compared easily, since Yoshida et al. tested the values (amount) of attributes while we inquired the reproduction quality of attributes.

4.8 Summary

In this chapter, we have presented an overview of image attributes for tone mapping that should facilitate access to the existing tone mapping literature. Since the attributes are intimately related, we have proposed a scheme of relationships between them. Moreover, we have proposed a measure for the *overall image quality*, which can be expressed as a combination of these attributes based on psychophysical experiments. We have verified the proposed ideas by means of two different psychophysical experiments.

The presented overview of image attributes is helpful for getting into the tone mapping field, or when implementing or developing a new tone mapping method. On the other hand, the identification of the relationships between the attributes is very useful for the subjective comparison of tone mapping methods. For example, we have found that overall brightness need not really be observed when there are the other attributes available. It also simplifies the comparison process by reducing the actual number of attributes that can be used to evaluate a tone mapping method. Finally, it represents the initial effort to design a truthful, objective comparison metric for high dynamic range images.

Using the results of two different experimental studies, with three typical real-world HDR scenes and 14 different tone mapping methods evaluated, this chapter presents one of the most comprehensive evaluations of tone mapping methods yet. Although there has been a lot of nice results in the field of local TM methods published, our results imply that the global

part of a tone mapping method is most essential to obtain good perceptual results for typical real world scenes.

An interesting and important result of the two different testing methodologies used (rating with reference and ranking without reference) is that almost all of the studied image quality attributes can be evaluated without comparison to a real HDR reference.

5 Hybrid Approach to Tone Mapping

"Where there is much light, the shadow is deep."

(Johann Wolfgang von Goethe)

Encouraged by the results of the evaluation of tone mapping methods presented above and also by the results of other perceptual studies, we propose a fast and simple yet powerful general hybrid approach to tone mapping in this chapter[1]. Our approach is based on combination of arbitrary global and local tone mapping methods. Since mentioned perceptual studies have shown high importance of preservation of global image attributes, we apply the global method first to reproduce overall image attributes correctly. At the same time, an enhancement map is constructed to guide a local operator to the critical areas that deserve enhancement. Based on the choice of involved methods and on the manner of construction of an enhancement map, we show that our approach is general and can be easily tailored to miscellaneous goals of tone mapping. An implementation of proposed hybrid tone mapping produces good results, it is easy to implement, fast to compute and it is comfortably scalable, if desired. These qualities nominate our approach for utilization in time-critical HDR applications like interactive visualizations, modern computer games, HDR image viewers, mobile device HDR applications, etc.

5.1 Introduction

Merits of High Dynamic Range Imaging (HDRI) are currently widely recognized in computer graphics, high-quality photography, computer vision, etc. However, HDRI becomes popular in interactive and real-time applications as well. Data visualization, computer games and other interactive applications gain new qualities thanks to HDRI. The reproduction of high dynamic (HDR) data on the low dynamic (LDR) output devices requires the reduction of dynamic range, commonly referred to as a tone mapping.

Many so-called tone mapping methods were proposed in the past [Devl02, Rein05b]. We can classify the existing approaches according to the transformation they apply to convert input luminances to the output values. *Global tone mapping methods* apply a tone reproduction curve (TRC), i.e. a function. Therefore, they transform particular value of the input luminance to one specific output value. *Local tone mapping methods* (operators) (TMO) may on the other hand reproduce particular input luminance to different output values depending on the surrounding pixels.

Although many sophisticated local tone mapping operators were published, these are typically not very generic approaches and just a few of these methods are suitable for interactive

[1] Recall that the topic of high dynamic range (HDR) tone mapping is introduced on page 3 and illustrated in Figure 1.2.



Figure 5.1: Global methods reproduce overall image attributes well, but wash away important details (left). Local approaches excel in reproduction of details, but may reproduce overall image attributes poorly (right).

and computationally limited applications. Even worse, hardly any of them can be marked as general and scalable. Besides, recent perceptual studies concerning the reproduction of HDR images [Yosh05, Ledd05, Cadi06, Akyu07, Cadi08] have shown high importance of preservation of global image attributes (overall brightness, overall contrast).

Generally speaking, global methods reproduce overall image attributes well, they are fast to compute, and easy to implement, but may wash away important details. Local approaches excel in reproduction of local contrast (details), but they are computationally intensive and may reproduce overall image attributes poorly, see Figure 5.1. Motivated by the mentioned findings we present a fast and simple yet powerful general hybrid approach to tone mapping issue. This approach takes advantages of both global and local tone mapping approaches to overcome mentioned limitations. Moreover, since the aims of tone mapping can differ among particular applications, the proposed approach can be easily tailored to the miscellaneous goals.

The rest of this chapter is organized as follows. In Section 5.2, we overview the previous work and we focus particularly on linear tone mapping methods. In Section 5.3, we introduce and describe the new hybrid tone mapping idea. In Section 5.4, we present two exemplary implementations of the hybrid tone mapping approach and show and discuss the results. Finally, in Section 5.5, we conclude and summarize the hybrid tone mapping approach.

5.2 Related Work

The areas of high dynamic range imaging and the tone mapping are currently quite complex. Refer to the state of the art by Devlin [Devl02] or to the book by Reinhard et al. [Rein05b] for an overview. Since we concentrate mainly on interactive applications and computationally efficient methods, we survey briefly the time-dependent methods suitable for interactive applications.

5.2.1 Time-Dependent Tone Mapping

Adaptation mechanisms of human visual system show time-dependency. Human needs some time to adapt to low luminance levels after entering the dark cinema from the sunlit street, for example. However, involvement of time-dependency is profitable also for non-perceptual applications of tone mapping, since it avoids flickering and other time-dependent artifacts.

One of the first time-dependent tone mapping methods was proposed by Pattanaik et al. [Patt00]. The method is based on statical color model by Hunt that is extended of time adaptation. Authors use global s-shaped curve for mimicking the response of both direct and inverse models of visual perception.

Scheel et al. [Sche00] shown that assumptions that were made for the tone mapping of still images had to be modified for the interactive context. Accordingly, authors introduced novel interactive method based on previous global approaches [Ward94, Lars97], however they did not consider time course of visual adaptation.

Durand and Dorsey [Dura00] used the global tone mapping method by Ferwerda et al. [Ferw96] for interactive tone mapping. Authors modeled visual adaptation course over the time for rods and cones.

Fairchild and Johnson [Fair03] adapted the iCAM model [Fair02] to account for time-dependent adaptation effects. The time-dependency provide two exponential filters that modify adaptation level. The used model is local (filtering using wide Gaussian curve) and therefore computationally intensive. Since the filtering kernel is very large, the properties of the iCAM outputs resemble global tone mapping results.

Ledda [Ledd04] proposed strictly local time-dependent approach. The method is based on initial effort by Pattanaik [Patt00], but adds local processing using bilateral filter. Time-dependency is modeled using exponential filters for rods and cones.

Krawczyk et al. [Kraw05] implemented a unified model accounting for various perceptual effects on graphics hardware. The model is based on the simplified local method of Reinhard et al. [Rein02] and it cleverly exploits necessary data structures of the tone mapping method to simulate veiling luminance, scotopic vision, and visual acuity.

The surveyed time-dependent approaches apply either global curves and thus they potentially destroy subtle details or they apply local methods and thus they are computationally demanding (except the GPU-aware approaches [Kraw05]). Furthermore, interactive applications often need to do some sort of load balancing, however there is an unanswered question, how to scale the time-dependent methods properly.

5.2.2 Linear Tone Mapping

As we have noticed, global methods reproduce overall image attributes well, see Figure 5.1. The group of global methods comprise a subset of linear tone reproduction curves. Despite of the simplicity of *linear* tone reproduction curves, the approaches utilizing linear (or close-to linear) mapping have many advantages that deserve our attention.

Since the linear methods scale image intensities by a constant (scale factor), they do not change scene contrasts for display. This is probably the reason why these methods [Cadi06, Cadi08] show to perform well in perceptual reproduction of the overall image attributes.

Linear tone mapping methods transform the input HDR image to the output image values using the scale factor, $L_d = mL_w$, where m is the *scale factor*, L_w is the input luminance, and L_d is the output value in the interval of [0, 1].

The simplest linear approach is the *maximum luminance* mapping, where we map the maximal input luminance to the maximal output value (e.g. to 1): $m = 1/L_{w\max}$, where $L_{w\max}$ is the maximal input luminance. As the maximal luminance is usually enormous in case of HDR images, this approach produces typically too dark and valueless results. *Mean value* mapping approach gives more reasonable outputs by mapping the *average* input luminance to the average output scale: $m = 0.5 \cdot 1/L_{\text{avg}}$, where L_{avg} is the average input luminance.

Ward's contrast based scale factor [Ward94] focuses on the preservation of *perceived contrast*. The computation of the scaling factor is based on Blackwell's [CIE81] psychophysical contrast sensitivity model. Almost the same principle of contrast preservation is exploited also in the work of Ferwerda et al. [Ferw96].

Another linear approach was introduced by Neumann et al. [Neum98a]. They propose the minimum *information loss* method that tries to mimic the photographer's practice to lose a minimum amount of information. The method automatically selects ideal clipping interval to obtain a minimum of detail-lost areas. The automatic selection of the interval is done by means of logarithmic image histogram.

Mapping using *s-shaped curve* [Patt00] is formally not a linear approach, but practically it can produce results that are very close to the linearly mapped results. S-shaped curves resemble transfer curves of classical photographic media.

5.3 Hybrid Tone Mapping

We propose a general idea of new hybrid tone mapping approach in this section, while two exemplary implementations are described in the next section. We do not concern ourselves with the inherent physically-induced issues of real output devices (gamma correction, etc.), in this chapter[2]. In the following, we assume (without the loss of generality) that the output device range is linear and is limited to the interval of [0, 1]. Moreover, we do not concern ourselves with the reproduction of colors here. All of the proposed methods aim to reduce dynamic range of the input image and the color information is left untouched. However, there is no obstacle to involve any method with a specific color processing into hybrid tone mapping approach if desired.

[2]Nevertheless, all the results presented in this chapter have been finally gamma corrected, using the value of $\gamma = 2.2$ as usual.

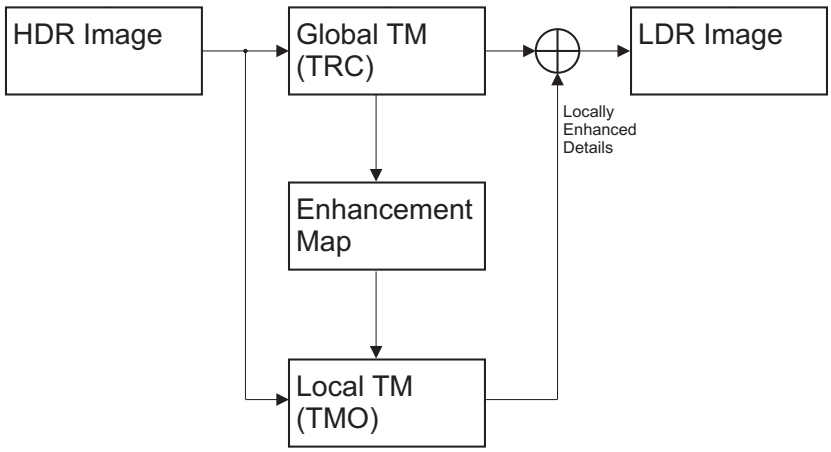


Figure 5.2: Overview of the hybrid tone mapping approach.

The overview of general hybrid tone mapping framework is shown in Figure 5.2. The process consists of three steps. First, the input high dynamic range image is transformed using global tone reproduction curve. This mapping produces the base of the output low dynamic range image and (if necessary) outputs desired values for the construction of an *enhancement map*. In the second step, the enhancement map is constructed considering the aim of the tone mapping (see Figures 5.3, 5.7). Finally, the enhancement map is used to guide a local tone mapping method to reconstruct subtle details. The locally enhanced details are then added back to the globally transformed image to improve the final image while maintaining good overall reproduction of image attributes. Due to the enhancement map, the local tone mapping method is applied merely to the critical areas of the original image and therefore many computational resources are saved.

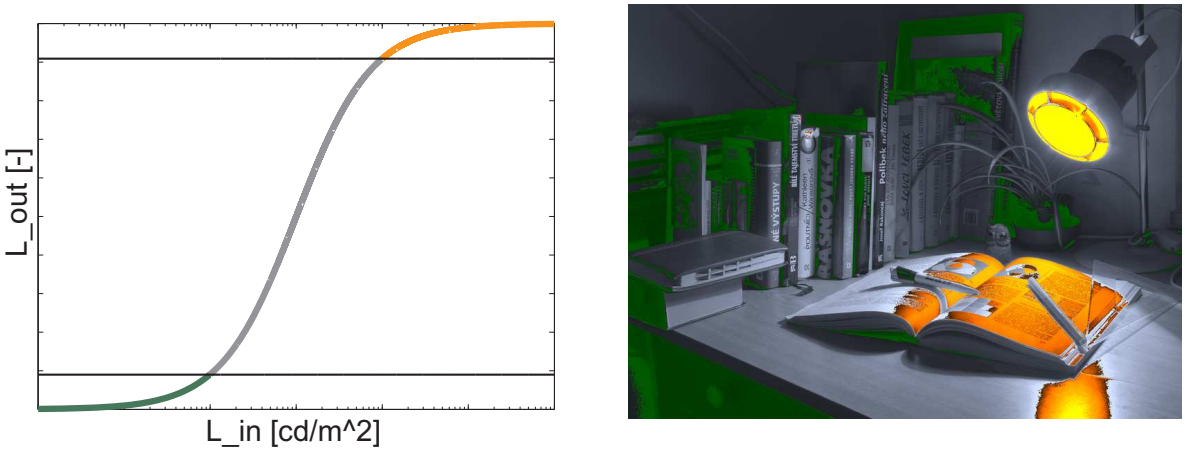


Figure 5.3: Right: example of an enhancement map (orange and green areas) for the perceptually plausible hybrid approach. The luminances in orange and green colored areas are potentially vanished due to the excessive compression of the reproduction curve (left) and due to the insufficient numerical precision.

5.3.1 Construction of Enhancement Map

Besides the choice of the involved methods, the manner of the construction of the enhancement map is an essential heart of the hybrid tone mapping. The enhancement map therefore has a profound effect on the properties of the hybrid tone mapping approach as a whole.

The enhancement map is generally a map of float numbers with the same dimensions as the original HDR image. In the examples shown in Section 5.4, the enhancement map is constructed using a sort of thresholding of input luminance values. However, there is no obstacle to construct the enhancement map in a different way (for example based on input gradients, etc.). When we use a linear transform function in a global part of hybrid mapping, the construction of the enhancement map is very simple and effective. As we know the analytical equation of the transfer curve, the clipping values are then also directly known and the enhancement map is constructed effortlessly by thresholding. For the other approaches (log scaling, s-shaped curves, etc.), we propose to use the clip values of 5th and 95th luminance percentiles. However, the manner of the enhancement map construction can reflect a special goal of implemented method, as we show in Section 5.4.3.

Due to the abrupt changes of the global mapping to the local mapping (and vice versa), artifacts might appear on the borders of the enhancement map. To avoid the artifacts, we propose to construct and apply the enhancement map as follows (we show the process for only one clipping value L_{clip}):

$$EM(x,y) = \begin{cases} 0 & \text{if } TRC(L_w) < L_{\text{clip}} \\ \min(1, TRC(L_w) - L_{\text{clip}}) & \text{otherwise,} \end{cases} \tag{5.1}$$

where $EM(x,y)$ is the enhancement mask value, $TRC()$ is a global tone mapping method, L_w is the input luminance, and L_{clip} is the clipping luminance. The final output value L_d is then computed as a weighted sum of a global method and a local method outputs:

$$L_d(x,y) = EM(x,y) \cdot TMO(L_w) + \Big(1 - EM(x,y)\Big) \cdot TRC(L_w), \tag{5.2}$$

where $TMO()$ is the involved local method. An alternative way to reduce artifacts is to blur the enhancement map, and there are perhaps yet another possibilities.

If the computational time is not an issue, we can involve more sophisticated criteria to the construction of an enhancement map, e.g. the human visual system properties. Using the visual attention model, for example, we can pass the computational resources to the visually important areas of the image. Another possibility is the usage of contrast sensitivity function (CSF) during the construction of the enhancement map. In this case, the effort of local method will be directed to the areas, where the detail is (at least potentially) visible for a human observer.

5.4 Use Cases and Results

The proposed idea of hybrid tone mapping is general – virtually any combination of existing (and potentially forthcoming) methods is possible. However, the choice of the involved methods has to reflect the intent of the resultant combination (e.g. an aesthetic view, a cognitively

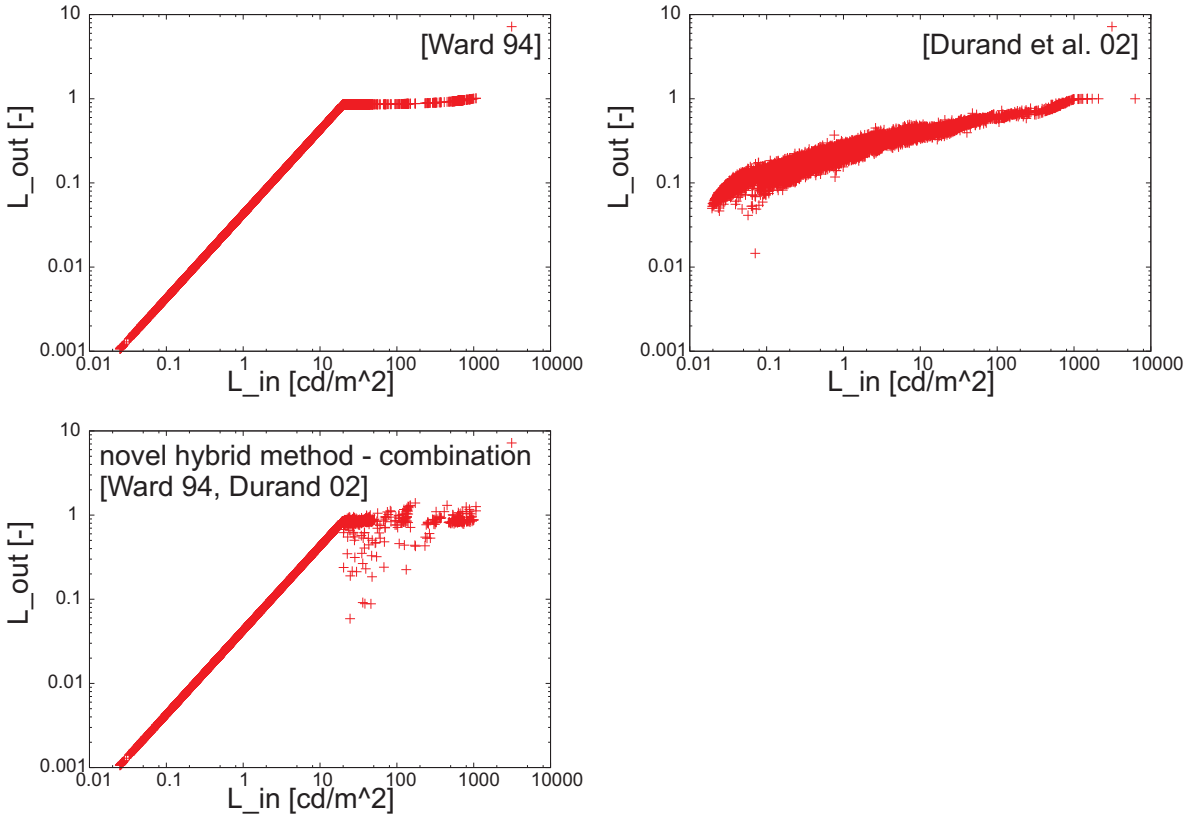


Figure 5.4: Tone mapping transforms, top left: global method [Ward94] maps input HDR values via linear function – note the clipping of high luminances, top right: local method [Dura02] applies different transform to different pixels – the reproduction of overall image attributes is poor, bottom left: our hybrid approach [Ward94, Dura02] combines merits of both the global and local approaches.

rich depiction, or a perceptually plausible reproduction). A combination of methods that is excellent in reproducing details can fail miserably when we aim in reproducing the perceptual experience of an observer. In this section, we show two different examples of the hybrid approach: perceptually plausible approach and cognitively rich approach. We show and discuss actual outputs of these methods and we also exhibit the performance values comparing to original local methods.

5.4.1 Fast Perceptually Plausible Approach

For such interactive applications where the perceptually convincing reproduction is desired and where the computational resources are limited (e.g. in computer games), we propose fast and simple implementation of hybrid tone mapping approach as follows.

At the post of global method, we use the linear mapping by Ward [Ward94]. This method was proven to give reasonable results for natural scenes [Cadi06] (as described in previous chapter), and the computational demands of the method are minimal.

Since the global part is purely linear, we can construct the enhancement map directly by thresholding of input luminance values. The exact clipping values are known: $L_{\text{clipLO}} = 0$ and $L_{\text{clipHI}} = 1$ for an original method. We can modify the approach by shifting the transform curve if desired (to allow the user to adjust brightness or contrast), but even in this case, the clipping values are easily found analytically. Pixels with luminance values outside of the linear interval would be clipped and therefore the information would be lost there. Therefore, these pixels form the enhancement map.

Having the enhancement map, we apply the bilateral filtering method [Dura02] just on the areas marked in the map. Bilateral filter separates the original luminance map to the base layer and the detail layer. We use the detail layer to enhance the result of the global tone mapping method. For acceleration of the local filtering, we utilize graphics hardware (GPU) [Fial06, Kraw05]. Figure 5.4 compares the transforms of the original methods and the hybrid approach and Figures 5.5, 5.6 show the results in the form of output images.

5.4.2 Time-Dependent Hybrid Mapping

It is usually advantageous to model the course of visual adaptation over the time for interactive applications. Time-dependency of tone mapping is twofold profitable: it increases the perceptual quality on one hand, and it also avoids temporal image artifacts on the other hand.

The way of implementation of time-dependency is influenced by the goal of the whole tone mapping method and it is not necessary to realize it at all, in some cases. In accordance with other authors [Dura00, Patt00] we use an exponential decay function in our perceptually plausible hybrid approach (described in Section 5.4.1) to model the light adaptation. We omit the simulation of long-term dark adaptation due to its subtle and slow effect and due to efficiency reasons. We modulate the adaptation level $L_{a(w)}$ in the Ward's method [Ward94] by the exponential function for smooth transitions when tone mapping a dynamic environment: $dL_{a(w)}/dt \approx (L_{a(w)} - L_{a(w)(t)})/\tau$, where $L_{a(w)}$ is the visual adaptation for static image, $L_{a(w)}(t)$ is the actual adaptation and $\tau = 0.1$ is a time constant that mimics the speed of adaptation.

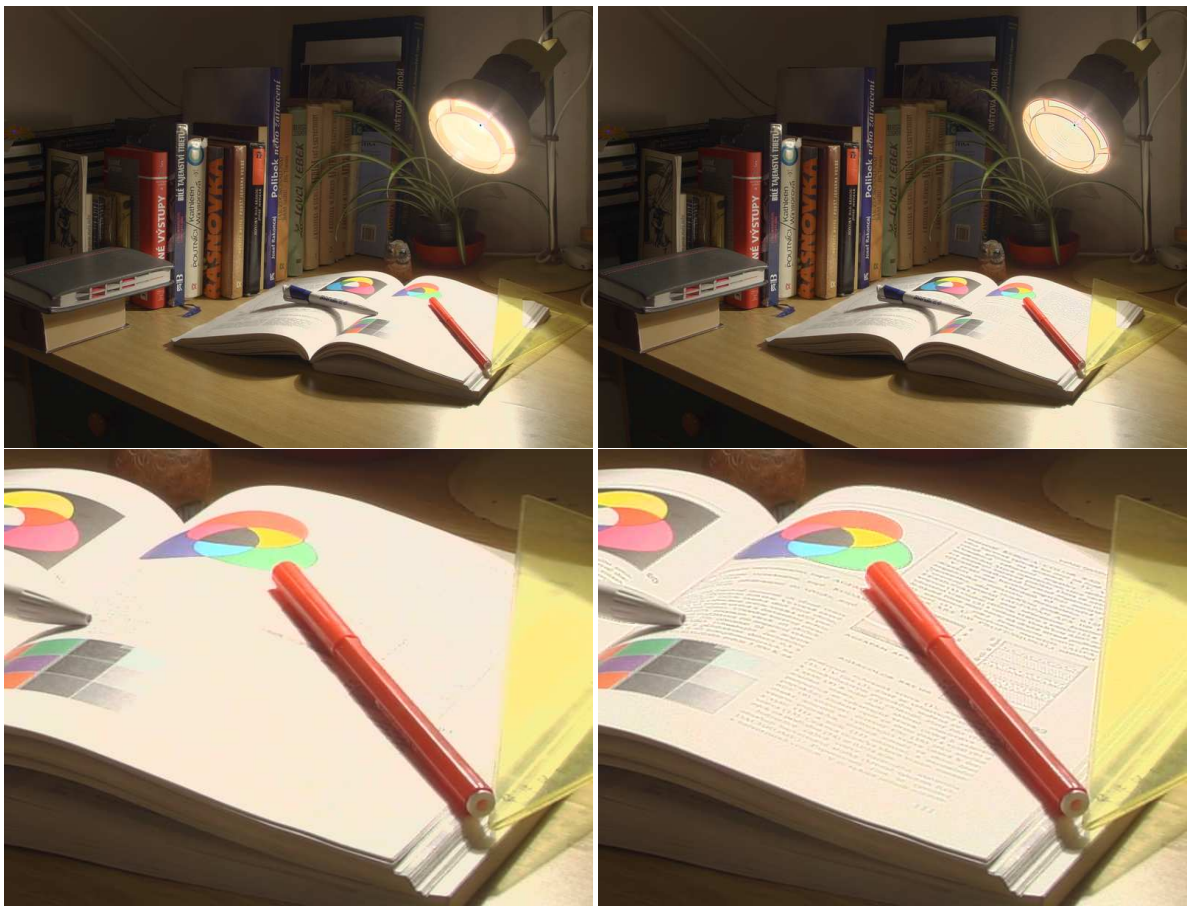


Figure 5.5: Top left: result of purely global method [Ward94] exhibits well reproduced overall contrast, however shows the lack of subtle details, top right: result of the new hybrid approach [Ward94, Dura02] preserves the overall contrast accurately, and adds the lost details. Bottom: close-ups of the book, note the reproduced details in the hybrid approach result.

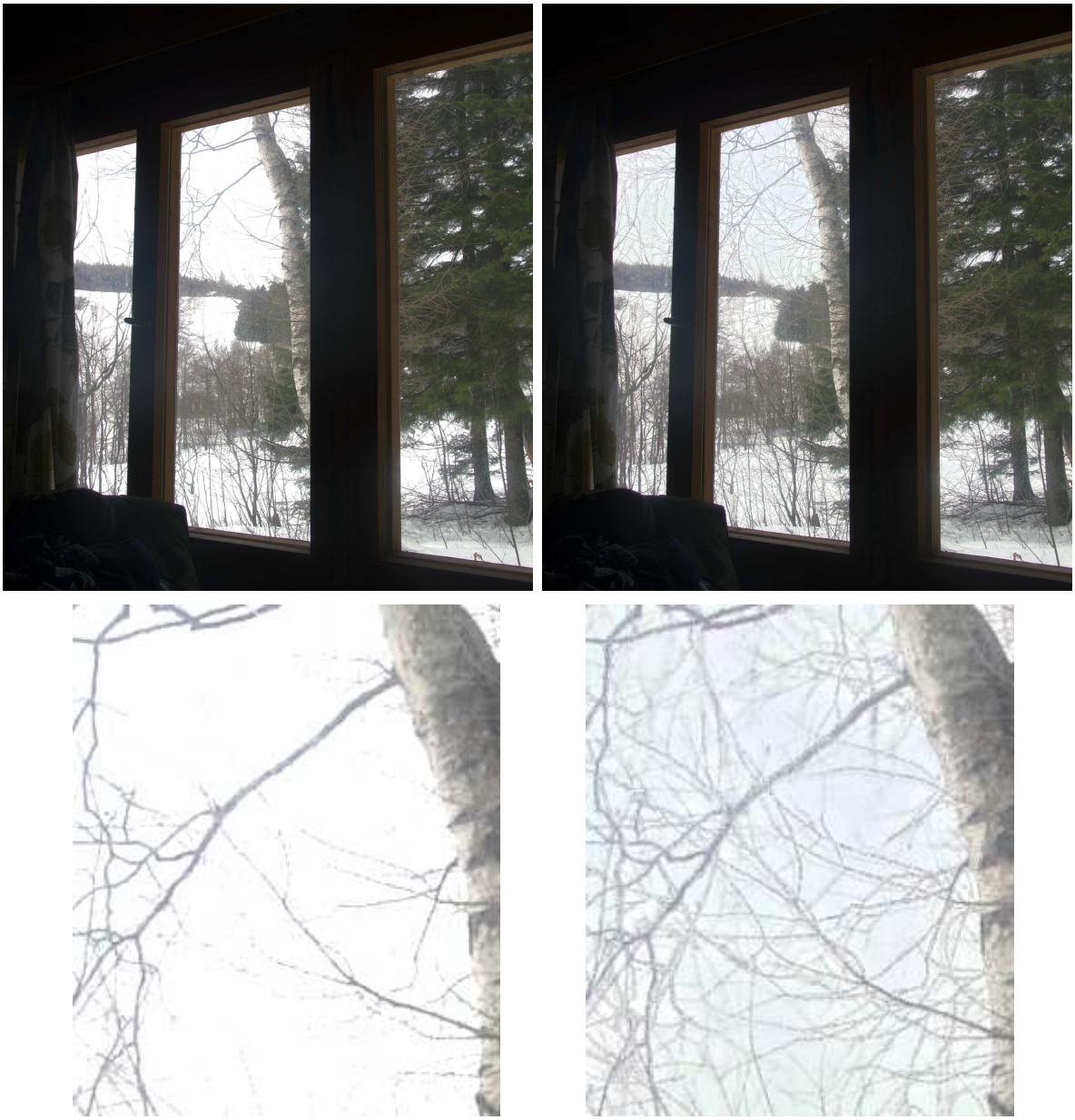


Figure 5.6: Top left: global method [Ward94] reproduces correctly the overall impression of the scene, however it washes away details seen through the window. Top right: hybrid approach [Ward94, Dura02] enhances the image without affecting the overall image attributes. Bottom: close-ups of the tree seen through the window, note the revival of the birch twigs and the bluish color of the sky in the hybrid approach result (bottom right).

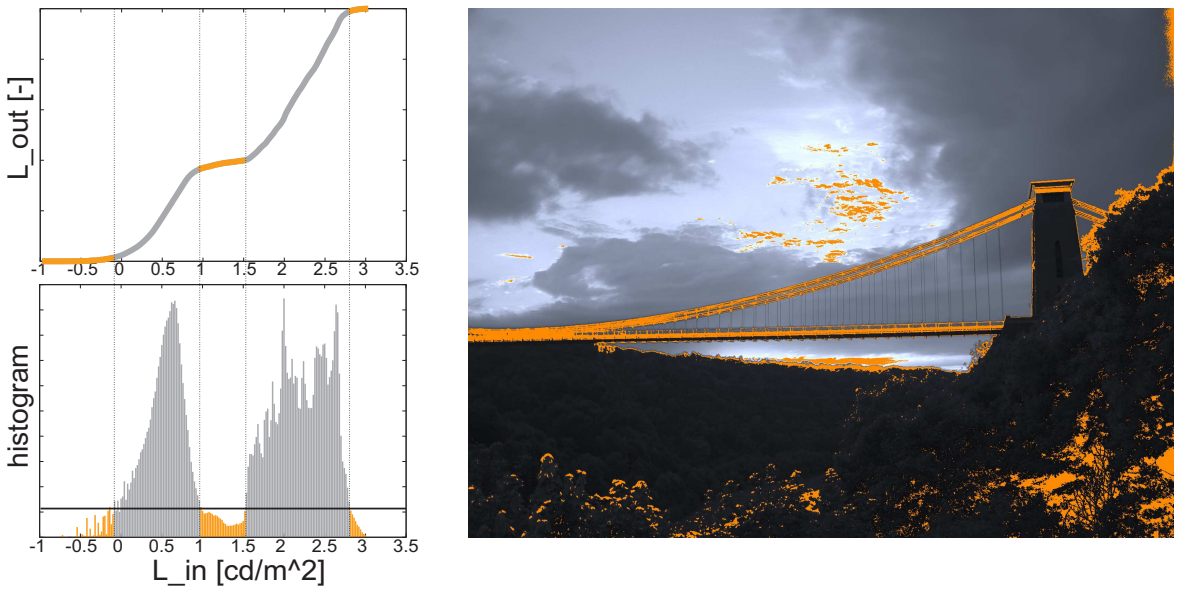


Figure 5.7: Illustration of the cognitive hybrid approach. Top left: cumulative transform function [Lars97], Bottom left: histogram, with detected areas (orange) for local enhancement. Right: corresponding enhancement map (orange areas). *Source HDR image (Bristol bridge) courtesy of Greg Ward.*

Similar approach is amenable in many other hybrid tone mapping implementations, because we can usually smoothen the response of particular parameter of involved global tone mapping method. If the computational cost is the main limitation, the time-dependency may be omitted temporarily with reasonable loss of reproduction quality.

5.4.3 Cognitive Approach

As an example of cognitive (perceptually plausible detail-oriented) hybrid tone mapping approach, we propose the combination of the histogram adjustment global tone mapping method [Lars97] enhanced by the locally applied bilateral filtering [Dura02], see Figure 5.7.

The histogram adjustment method grants most of the available device contrast to the areas of abundant luminance values in the input HDR image. Generally speaking, large areas in an input image are given more contrast (thus subtle details present at these areas may become visible) at the expense of tiny areas. This advanced 'distribution of contrast' is achieved thanks to cumulative function derived from formerly constructed image histogram, see Figure 5.7. The cumulative function is then used as a tone reproduction curve to transform input luminance to output values.

In accordance with the choice of involved methods, we propose to construct also the *enhancement map* seeking the same goal of cognitively rich (detailed) output image. We detect the areas of small local contrast on the chart of cumulative function constructed in the previous step – these areas represent pixels, where the detail is potentially vanished. It is possible to use the second derivative of cumulative function (note that the derivative is never a negative number in our case), or histogram thresholding, for the detection. Sections that show low



Figure 5.8: Further results of the new hybrid approach. Left: hybrid approach [Ward94, Dura02] – perceptually plausible reproduction of Memorial church image, right: more cognitive rendition of the same input image by hybrid combination of histogram adjustment and trilateral filtering [Lars97, Chou03]. *Source HDR image courtesy of Paul Debevec.*

slope (low values of the second derivation) are selected to construct the enhancement map (see Figure 5.7) – each pixel with the luminance value within the boundary of a selected section becomes a member of enhancement map.

Finally, we apply the local trilateral filtering method [Chou03] on the original input pixels that are present (or masked) in enhancement map. Similarly to the bilateral filter, the trilateral filter produces blurred image, but preserves significant luminance edges. We obtain local details by dividing the original image by the blurred image. Finally, we enrich the result of histogram adjustment method by these details. See Figure 5.8 where two renditions of one input image (Memorial Church) are shown to compare the two presented hybrid implementations.

	Input [Px]	Enhancement map [Px, %],	Speedup [-]
Perceptual method (Sec. 5.4.1)	1461568 Px	96152 Px, 0.0014%	118.5
Cognitive method (Sec. 5.4.3)	1461568 Px	480080.3 Px, 0.132%	41.7

Table 5.1: Comparison of performances of two different implementations of hybrid tone mapping (average results over 10 HDR images). The enhancement map percentage values denote the average ratio of the enhancement map size over the input image size, the speedup values mean the average acceleration of hybrid approach against the original, completely local approach.

5.4.4 Performance Results

Since one of the goals of the hybrid approach is the reduction of computational complexity, we present here the numerical performance results, see the Table 5.1[3]. The table shows average values measured at the group of 10 input HDR images. The perception-targeted fast hybrid approach is on average 3 times faster than the cognitive method and more than 118 times faster than the original bilateral filtering method. The reported speedup is gained thanks to the enhancement map. Since the enhancement map contains usually just a small portion of the original image pixels (as shown in the table), the time-demanding local approach is applied locally, to the small (necessary) part of the image.

Generally speaking, our technique induces very small additional load to the system leaving large space for other computations. This is very profitable in interactive applications like the computer games, etc. However, note that besides the performance improvement, hybrid tone mapping can enhance the quality of the output image as well (see Figure 5.5).

In the imminent future, we can expect the need of dynamic range reduction even on various portable devices and on small and computationally elementary machines. The hybrid tone mapping will be reasonable in this case as well, thanks to its *good scalability*. If we face the lack of computational power, we can modify (soften) the criteria of the construction of the enhancement map. Depending on these criteria, we are able to continuously balance the computational load spanning from the complete locally enhanced method up to the factual omitting of the local enhancement (i.e. purely global tone reproduction). Finally, the other possibility to decrease the time consumption is to omit the time-dependent processing, as noted in Section 5.4.2.

5.5 Summary

In this chaper, we presented a novel hybrid approach to the issue of reproduction of high dynamic range images on devices with limited dynamic range of luminance (i.e. the tone mapping issue). In our approach, we combine results of arbitrary global and local tone mapping methods. Recent perceptual studies concerning the reproduction of HDR images have shown high importance of preservation of overall image attributes. Motivated by these studies, we apply the global method first to reproduce overall image attributes correctly. At

[3]The performance values strongly depend on the selection and the implementation of the involved methods and we therefore present the average values for an overview.

the same time, we construct an enhancement map to guide a local operator to the critical areas that deserve enhancement.

We do not invent another complex tone mapping method, but we rather propose a general framework that utilizes already known ideas and combines existing and potentially forthcoming methods. We have shown that the presented hybrid approach can be easily tailored to miscellaneous potential goals of tone mapping (e.g. to get perceptually plausible images, to get detail-rich depictions). Our experiences indicate that an implementation of proposed general hybrid tone mapping approach typically produces reasonable results. Further, the technique is easy to implement, fast to compute and it is easily scalable, if desired. These qualities nominate our approach for utilization in time-critical HDR applications like interactive visualizations, modern computer games, HDR image viewers on mobile devices, etc.

6 Perceptually-Plausible Conversion of Color Images to Grayscale

---

"The only things in my life that compatibly exists with this grand universe are the creative works of the human spirit."

(Ansel Adams)

So far, we have concerned ourselves with the problem of reduction of high dynamic range of luminances, which is in fact the transformation of values in one dimension of space. In this chapter, we compete with even more difficult puzzle – how to convert color images to grayscale to obtain perceptually plausible results. Color to grayscale transformation is basically a reduction of (at least) three dimensional data into just one dimension, see Figure 1.3 on page 3 for illustration. It is clear that some loss of information is inevitable in this case.

Color images often have to be converted to grayscale for reproduction or for subsequent processing. Conversions which take into account local color changes are better at preserving details than global methods such as simply selecting the luminance channel. However, color changes across an image are expressed as color difference formulas with no direction, and the challenge is to find new formulas which can be used to determine an equivalent and consistent set of transitions between different gray levels. We present two new local grayscale conversion techniques in this chapter: one is based on the CIELab color difference formula, and the other on the Coloroid system and its experimental background. Using one of the new formulas, we obtain an inconsistent gradient field from a color image, find the closest consistent gradient field, and then integrate that to produce a grayscale image. The complexity of this method is linear in the number of pixels, making it suitable for high-resolution images; the results are pleasing and perceptually plausible.

6.1 Introduction and Related Work

Although today's digital cameras take color images, many laser printers still use one color of toner, and most of the pictures in the world's newspapers are in black and white. Conversion from color to grayscale therefore remains an important piece of the digital imaging puzzle. In conventional black-and-white photography tones are determined by the spectral sensitivity of the emulsion, and can be varied in an active way by selecting different filters to enhance one part of the spectrum. This analog multispectral technique can now be emulated by digital image processing, and this approach to the problem can be considered as a 'global' technique.

An alternative, spatial approach to color-to-grayscale conversion has been proposed by Bala and Eschbach [Bala04]. They preserve chrominance edges locally by introducing high-frequency chrominance information into the luminance channel. A spatial high-pass filter is applied to the chromatic channels, the output is weighted with a luminance-dependent term,

and the final result is added to the luminance channel. Grundland and Dodgson [Grun05] proposed the Decolorize algorithm for contrast enhancement as well as color-to-grayscale conversion. They perform a global grayscale conversion by expressing grayscale as a continuous, image-dependent, piecewise linear mapping of the RGB primary colors and their saturation. Three parameters are used to control contrast enhancement, scale selection and noise suppression, and image-independent default values for these parameters have been proposed [Grun05]. A somewhat different approach was taken by Gooch et al. [Gooc05], who introduced the pioneering local algorithm known as Color2Gray. In this gradient-domain method, the gray value of each pixel is iteratively adjusted to minimize an objective function based on local contrasts between all the pixel pairs. The method requires three tunable parameters (θ, α, μ), for which image-independent values are not provided. Moreover, the complexity of the method is very prohibitive $O(N^4)$ (see Figure 6.1), which can be improved by limiting the number of considered differences (e.g. by color quantization), however at the cost of possible artifacts. Mantiuk et al. [Mant06] showed an application of their contrast processing framework for conversion of color images to grayscale. Their method is based on Gooch's [Gooc05] idea, but involves a more efficient approach, in which the close neighborhood of a pixel is considered on fine levels of a pyramid while far neighborhood is covered on coarser levels. Authors claim that this enables them to convert bigger images and perform computations faster, unfortunately no complexity analysis is reported. Another conversion to grayscale that aims to preserve the contrast while maintaining consistent luminance was introduced by Rasche et al. [Rasc05], who approached the problem by means of a constrained multidimensional scaling, but the execution time of this approach is highly sensitive to the number of colors. This can be ameliorated by color quantization [Rasc05], but it results in undesirable artifacts, see Figure 6.2, making the use of the method proposed by Rasche et al. very questionable for photographic images.

Color-to-grayscale conversion is a mapping of three-dimensional set with strong spatial coherence to a one-dimensional space, and necessarily involves some loss of information. We believe that global transformation approaches cannot be a full answer to the problem, although they have been shown to offer some fast and acceptable results. Adaptive local methods hold the promise of much better solution, although they suffer from theoretical problems in perceptual modeling, as well as computational difficulties. Finding the conversion that provides the greatest perceptual plausibility is still an open problem. For instance, Fairchild and Pirrotta [Fair91] have proposed a formula to convert CIE luminance into perceived lightness by taking into account the Helmholtz-Kohlrausch effect, which makes chromatic colors appear lighter than equiluminant achromatic tones. This formula seems to offer an alternative to the approach of Gooch et al. [Gooc05], but it does not provide an estimate of the local gray gradient for arbitrary color pairs, making it inadequate to convert real images with arbitrary colors. But human aesthetic and perceptual abilities, which are the result of complicated neural activity, may not prove ultimately reducible to simple formulas. Computer graphics and image processing has tended to avoid the methodological quandaries posed by working directly with empirical observations from psychophysical experiments, preferring perceptual models based on analytical approximations.

In this chapter, we present two new local color-to-grayscale conversion techniques (i.e. solutions to both analytical and experimental approaches). The analytical solution is a generalization of the CIELab formula, which introduces a signed power function to give a signum to the weighted Lab components. The experimental solution aims to obtain the best perceptual gray

gradient equivalence by exploiting the Coloroid system and its experimental background. As opposed to the use of either luminance or chrominance to compute gray gradients [Gooc05], our two new approaches take into account both quantities, giving (signed) chrominance between 2 and 3 times less weight than luminance. The main contributions of this chapter are: (1) the two gray-gradient formulas mentioned above; (2) a new and efficient way of converting an inconsistent gradient field into a consistent one, which can be directly converted into an image by a simple 2D integration; (3) pleasing grayscale images that preserve high-frequency details and contrast effects without artifacts; (4) an absence of tuning parameters, much faster operation than previous local approaches, and easier implementation. We believe that the benefits of this algorithm will inspire the wider use of gradient domain methods in computer graphics and image processing.

The rest of this chapter is structured as follows: we first describe the Coloroid color system and present our new observations based on Coloroid in Section 6.3. In Section 6.4 we introduce an efficient gradient-based color-to-grayscale conversion algorithm that uses a new gradient inconsistency correction method. A selection of results produced by this algorithm are presented and discussed in Section 6.5. Finally, in Section 6.6, we give a summary of the proposed methods.

6.2 A Simple New CIELab-Based Formula

Before turning to the Coloroid formula, we studied the possibility of extending the Color2Gray method [Gooc05], eliminating the artifacts and to reducing the computational cost. Using the CIELab values, Color2Gray computes a warm-cold hue transient value, which is then multiplied by the chroma, and finally modified by a stretched tanh function (the 'crunch') to obtain the chrominance. The signed gray difference that is used is either the chrominance or the luminance, depending on which has the largest absolute value. However, this approach can result in a strongly inconsistent gradient field, e.g. a large negative value can appear immediately after a large positive one. To 'blur' this kind of artifact requires a large neighborhood, and increases the effective complexity to $O(N^4)$.

To overcome these shortcomings, we abandoned the use of a simple maximum, and formulated a continuous function using the CIELab space. If we consider the max function to be the $n = +\infty$ power-norm, then the 3rd power-norm has a similar effect, and it is also similar to a square root, we use it as follows:

$$\Delta = (\Delta L^3 + w_a \cdot \Delta a^3 + w_b \cdot \Delta b^3)^{1/3}. \tag{6.1}$$

This formula directly conveys the sign of Δ, which is the gray difference. The weights w_a and w_b applied to the chrominance channels are in the range $[0.3 - 0.6]$, so as to reduce the chrominance-luminance ratio. The resulting luminance gradient is smaller than a CIE color difference value, which can exceed 200.

In contrast to method published by Gooch et al., the proposed technique takes just a fraction of a second to process an image, and the results, as shown in Figure 6.1, preserve more details. In this figure, the gradient field was corrected using the method to be described in Section 6.4.2. This method can be simply extended with 2 weight factors for each opposing color channel, with different values for positive and negative a and b (red-green and yellow-



Figure 6.1: Comparison of our method using the CIELab formula with the CIE Y equivalent and the method by Gooch et al. [Gooc05]. Top left: original color image. Top right: the CIE Y equivalent. Bottom left: the result of Gooch et al. (parameters: $\theta = 45°$; $\alpha = 10$; $\mu =$ entire image; computation time ≈ 850 seconds). Bottom right: the result of our adaptive color-to-grayscale conversion (computation time ≈ 0.1 seconds). *Color image (312 × 255 pixels, 255 colors): Impressionist Sunrise by Claude Monet.*

blue), meaning that 4 weight factors are needed in total. Other powers p can be used in Equation (6.1) instead of 3, generalizing to the the signed power function:

$$\text{powsign}(x, p) = \text{sign}(x) \cdot |x|^p, \tag{6.2}$$

so that $p = 2$ corresponds to a signed version of the CIELab formula, because $\text{powsign}(-5, 2) = -25$.

6.3 Color-to-Grayscale Conversion Based on the Coloroid System

6.3.1 The Coloroid System

While luminance obviously plays a role within the human visual system, color-to-grayscale conversion contains additional elements, because local luminance and chrominance changes are

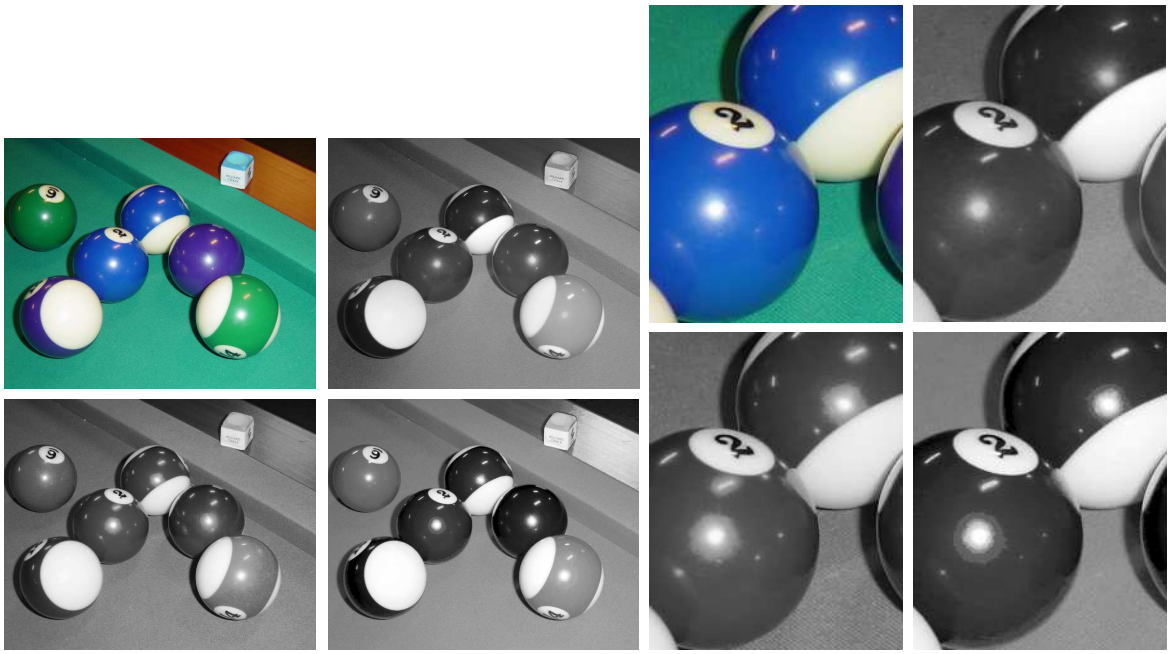


Figure 6.2: Comparison of our method using the Coloroid formula with Gooch's method [Gooc05] and Rasche's method [Rasc05] (left: entire results, right: close-ups). Top left: original color image courtesy of Karl Rasche. Top right: the result of our adaptive color-to-grayscale conversion. Bottom left: the result of Gooch et al. Bottom right: the result of Rasche et al., notice the quantization artifacts.

not transformed to luminance gradients in the brain. However, some perceptually justifiable values can be determined through observations using an appropriate arrangement of color and gray samples, and the Coloroid system [Nemc87, Neum05] appears to be the appropriate tool for this purpose.

The Coloroid system involves a color ordering and a continuous color space developed to describe and determine aesthetic relationships between colors. It is based on a uniquely large number of observations, by 80 thousand observers who made 26 million elementary choices. The basic perceptual quantities measured, (i.e. hue, saturation and lightness), have a very good correlation with the Coloroid model parameters, as follows:

- surfaces of constant hue (A) form a plane, containing a neutral axis and a hue-dependent limit-color, unlike most other systems, which have curved hue surfaces (e.g. the Munsell system)
- the saturation $T = C \cdot p$, where p is the content of the limit-color, and the constant C depends on the hue
- the lightness $V = 10 \cdot Y^{1/2}$. Unlike spaces based on a ds line-element, the Coloroid does not contain a cube root or a logarithm in the lightness formula

Figure 6.3 (left) shows the circle of 48 limit-colors, while Figure 6.3 (right) shows the continuous 3D limit-color line. Figure 6.4 (left) demonstrates the typical shape of the Coloroid

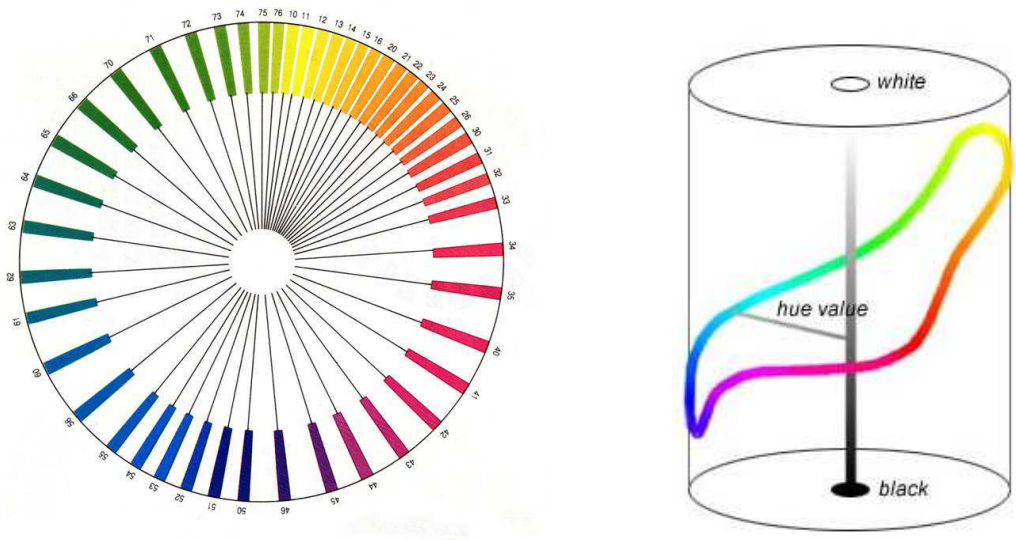


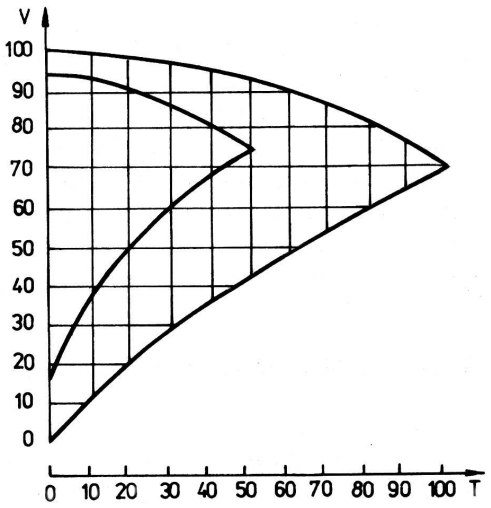
Figure 6.3: Left: the 48 limit-colors of the Coloroid system. Right: a curve of Coloroid limit-colors in 3D.

gamut at a fixed hue value. The lightness of the most saturated point depends on the hue, as we see in Figure 6.3 (right). The two Coloroid gamuts represent two limit-color selections: the larger corresponds to the spectrum or purple limit-colors and the smaller to the most saturated solid colors, which will be used in this chapter. The concepts and formulas of Coloroid can be found elsewhere [Nemc80, Nemc87, Hunt92], and there has also been an in-depth survey of application areas [Nemc04].

6.3.2 Observation Conditions of the Coloroid Experiments

Coloroid emulates real-life observation conditions. These 'natural' observation conditions differ from other systems due to the use of a wide viewing angle with the semi-adapted eye. The experiments were made in indoor conditions under north-light illumination (approx. 1,800 lux). Observer saw a set of large color samples (between 5×5 and 10×10 cm) simultaneously within a wide viewing angle. Then one of the colors was replaced with a gray sample, and the observer had to decide about the perceptual uniformity of the set (providing a yes or no answer) within a short time that did not permit adaptation of the eye. In other tests the only answers allowed were "bigger, smaller or approximately equal".

We generated a sequences of colors with fixed hue (A) and lightness (V), but uniformly increasing saturations (T). Then we used the Coloroid arrangements to evaluate these sequences, replacing the given color with the appropriate gray sample for every saturation to conserve perceptual uniformity. The unexpected result is that the equivalent gray difference does not change monotonously! For example, on the $A = 60$ hue page at the lightness $V = 85$ in the range of solid colors in the Coloroid space, we obtained the following values:



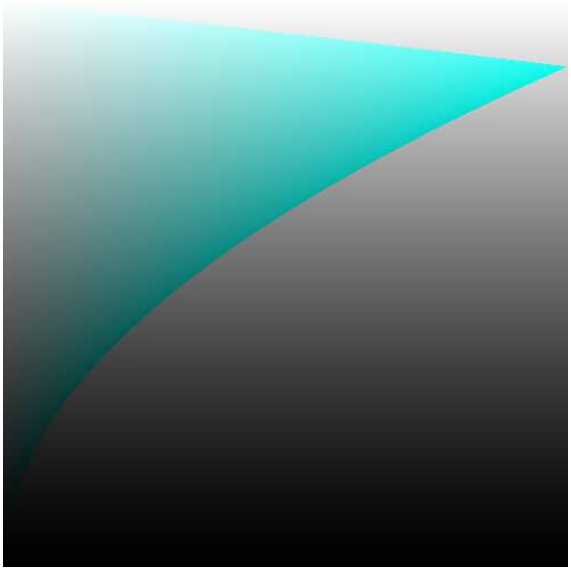


Figure 6.4: Left: shape of the Coloroid gamut for a fixed hue. Right: turquoise hue plane in the Coloroid space.

Relative saturation difference	1	2	3	4	5
Relative Δ-gray	1	2	4	0	-5

6.3.3 Perceptually-Equivalent Color and Grayscale Attributes

We now present our own experiments in color-to-grayscale conversion in more detail. First, we studied the average relative luminance difference of hue pairs. We first selected seven basic colors, one from each Coloroid hue group: Table 6.1 shows their Coloroid hues A, their characteristic wavelengths λ, and the angles φ that correspond to the $D65$ white point in the CIExy chromacity diagram. The observers' task was to rank the color and gray samples, or to fit missing samples into color series with a short adaptation time (i.e. the semi-adapted eye) [Nemc04]. The resulting 7×7 matrix (see Table 6.2) contains the relative gray-equivalent differences. The matrix has a zero diagonal and is anti-symmetric. The maximum value is scaled to 10. The largest perceived difference is between the A = 50 blue hue and the A = 10 yellow hue. For arbitrary (A1, A2) saturations, we applied 4 linear interpolations, which preserved the anti-symmetric property of the perceived differences. In Section 6.3.4, we will describe how to generalize gray changes for arbitrary saturations based on hue pairs.

The second series of observations was aimed at discovering the gray-equivalence of the saturation increase. We investigated the effect of the saturation increase for all the hues of Table 6.2 (A=10, 20, 30, 40, 50, 60, 70) at different constant luminance levels (V=45, 65, 85). As above, all results were scaled to a maximum value of 10.

We interpolated between the seven selected hues using the maximum absolute solid-color saturation values of the 48 Coloroid pages. Thereafter, we applied relative saturations at every hue and luminance level. Relative saturation is defined so that it has a maximum value of 5 at the Coloroid gamut border [Nemc04]. The relative saturation for an arbitrary color is

	A	λ	φ
1	10	570.836	58.040
2	20	582.640	32.898
3	30	602.717	5.533
4	40	-504.836	-46.209
5	50	450.000	-116.628
6	60	490.371	-174.503
7	70	536.295	103.890

Table 6.1: Definition of the seven basic Coloroid hues

A/A	10	20	30	40	50	60	70
10	**0.0**	-2.5	-5.0	-7.0	-10.0	-5.0	-2.0
20	2.5	**0.0**	-2.5	-5.0	-8.0	-3.0	1.5
30	5.0	2.5	**0.0**	-3.0	-5.0	-3.0	3.5
40	7.0	5.0	3.0	**0.0**	-2.5	1.0	4.0
50	10.0	8.0	5.0	2.5	**0.0**	4.0	8.5
60	5.0	3.0	3.0	-1.0	-4.0	**0.0**	3.0
70	2.0	-1.5	-3.5	-4.0	-8.5	-3.0	**0.0**

Table 6.2: Relative gray-equivalent differences of the basic hue pairs

obtained by trilinear interpolation between the Coloroid limit-colors, taking proper account of the zero saturation of the black and white points.

The gray changes equivalent to hue and saturation differences are independently scaled on relative scales. We made 40 additional experiments, changing two attributes simultaneously to calibrate the relative scales to each other using linear regression. Two color pairs from this set of observations are illustrated below (where Δ-gray is the observed equivalent absolute gray difference):

Δ-gray$=-1.0$ (A1= 70.0, T1= 15.0, V1= 67.0) (A2= 24.0, T2= 15.0, V2= 67.0)
Δ-gray$=+4.0$ (A1= 30.0, T1= 32.0, V1= 47.0) (A2= 50.0, T2= 32.0, V2= 47.0)

6.3.4 The Coloroid-Based Formula

The perceptually-based Coloroid gray-gradient formula is central to our method. Unfortunately, it cannot be given in a closed form, since it consists of tables of observations and appropriate interpolation rules. The structure of the formula and the meanings of the terms involved are described below.

Bilinear interpolation is used since the relative gray-equivalence of pure hue changes is only given for 7 basic hues (see Tables 6.1 and 6.2). Therefore we linearly interpolate φ values to obtain the hues of intermediate colors [Hunt92]. For an arbitrary hue-pair (A_1, A_2), this allows us to obtain an H value in the interval $[-10, 10]$. The hue term is additive in this model and depends sub-linearly on the saturation. In the gradient term, H is computed

using a weight factor of the following form:

$$h(A_1, T_1, A_2, T_2) = w_h \cdot H(A_1, A_2) \cdot \sqrt{u(T_{1\text{rel}}) \cdot u(T_{2\text{rel}})}, \tag{6.3}$$

where T_{rel} is the relative saturation scaled to [0,5] for every hue plane and at every luminance level, computed from the maximum solid color saturation. This equation contains the geometrical mean of the two u-factors, and will be zero if at least one of the two colors is neutral. The $u(x)$ term, where $x = 2 \cdot T_{\text{rel}}$, is defined as follows:
$u(x) = 0.5 \cdot x$, iff $x < 0.5$, and
$u(x) = \sqrt{x} - 0.5$, otherwise.

Determining a saturation dependent gray-equivalent conversion is more complicated, since both hue and luminance are involved. We have to evaluate the relative gray change for any arbitrary pair of colors caused by a saturation change. We made observations for the 7 basic hues (Table 6.1) using the perceptually uniform Coloroid V values, at luminance levels of 45, 65 and 85. At the black and white points the change is zero (i.e. at $V = 0$ and $V = 100$). The suggested gray change for a color (A,T,V) due to the saturation term is also scaled to the interval $[-10, 10]$, but with a different weighting. We made additional observations to fit these two different scalings to each other. The effect of an increase in saturation may result in the positive or negative gray changes for a fixed hue and luminance. We use the φ values of the 7 basic hues and all the data for the most saturated solid colors [Nemc04], which are versions of the 48 limit-colors, and the 5 luminance levels mentioned above. Let us denote the gray change corresponding to the saturation of one color as $S(A, T, V)$. To compute S, we must first obtain T_{rel}, as we did for hue. Then we apply a trilinear interpolation using the neighboring φ, T_{rel} and the V values. For two colors, the signed gray change can now be obtained in the form[1]:

$$S(A_1, T_1, V_1, A_2, T_2, V_2) = w_s \cdot [S(A_2, T_2, V_2) - S(A_1, T_1, V_1)]. \tag{6.4}$$

The apparent part of the gray gradient is the unweighted luminance difference:

$$dL(L_1, L_2) = L_2 - L_1. \tag{6.5}$$

The color difference (gradient) can then be obtained by summing the luminance (6.5), saturation (6.4) and hue (6.3) formulas:

$$\Delta_{1,2} = dL(L_1, L_2) + S(A_1, T_1, V_1, A_2, T_2, V_2) + h(A_1, T_1, A_2, T_2). \tag{6.6}$$

6.4 Gradient Inconsistency Correction

6.4.1 The Basic Idea

Our adaptive color-to-grayscale conversion consists of three steps, as illustrated in Figure 6.5. In the first step, we regard the color and luminance contrasts as a gradient field which we construct using the formulas described in the previous sections. Then, instead of using a Poisson solver or conjugate gradient method, which work with density values, we correct the gradi-

[1]The database used to obtain Equation (6.4) can be downloaded from the web pages accompanying the thesis: `http://www.cgg.cvut.cz/~cadikm/diss`.

ent field using new fast and effective *gradient inconsistency correction method*, which uses an orthogonal projection technique to find the nearest consistent gradient field. Finally, we integrate the corrected gradient field and transform the values to the display range, which yields the grayscale image.

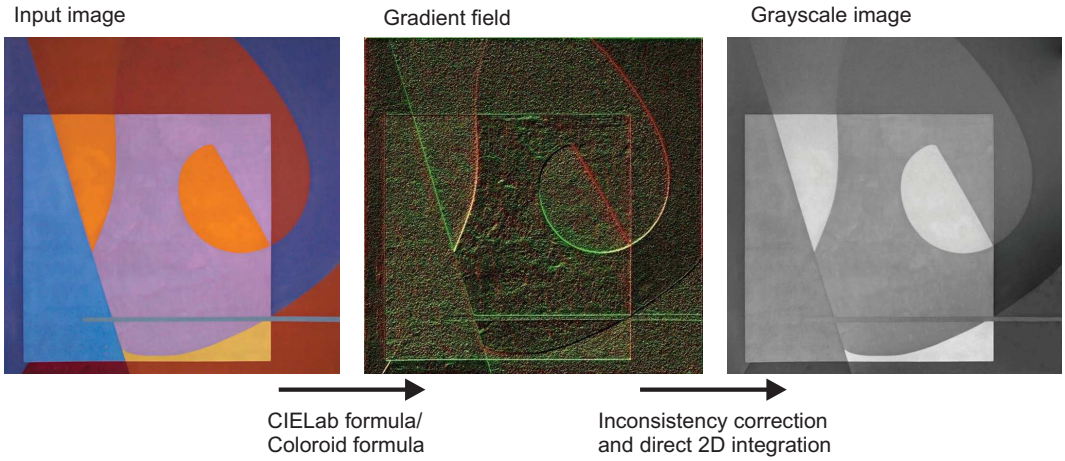


Figure 6.5: The three steps of the proposed adaptive color-to-grayscale conversion.

6.4.2 Gradient Inconsistency Correction Method

Gradient domain imaging methods generally change the original gradient field of an image, or generate an artificial gradient field from a set of images. The key issue in that approach is the backward transformation, which is the search for an image with a prescribed gradient field. In general, there is no exact solution to this problem and furthermore, the set of manipulated artificial gradient vectors is not a consistent gradient field, we cannot obtain the unknown image by two-dimensional integration (see Figure 6.6 – center).

Existing methods try to answer the following question: *which image has the nearest gradient field to the given inconsistent one?* The well-known and widely used multigrid Poisson solver, the FFT method, and a range of iterative methods minimize the sum of elementary quadratic error terms containing the finite difference of unknown pixels and the components of the gradient. Perhaps the most efficient and elegant technique is the conjugate gradient method with locally adapted hierarchical basis preconditioning [Szel06].

We approach the problem of an inconsistent gradient field with a new question: *what is the nearest consistent gradient field to an existing inconsistent one?* If we can find the nearest consistent gradients, the image can then be obtained by a simple two-dimensional integration requiring only one addition per pixel (see Figure 6.6 – right).

In the classical methods the unknowns are the pixel luminance values. Our new approach uses twice as many unknowns, namely the x and y components of each gradient vector. Consistency has a simple pictorial meaning: along a loop that starts and ends at any given pixel, the total gradient changes must be zero (see Figure 6.7 – left). Thus, an equation in terms of 4 gradient components can be constructed at every pixel. For an image of resolution $N \times M$, the total number of these equations is equal to the number of elementary loops

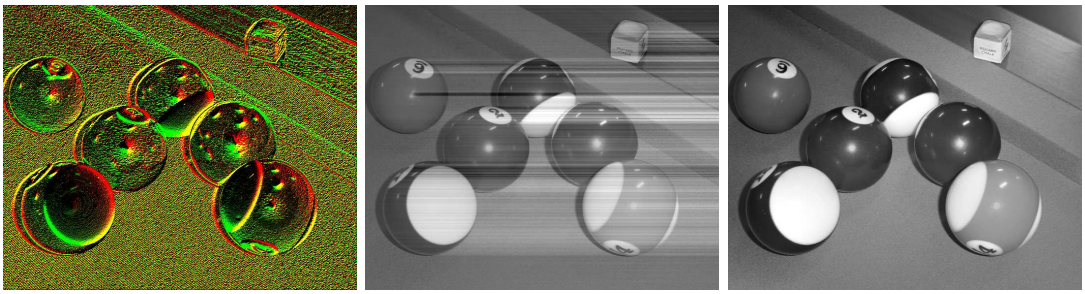


Figure 6.6: Left: inconsistent gradient field. Center: the result of direct integration of an inconsistent gradient field. Right: direct integration of a gradient field after inconsistency correction.

$((N-1)\cdot(M-1))$. The number of unknown gradient terms is $L=(N-1)\cdot M+(M-1)\cdot N$, which is approximately $2\cdot N\cdot M$.

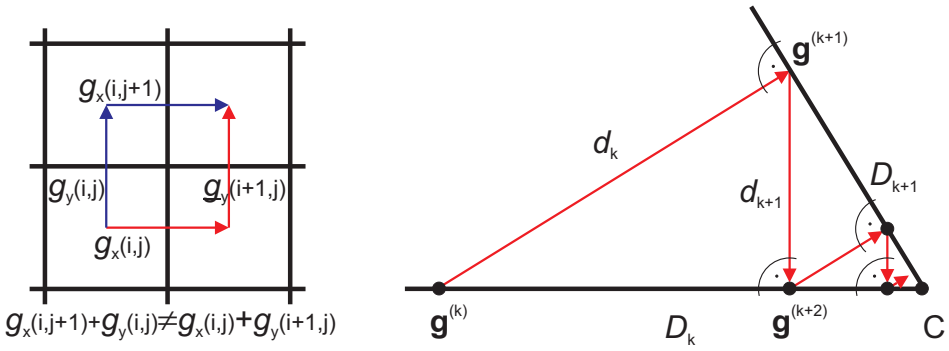


Figure 6.7: Left: the sum of the gradients around a pixel should be zero, but this is not in the case for an inconsistent gradient field (g denotes the gradient). Right: a simplified 2D example of orthogonal projection ($\mathbf{g}$ denotes the gradient field). Point $\mathbf{C}$ corresponds to a consistent gradient field, that we approach iteratively through $\mathbf{g}^{(k)}$, $\mathbf{g}^{(k+1)}$, $\mathbf{g}^{(k+2)}$, etc. using alternating orthogonal projections. D_k is the Euclidean distance from $\mathbf{g}^{(k)}$ to $\mathbf{C}$ and d_k is the length of the projection.

The possible inconsistent gradient terms can be described in a space of $L\approx N\cdot M$ dimensions, while the search for the nearest consistent field is conducted in the $N\cdot M$-dimensional linear subspace of consistent gradient fields. The Euclidean metric is used, as it provides the most natural notion of a 'nearest point'. In a higher dimension the problem is similar to the search for the nearest point on a line or plane to a given point in three-dimensional space. Summing the appropriate set of elementary equations with 4 gradient terms, we obtain the equations of a set of arbitrary closed curves. On all of these curves, or on all closed 'Manhattan lines' containing vertical and horizontal elementary intervals, the sum of the gradients has to be zero.

Almost all other gradient methods face the problem of the contradiction of gradients at different resolution levels. For instance, in their tone mapping method, Fattal et al. [Fatt02] used a multi-scale technique to construct an artificial gradient field. However, this approach

also changes large parts of the image which have a low dynamic range, although these regions should remain invariant. Gooch et al. [Gooc05] basically compared every pixel to every other pixel, in $O(N^4)$ time, so as to avoid the problem of contradictors resolutions to obtain a satisfactory overall appearance, despite a highly inconsistent gradient field. Our new consistency-correction method with simple 1-neighbor gradients also gives an attractive appearance, while implicitly solving the problem of contradictory resolutions in a new way.

6.4.3 The Algorithm

To find the nearest point in a higher-dimensional space, or to project a point orthogonally in a sub-space, generally requires a time-consuming algorithm. In a lower dimension, the Gram-Schmidt orthogonalization gives a closed-form solution. Fortunately, for sparse matrix problems an iterative method tends to the nearest point with very simple elementary projection steps. Figure 6.7 (left) shows the pixel (middle point) and two different ways around the pixel: the total changes to the field along these two paths have to be the same. After image manipulation, or for an artificially prescribed gradient field, this consistency does not hold. The sum of the gradients (E) on the closed curve containing the red and blue paths is defined as:

$$E_{ij} = g_x(i,j) + g_y(i+1,j) - g_x(i,j+1) - g_y(i,j), \tag{6.7}$$

where i, j are the coordinates of the pixel, g_x and g_y are horizontal and vertical gradients, respectively. We will refer to the red and blue closed paths in Figure 6.7 (left) as elementary gradient loops. For inconsistent loops, $|E_{ij}|$ is non-zero, i.e. $|E_{ij}| \geq \varepsilon > 0$. We have to change the values of the field encountered along all the elementary gradient loops, until every E_{ij} converges to zero. In the approximately $(2 \cdot N \cdot M)$-dimensional space of gradient components, the set of all the elementary gradient loops defines $((N-1) \cdot (M-1))$ equations, each of which contains only 4 non-zero coefficients, according to Equation (6.7).

Using the vector notation $\mathbf{N}_{ij} \cdot \mathbf{g} = E_{ij}$, where $\mathbf{g}$ denotes the vector describing the whole gradient field and the normal vector $\mathbf{N}_{ij} = (0, .., 0, +1, +1, -1, -1, 0, .., 0)$ contains 4 non-zero elements in the space of approximately $(2 \cdot N \cdot M)$ dimensions that is occupied by the gradient vectors. The equations of the elementary loops are:

$$\mathbf{N}_{ij} \cdot \mathbf{g} = 0, \qquad (i = 1, ..., N-1; j = 1, ..., M-1). \tag{6.8}$$

Therefore, the fundamental step of gradient inconsistency correction method (see Algorithm 1) is:

$$\mathbf{g}^{(k+1)} = \mathbf{g}^{(k)} - \frac{1}{4} \cdot E_{ij} \cdot \mathbf{N}_{ij},$$

where $\mathbf{g}$ is the vector describing the whole gradient field, E_{ij} is the sum of gradients along the elementary loop at pixel ij, and $\mathbf{N}_{ij}$ is the normal vector corresponding to the elementary loop at pixel ij. The orthogonal projection method converges to the nearest point of a subspace, in our case to one of the consistent gradient fields. This subspace is the common part of all hyper-planes defined by the elementary gradient loop equations. If we select the equation with the largest error, and project the current gradients on to this hyper-plane in the direction of $\mathbf{N}$, we approach the nearest consistent point $\mathbf{C}$ (see Figure 6.7 – right). The new distance D_{k+1} can be expressed as a function of the old distance and the length d_k of

the projection, characterizing the local inconsistency:

$$D_{k+1}^2 = D_k^2 - d_k^2.$$

It is easy to prove that the distance D_k converges to zero at a geometrical rate, both in the cyclical and in the maximum error selection case (maximum error selection algorithm is described in the next paragraph and it is discussed in Section 6.5).

Algorithm 1: Inconsistency correction

correct (gradient_field *grad*, double ω, double ε) {
repeat
  $E_{\max}$=0;
  for $j = 1$ to M-1 **do**
    for $i = 1$ to N-1 **do**
      $E = grad.X[i][j] + grad.Y[i+1][j] - grad.X[i][j+1] - grad.Y[i][j]$;
      if $|E| > E_{\max}$ **then**
        $E_{\max} = |E|$;
      end if
      $s = 1/4 \cdot E \cdot \omega$;
      $grad.X[i][j] = -s + grad.X[i][j]$;
      $grad.Y[i+1][j] = -s + grad.Y[i+1][j]$;
      $grad.X[i][j+1] = s + grad.X[i][j+1]$;
      $grad.Y[i][j] = s + grad.Y[i][j]$;
    end for
  end for
until $E_{\max} < \varepsilon$
}

Algorithm 1 corrects the absolute error $|E|$ of the elementary gradient loops, as illustrated in Figure 6.7 (left). This correction is carried out *cyclically* over the image. A more efficient alternative approach (*maximum error selection*) is always to select and correct the elementary gradient loop having the current maximum error $|E|_{\max}$, see Figure 6.11 and the discussion in Section 6.5. On the other hand, the cyclical correction of the elementary gradient loops does not require a data structure to store the error values and to find the maximum error, i.e. it is simpler to implement.

The standard orthogonal projection method would require the parameter ω in Algorithm 1 to be set to 1. A further possibility is to use *over-projection* by selecting ω such that $1 < \omega < 2$. This is locally less efficient, but significantly increases the overall rate of convergence. In this context, $\omega \in [1.6, 1.999]$ is convenient for most images, while $\omega = 1$ corresponds to the original convergence rate (i.e. using classical orthogonal projection instead of over-projection). Our experience shows that $\omega = 1.99$ is usually the best choice giving approximate speedup of $7-15$ compared to the classical orthogonal projection. All the variations of Algorithm 1 perform the sequence of orthogonal projections shown in Figure 6.7 (right). Fortunately, in our problem nearly all the hyperplanes defined by the equations of the elementary loops are orthogonal, since they do not share any common variables. Only the neighboring gradient loops have a single common variable. The angle between their hyperplanes is equal to $\arccos(-1/4)$.

Now that we have a consistent gradient field, the final image can be constructed by simple *two-dimensional integration*, as shown in Algorithm 2. We believe the reader should be able to implement this simple but efficient method very easily.

Algorithm 2: Double integration

integrate (gradient_field *grad*, output_image *out*) {
$out[1][1] = 0$;
for $j = 1$ to M **do**
  if $j > 1$ **then**
    $out[1][j] = out[1][j-1] + grad.Y[1][j-1]$;
  end if
  for $i = 2$ to N **do**
    $out[i][j] = out[i-1][j] + grad.X[i-1][j]$;
  end for
end for
}

6.5 Results and Discussion

We have verified the performance of our new color-to-grayscale conversion on a large variety of color images and photographs. The method is particularly attractive for low-contrast, colorful images, as shown in Figure 6.1. For this kind of image, the simple CIE Y equivalent image has poor contrast (see Figure 6.1 – top right), making some important details invisible. Note, for instance, the appearance of the sun and its reflection. On the other hand, the adaptive method of Gooch et al. (see Figure 6.1 – bottom left) ensures a significantly better appearance. However, the high-frequency details are now missing, as well as some contrast effects (e.g. in the sky). Our simple CIELab formula produces an image that is rich in detail and local contrast (see Figure 6.1 – bottom right). Moreover, the overall appearance of this image is superior to that achieved by any of the competing approaches represented in this figure. The time required to compute the conversion using the method by Gooch et al. was more than 10 minutes, whereas our method took approximately 0.1 seconds. The processing time of a color-to-grayscale conversion using our approach is approximately 5 to 10 seconds per megapixel.

In Figure 6.2, we compare our results with previous approaches (method by Gooch et al. (Color2gray) [Gooc05] and the approach of Rasche et al. [Rasc05]). Close-ups demonstrate problems of color quantization used in Rasche's approach, since it results in undesirable quantization artifacts (Figure 6.2 – bottom right). Furthermore, while our result (Figure 6.2 – top right) is obtained in a fraction of second, Gooch's Color2gray (Figure 6.2 – bottom left) consumed more than seven hours.

Figure 6.8 shows different color-to-grayscale conversions applied to a test image containing isoluminant colors. The classical approach (CIE Y) results in constant luminance (see Figure 6.8 – center), while our approach transforms the chrominance differences into clearly noticeable grayscale differences (see Figure 6.8 – right), but still preserves the moderate contrast of the original color image. In Figure 6.9, we compare the results obtained using the proposed CIELab and Coloroid formulas. Even though the differences between the CIE Y,

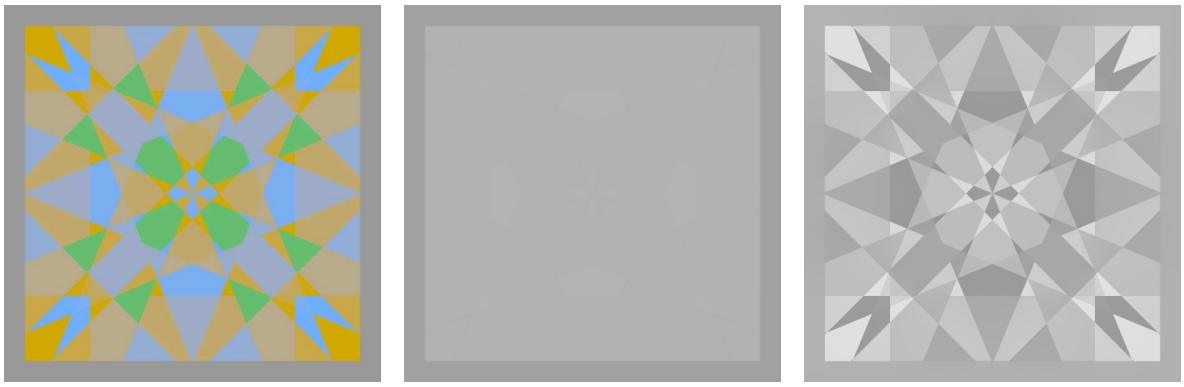


Figure 6.8: An artificial isoluminant image. Left: original color image. Center: CIE Y equivalent. Right: the result of our adaptive color-to-grayscale conversion (CIELab formula).

CIELab and Coloroid results are quite subtle, there are noticeable variations in the reproduction of several patches of color from the original image. We believe that the Coloroid result (bottom right) is closest to the human perception of the original color image (top left).

Further examples are provided in Figure 6.10. The images in the top row exhibit an obvious improvement in the amount of detail visible in the sky area. The sun is clearly delineated in our result, but it almost disappears in the classical grayscale conversion. The examples in the lower four rows show the enhanced visibility of many features of the original color images (e.g. the blossom of the plant, the form in the abstract painting, and the colored stripes on the cookers). In some cases, the result of the classical conversion is satisfying, and represents a credible alternative to our approach. But, for many images, our method distinctly outperforms the classical CIE Y formula.

We implemented and tested two different versions of the iterative algorithm that we use to obtain the nearest consistent gradient field. The first was the *cyclical iteration algorithm*, which cyclically corrects elementary gradient loops. The second – *maximum error selection*, a Fibonacci heap-based algorithm selects the most inconsistent elementary loop at each iteration. This algorithm uses a version of the Fibonacci heap [Fred87] as auxiliary data structure, which slightly increases the computational cost, but the algorithm is in particular effective in handling the most inconsistent areas. In Figure 6.11 the problematic areas are marked with a darker color, and the distribution of the computational effort is focused on these darker areas, drastically reducing the number of iterations, compared to the cyclical algorithm. In processing images containing between 0.5 and 1.5 million pixels, the Fibonacci heap algorithm ran on average 15 times faster than the cyclical algorithm: depending on the magnitude and spatial distribution of the gradient inconsistency, the speedup factors varied between 6 and 54.

6.6 Summary

We have investigated local adaptive approaches to the color-to-grayscale conversion. The human vision system does not convert color differences directly to signed luminance gradients, and so this kind of transformation is essentially artificial and no perceptually ideal solution exists. In this chapter we presented two novel and efficient ways to construct a gradient field

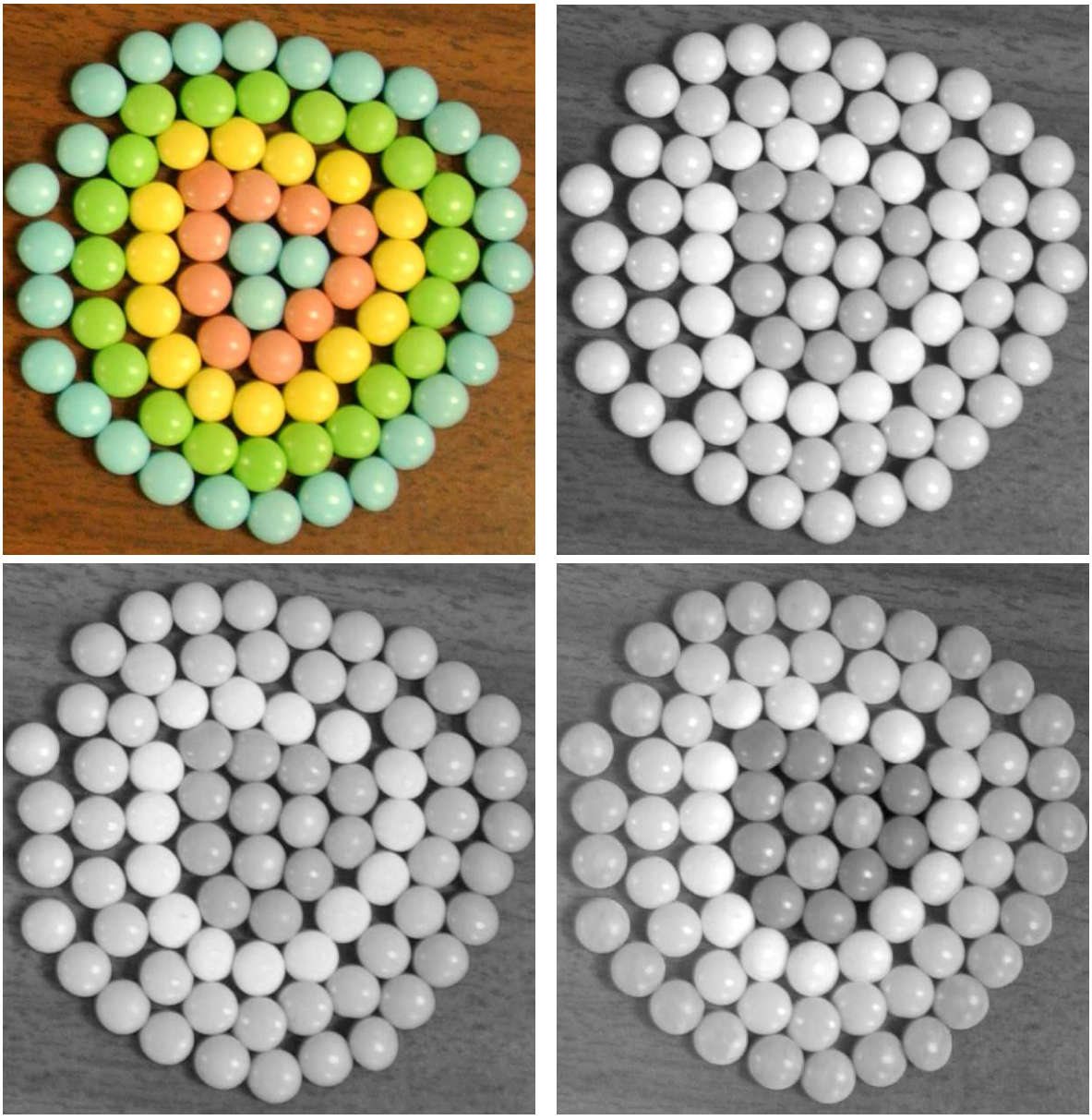


Figure 6.9: Comparison of the results from the new method obtained using two different formulas. Top left: original color image. Top right: CIE Y equivalent. Bottom left: our result using the CIELab formula. Bottom right: our result using the Coloroid formula. *Color image courtesy of Karl Rasche.*

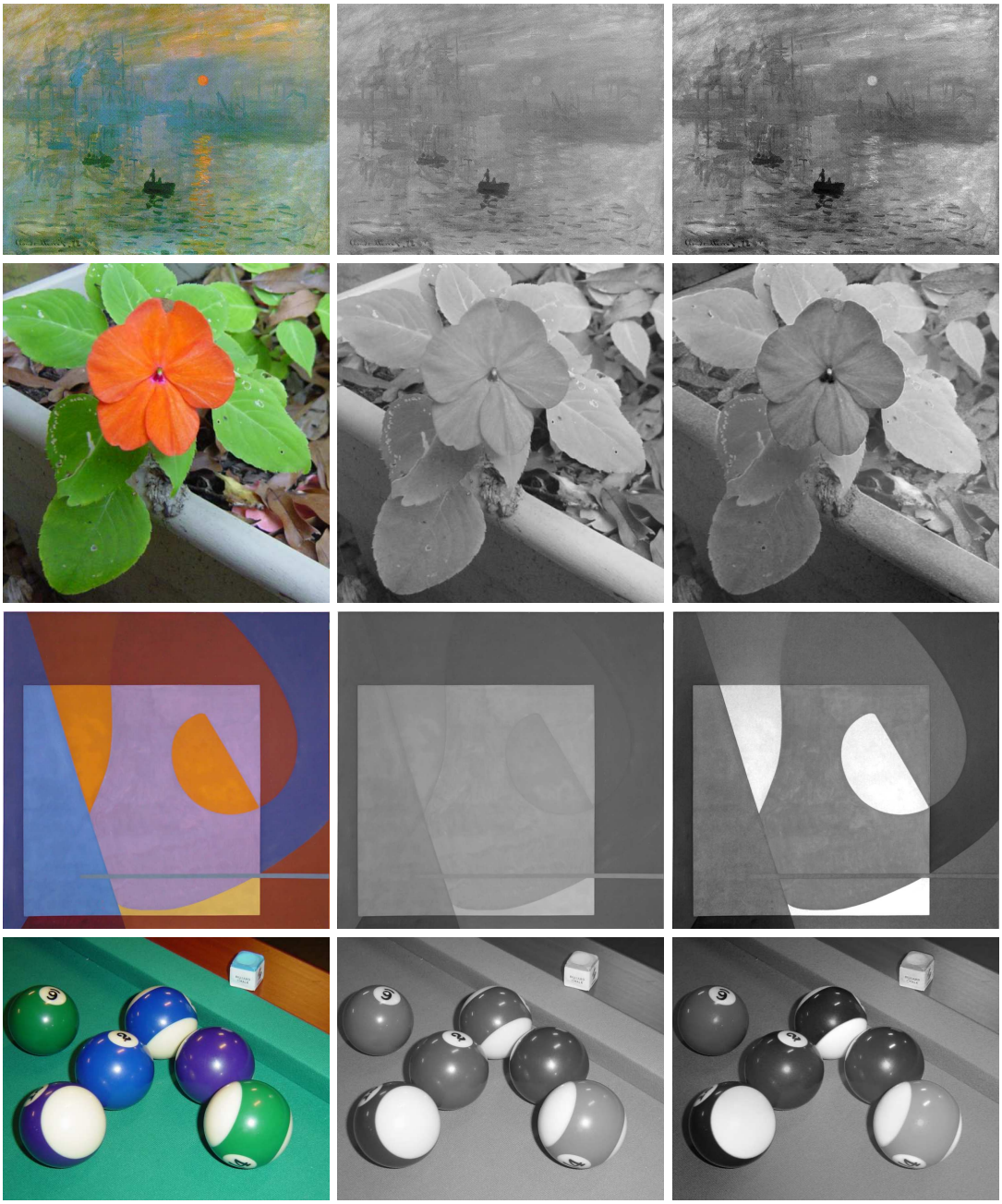


Figure 6.10: Left column: original color image. Middle column: CIE Y equivalent. Right column: result of our adaptive color-to-grayscale conversion (Coloroid formula). *Color images courtesy of Amy Gooch, Karl Rasche, and Antal Nemcsics.*

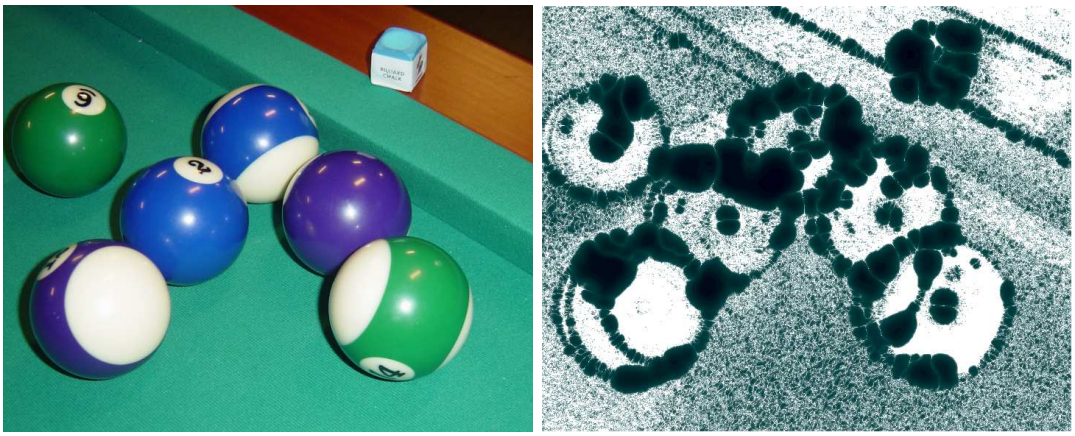


Figure 6.11: Maximum error selection using the Fibonacci heap. Left: input color image. Right: maximum error selection, the darker color shows frequently corrected areas of inconsistent gradients. An equivalent image for the cyclical method would all be dark gray.

from a color image, leading to a grayscale image of pleasing appearance. The first formula operates in the CIELab color space, while the second uses the Coloroid system, which is based on a strong experimental background. In this second approach we formulate, and obtain, perceptually justifiable gray gradients equivalent to different color attributes, using efficient experimental arrangements.

We have introduced a new gradient correction method for solving the gradient field inconsistency problem. This method differs from all existing techniques (Poisson, conjugate gradient, FFT, etc.), which work with luminance values. Our technique works with unknown gradients to obtain a consistent gradient field. Given an inconsistent gradient field, it can find the nearest consistent field in the linear subspace of consistent gradient fields. Having obtained a consistent gradient field, the final image can be produced by two-dimensional integration. The method has linear complexity in the number of pixels, which makes it suitable for high-resolution images.

7 Conclusions

"What we play is life."

(Louis Armstrong)

In this thesis, we presented experimental subjective evaluations of image quality assessment metrics and High Dynamic Range (HDR) tone mapping methods. Furthermore, we proposed new human perception motivated approaches to HDR tone mapping and to color to grayscale conversion. This chapter summarizes our contributions and discusses ongoing and potential directions of future work.

7.1 Summary of Results

We concerned ourselves with image quality assessment metrics, HDR tone mapping methods, and color to grayscale conversions in this thesis. In all these areas of digital imaging, exploiting the knowledge of the human visual system (HVS) and visual perception is particularly important. Since methods we deal with usually involve a model of HVS, it is very advantageous and necessary to conduct experimental subjective analyses of these methods to validate and evaluate them properly. Besides exposing strengths and weaknesses of inquired methods, the experimental evaluations also attain a deeper knowledge of examined fields which can advance current state of the art and which can furthermore result in proposals of new approaches.

7.1.1 Perceptual Evaluation of Image Quality Metrics

In Chapter 3, we focused on methods for image quality assessment. We presented results of the *evaluation of two principal approaches to the image quality assessment*. We evaluated the error sensitivity based approach and the structural similarity based approach using the representative models of each group: the Visible Differences Predictor (VDP) and the Structural SIMilarity index (SSIM). Inputs to the subjective experiments included *uniformly compressed* images and images compressed non-uniformly with *regions of interest* (ROI). We assessed the performance of the selected models using subjective opinion scores obtained in both the uniform and the ROI compression experiments. We discussed obtained results, and pointed out the similarities and differences between the methods. Besides, we presented a survey of published models of the two inquired groups.

Our results imply that the structural similarity approach overcomes the traditional error sensitivity based approach. Since the implementation of a structural similarity model is typically more straightforward than the implementation of an error sensitivity based model, we concluded that the structural similarity is a significant alternative to the thoroughly verified error

sensitivity. Nevertheless, both inquired approaches perform poorly in the assessment of ROI compressed images.

7.1.2 Perceptual Evaluation of Tone Mapping Methods

In Chapter 4, we presented a study about the effect of *basic image attributes* in the HDR tone mapping. We proposed a scheme of relationships between these attributes and we presented results of *subjective psychophysical experiments* that we have performed to prove the proposed relationship scheme. Furthermore, we presented an *evaluation of existing tone mapping methods* with regard to these attributes. Differently from previous evaluations [Drag03b, Yosh05, Kuan06b], we adopted both the direct (with reference) comparison of tone mapped images to the real scenes, and the self-contained ranking without real references. This enabled us to confront the results from these two experiments.

An interesting and important result of the two different testing methodologies used (rating with reference and ranking without reference) is that almost all of the studied image quality attributes can be evaluated without comparison to a real HDR reference. With 14 inquired methods, and three typical real-world HDR scenes, the presented study is one of the most comprehensive evaluations of tone mapping methods. Another important outcome of our study is that the proper global part of a tone mapping method is essential to obtain good perceptual results for typical real world scenes. Moreover, our effort is also useful to get into the tone mapping field or when implementing a tone mapping method, and it also sets the stage for well-founded quality comparisons between tone mapping methods.

7.1.3 Hybrid Approach to Tone Mapping

In Chapter 5, we proposed a novel simple yet powerful *general hybrid approach to tone mapping* issue. In our approach, we combine outputs of arbitrary global and local tone mapping (TM) methods. Motivated by the results of conducted subjective experiments that have shown high importance of preservation of global image attributes, we apply the global method first to reproduce overall image attributes correctly. At the same time, we construct an enhancement map to guide a local operator to the critical areas of an image that deserve enhancement. We do not invent another complex TM method, but we rather propose a general framework that utilizes already known ideas and combines existing and potentially forthcoming methods to obtain perceptually justifiable results. Based on the choice of involved methods and on the manner of construction of the enhancement map, we show that our approach is general and it can be easily tailored to miscellaneous goals of tone mapping. An implementation of proposed hybrid tone mapping produces plausible results, it is easy to implement, fast to compute and it is comfortably scalable. These qualities nominate our approach for utilization in perceptually oriented and time-critical HDR applications like HDR image viewers, modern computer games, interactive visualizations, mobile device HDR applications, etc.

Later, Artusi et al. [Artu07b] published a concept of *selective tone mapper* which uses a model of visual attention to direct local TM methods to perceptually important parts of an image, while a global TM method is used for the remainder. Artusi et al. propose a generic GPU-aware implementation that can utilize any existing GPU TM methods. Factually, this work is an implementation and verification of the hybrid approach as was proposed in Chapter 5:

authors utilize the Canny edge detection in the construction of the enhancement map (in their notation: important areas) and then, in accord with the hybrid approach, they apply local TM method just to these important parts of the image. Since the reported computational and quality performance results are very promising they encouraged authors of the selective tone mapper to patent their work [Artu07a].

7.1.4 Perceptually-Plausible Conversion of Color Images to Grayscale

In Chapter 6, we proposed a *new perceptually-plausible conversion of color images to grayscale*. As the method operates in gradient domain, we proposed two novel and efficient ways to construct a gradient field from a color image. The first approach operates in the CIELab color space, while the second uses the Coloroid system, which is based on a strong experimental background. In this second approach we formulate, and obtain, perceptually justifiable gray gradients equivalent to different color attributes, by means of efficient experimental arrangements.

Using one of the new approaches, we obtain an inconsistent gradient field from a color image. As the inconsistent gradient field does not correspond to any real image, the problem is how to transform it to an output grayscale image. For solving this gradient field inconsistency problem, we have introduced a *new gradient inconsistency correction method*. This method works with unknown gradients to obtain a consistent gradient field. Given an inconsistent gradient field, it can find the nearest consistent field in the linear subspace of consistent gradient fields. Having obtained a consistent gradient field, the final image can then be produced by two-dimensional integration. The method has linear complexity in the number of pixels, which makes it suitable for high-resolution images.

It is advantageous to apply the proposed method in all cases where color images have to be converted to grayscale for reproduction or for subsequent processing. The important benefit is that much more color features and contrasts are saved in comparison to usually applied conversion taking luminance values. Our experiments show that in comparison to existing approaches [Gooc05, Rasc05], the new method produces compelling results and it is much faster to compute.

7.2 Suggestions for Further Research

This section describes suggestions for further research on the topics discussed in the thesis. As we have seen, the utilization of human perception properties is profitable in various imaging applications, and finding novel unforeseen uses of HVS models in computer graphics and digital imaging is an interesting and valuable forthcoming work. To succeed, computer graphics researchers have to actively and closely communicate with vision scientists, psychologists, physiologists, and other researchers.

7.2.1 Perceptual Evaluation of Image Quality Metrics

Besides the two evaluated approaches to the image quality assessment, several ad-hoc non-perceptual methods exist [Shei02, Marz02]. These methods are typically based on the assumption that we know the types of image distortions in advance. It would be interesting to

inquire such methods in future evaluations. Moreover, it would be exciting to perform deeper study involving the *majority of approaches surveyed* in the beginning of Chapter 3. Then, one could assess the bottlenecks and powerful features of individual image quality metrics as well as similarities and differences between them in a finer detail.

According to evident specific benefits of image quality assessment and HDR tone mapping (i.e. the mature methodology and numerous published approaches in the case of IQA, and a real-world dynamic range of luminances and modeling of advanced spatial visual phenomena in HDR tone mapping case), IQA and HDR tone mapping are naturally starting to overlap. Both IQA and tone mapping methods try to predict the appearance of an image (scene) to the observer. IQA methods then use this information to assess the quality of the input image, while TM methods use it for perceptual luminance range reduction. Originally, IQM models were strictly LDR, but these times are over now, notice e.g. an HDR version of the visible differences predictor [Mant05]. Furthermore, tone mapping method called iCAM [Kuan06a] is capable to model image appearance and to predict the differences as observed by subjects. It seems so that the modeling of *image appearance* that takes into account an HDR image as a whole is the common future of image quality assessment and HDR tone mapping. Sophisticated image appearance models could be utilized in a broad range of imaging applications, such as IQM, visual image difference, HDR TM, and prediction of visual phenomena.

7.2.2 Perceptual Evaluation of Tone Mapping Methods

The next question to answer in the evaluation of tone mapping methods using image attributes is how to *numerically assess the quality of reproduction of particular image attributes*. Although some approaches operating on LDR images were proposed in the literature [Jans01, Matk05], the HDR field deserves further investigation and perceptual verification. It is necessary to conduct consequential subjective experiments targeted on individual image attributes to be finally able to computationally assess the overall quality of images produced by tone mapping methods.

Furthermore, we are not aware of any experimental *evaluation of time-dependent tone mapping methods*. This is evidently harder task than the evaluation of tone mapping methods operating on static images (especially the testing with a real-world reference), however the results would be appreciated in many areas of imaging, including game and movie industry, interactive visualizations, HDR players, etc.

7.2.3 Hybrid Approach to Tone Mapping

Proposed hybrid approach to tone mapping can be *parametrized* instead of the current solution using luminance percentiles (described in Section 5.3.1). Specifically, the construction of an enhancement map (Equation (5.1) on page 80) and the blending of local and global methods (Equation (5.2) on page 80) can easily be modified by user parameters. However, these parameters have to be carefully assessed in a subjective experiment, because it is of particular importance to preserve the perceptually plausible balance between the applied global and local methods.

The perception of images depends partially on the contents (semantics) of the image or the captured scene. Therefore every, even a subtle modification of an image can affect the quality

of reproduction of an image attribute (in both positive and negative sense). In the future, we will conduct subjective perceptual experiments to uncover and quantify the *effect of particular local enhancement method* (in relation to the manner of enhancement map construction) *on the quality of reproduction of image attributes*.

7.2.4 Perceptually-Plausible Conversion of Color Images to Grayscale

In future, we plan to further assess the performance of the proposed color to grayscale conversion algorithm, and we also intend to undertake more extensive experimentation, including *subjective testing on photographers and painters*, which is the input that we believe will lead to the best conversion. Moreover, we will explore a *multiscale solution* with the aim of making the method work in real time even for high resolution images.

An important and not really solved problem in color to grayscale conversions is an endorsement of the *temporal coherence*. When converting a color animation or series of color images sets (such as cartoons) into grayscale, several problems and possibilities arise: maintaining local consistence of gray levels between frames and/or global consistence throughout the animation, avoiding temporal artifacts, utilizing the temporal coherence for acceleration of the conversion, etc. Forthcoming method by Smith et al. [Smit08] applies global "absolute" mapping to overcome some of the mentioned issues. However the local adjustment which they subsequently apply may imply other existing problems.

As existing color to grayscale conversions operate on ordinary (LDR) images, another interesting improvement would be a perceptually plausible *direct conversion of HDR color images* to LDR grayscale images. Currently, one can solve this issue by applying a tone mapping method and a color to grayscale conversion successively. However, this approach may vanish important luminance and/or chrominance details. Direct conversion of color HDR images to grayscale LDR images therefore potentially grants better results.

Finally, we are currently investigating how to utilize *optical illusions*, for instance the lightness illusions [Adel99], e.g. the simultaneous contrast effect (see Figure 7.1) for improvement of the grayscale rendition.



Figure 7.1: The simultaneous contrast effect. The two inner squares are the same level of gray, but they appear to have different shades due to their backgrounds.

8 Bibliography

"All books are divisible into two classes: the books of the hour, and the books of all time."

(John Ruskin)

[Adel99] E. H. Adelson. Lightness perception and lightness illusions. In M. Gazzaniga, Ed., *The Cognitive Neurosciences*, pp. 339– 351, MIT Press, Cambridge, MA, 1999.

[Akyu07] A. O. Akyüz, R. Fleming, B. E. Riecke, E. Reinhard, and H. H. Bülthoff. Do HDR Displays Support LDR Content? A Psychophysical Evaluation. *ACM Transactions on Graphics (Proceedings of SIGGRAPH)*, Vol. 26, No. 3, p. 38, 2007.

[Ande03] M. J. Anderson and C. ter Braak. Permutation tests for multi-factorial analysis of variance. *Journal of Statistical Computation and Simulation*, Vol. 73, pp. 85–113, 2003.

[Artu07a] A. Artusi, B. Roch, A. Chalmers, and Y. Chrysanthou. US Patent application number GB 0709392.5: Selective Tone Mapper. 2007. UK Patent Office.

[Artu07b] A. Artusi, B. Roch, Y. Chrysanthou, D. Michael, and A. Chalmers. Selective Tone Mapper. Tech. Rep. TR-07-05, University of Cyprus, 2007.

[Ashi02] M. Ashikhmin. A tone mapping algorithm for high contrast images. In *13th Eurographics Workshop on Rendering*, pp. 145–156, Eurographics Association, 2002.

[Ashi06] M. Ashikhmin and J. Goyal. A reality check for tone-mapping operators. *ACM Transactions on Applied Perception*, Vol. 3, No. 4, pp. 399–411, 2006.

[Bala04] R. Bala and R. Eschbach. Spatial Color-to-Grayscale Transform Preserving Chrominance Edge Information. In *Color Imaging Conference*, pp. 82–86, IS&T – The Society for Imaging Science and Technology, 2004.

[Blak69] C. Blakemore and F. W. Campbell. On the existence of neurons in the human visual system selectively responsive to the orientation and size of retinal images. *Journal of Physiology*, Vol. 203, pp. 237– 260, 1969.

[Brad99] A. P. Bradley. A Wavelet Visible Difference Predictor. *IEEE Transactions on Image Processing*, Vol. 8, No. 5, pp. 717– 730, 1999.

[Burt83] P. J. Burt and E. H. Adelson. The Laplacian Pyramid as a Compact Image Code. *IEEE Transactions on Communications*, Vol. Com–31, pp. 532– 540, 1983.

[Cadi05] M. Čadík and P. Slavík. The Naturalness of Reproduced High Dynamic Range Images. In *Proceedings of the Ninth International Conference on Information Visualisation*, pp. 920–925, IEEE Computer Society, Los Alamitos, 2005.

[Cadi06] M. Čadík, M. Wimmer, L. Neumann, and A. Artusi. Image Attributes and Quality for Evaluation of Tone Mapping Operators. In *Proceedings of Pacific Graphics 2006*, pp. 35– 44, National Taiwan University Press, Taipei, Taiwan, 2006.

[Cadi07] M. Čadík. Perception Motivated Hybrid Approach to Tone Mapping. In *Proceedings of WSCG (Full Papers)*, pp. 129– 136, University of West Bohemia, 2007.

[Cadi08] M. Čadík, M. Wimmer, L. Neumann, and A. Artusi. Evaluation of HDR Tone Mapping Methods Using Essential Perceptual Attributes. *Submitted to Computers & Graphics*, 2008.

[Cala03] A. J. Calabria and M. D. Fairchild. Perceived image contrast and observer preference I: The effects of lightness, chroma, and sharpness manipulations on contrast perception. *Journal of Imaging Science & Technology*, Vol. 47, pp. 479–493, 2003.

[Camp66] F. W. Campbell, J. J. Kulikowski, and J. Levinson. The effect of orientation on the visual resolution of gratings. In *Journal of Physiology*, pp. 427– 436, 1966.

[Chen06] G.-H. Chen, C.-L. Yang, and S.-L. Xie. Gradient-Based Structural Similarity for Image Quality Assessment. In *Proceedings of International Conference on Image Processing (ICIP)*, pp. 2929–2932, IEEE, 2006.

[Chiu93] K. Chiu, M. Herf, P. Shirley, S. Swamy, C. Wang, and K. Zimmerman. Spatially nonuniform scaling functions for high contrast images. In *Proceedings of Graphics Interface '93*, pp. 245– 253, 1993.

[Chou03] P. Choudhury and J. Tumblin. The trilateral filter for high contrast images and meshes. In *EGRW '03: Proceedings of the 14th Eurographics workshop on Rendering*, pp. 186–196, Eurographics Association, 2003.

[CIE81] CIE. *An Analytical Model for Describing the Influence of Lighting Parameters upon Visual Performance*. Vol. 1: Technical Foundations, CIE 19/2.1, International Organization for Standardization, 1981.

[Cign98] P. Cignoni, C. Rocchini, and R. Scopigno. Metro: measuring error on simplified surfaces. *Computer Graphics Forum*, Vol. 17, No. 2, 1998.

[Daly93] S. Daly. The Visible Differences Predictor: An Algorithm for the Assessment of Image Fidelity. In A. B. Watson, Ed., *Digital Images and Human Vision*, pp. 179–206, MIT Press, Cambridge, MA, 1993.

[Debe97] P. E. Debevec and J. Malik. Recovering High Dynamic Range Radiance Maps from Photographs. In T. Whitted, Ed., *SIGGRAPH '97 Proceedings of the 24th annual conference on Computer graphics and interactive techniques*, pp. 369–378, ACM SIGGRAPH, Addison Wesley, Aug. 1997.

[Devl02] K. Devlin, A. Chalmers, A. Wilkie, and W. Purgathofer. STAR: Tone Reproduction and Physically Based Spectral Rendering. In D. Fellner and R. Scopignio, Eds., *State of the Art Reports, Eurographics 2002*, pp. 101–123, The Eurographics Association, September 2002.

[Drag03a] F. Drago, K. Myszkowski, T. Annen, and N.Chiba. Adaptive Logarithmic Mapping For Displaying High Contrast Scenes. *Computer Graphics Forum*, Vol. 22, No. 3, 2003.

[Drag03b] F. Drago, W. L. Martens, K. Myszkowski, and H.-P. Seidel. Perceptual evaluation of tone mapping operators. In *GRAPH '03: Proceedings of the SIGGRAPH 2003 conference on Sketches & applications*, pp. 1–1, ACM Press, New York, NY, USA, 2003.

[Dura00] F. Durand and J. Dorsey. Interactive Tone Mapping. In *Proceedings of the Eurographics Workshop on Rendering*, pp. 219–230, Springer Verlag, 2000. Held in Brno, Czech Republic.

[Dura02] F. Durand and J. Dorsey. Fast bilateral filtering for the display of high-dynamic-range images. In *SIGGRAPH '02: Proceedings of the 29th annual conference on Computer graphics and interactive techniques*, pp. 257–266, ACM Press, 2002.

[Erik98] R. Eriksson, B. Andren, and K. Brunstrom. Modeling the perception of digital images: A performance study. In *IS&T/SPIE Conference on Human Vision and Electronic Imaging III, SPIE volume 3299*, pp. 88–97, San Jose, CA, January 1998.

[Fair02] M. D. Fairchild and G. M. Johnson. Meet iCAM: An Image Color Appearance Model. In *IS&T/SID 10th Color Imaging Conference*, pp. 33–38, VA: Society for Imaging Science & Technology, 2002.

[Fair03] M. D. Fairchild and G. M. Johnson. Image appearance modeling. In *SPIE/IS&T Electronic Imaging Conference*, pp. 149–160, SPIE, Bellingham WA, 2003.

[Fair04] M. D. Fairchild, G. M. Johnson, J. Kuang, and H. Yamaguchi. Image appearance modeling and high-dynamic-range image rendering. In *APGV '04: Proceedings of the 1st Symposium on Applied perception in graphics and visualization*, pp. 171–171, ACM, New York, NY, USA, 2004.

[Fair05] M. D. Fairchild. *Color Appearance Models*. Wiley-IS&T, Chichester, UK, 2nd Ed., 2005.

[Fair91] M. D. Fairchild and E. Pirrotta. Predicting the lightness of chromatic object colors using CIELAB. *Color Research and Application*, Vol. 16, pp. 385–393, 1991.

[Fair98] M. D. Fairchild. *Color Appearance Models*. Reading, Mass.: Addison-Wesley, 1998.

[Fatt02] R. Fattal, D. Lischinski, and M. Werman. Gradient domain high dynamic range compression. In *SIGGRAPH '02: Proceedings of the 29th annual conference on Computer graphics and interactive techniques*, pp. 249–256, ACM Press, 2002.

[Fedo97] E. Fedorovskaya, H. de Ridder, and F. Blommaert. Chroma variations and perceived quality of color images of natural scenes. *Color Research & Application*, Vol. 22, No. 2, pp. 96–110, 1997.

[Ferw96] J. A. Ferwerda, S. N. Pattanaik, P. Shirley, and D. P. Greenberg. A Model of Visual Adaptation for Realistic Image Synthesis. In *SIGGRAPH '96: Proceedings of the 23rd annual conference on Computer graphics and interactive techniques*, pp. 249–258, ACM, New York, NY, USA, 1996.

[Fial06] O. Fialka and M. Čadík. FFT and Convolution Performance in Image Filtering on GPU. In *Tenth International Conference on Information Visualisation*, pp. 609–614, IEEE Computer Society Press, Los Alamitos, 2006.

[Font01] L. da Fontoura Costa and R. M. Cesar. *Shape analysis and classification: theory and practise*. CRC Press, USA, 2001.

[Fred87] M. L. Fredman and R. E. Tarjan. Fibonacci heaps and their uses in improved network optimization algorithms. *Journal of the ACM*, Vol. 34, No. 3, pp. 596–615, 1987.

[Free91] W. T. Freeman and E. H. Adelson. The design and use of steerable filters. *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 13, No. 9, pp. 891– 906, 1991.

[Gerv84] M. J. Gervais, L. O. Harvey, and J. O. Roberts. Identification confusions among letters of the alphabet. In *Journal of Experimental Psychology: Human Perception and Performance*, pp. 655– 666, 1984.

[Gonz02] R. C. Gonzalez and R. E. Woods. *Digital Image Processing*. Prentice-Hall, New Jersey, 2nd Ed., 2002.

[Gooc05] A. A. Gooch, S. C. Olsen, J. Tumblin, and B. Gooch. Color2Gray: salience-preserving color removal. *ACM Transactions on Graphics*, Vol. 24, No. 3, pp. 634–639, 2005.

[Grun05] M. Grundland and N. A. Dodgson. The Decolorize Algorithm for Contrast Enhancing, Color to Grayscale Conversion. Tech. Rep. UCAM-CL-TR-649, University of Cambridge, 2005.

[Hube59] D. H. Hubel and T. N. Wiesel. Receptive fields of single neurons in the cat's striate cortex. In *Journal of Physiology*, pp. 574– 591, London, 1959.

[Hunt92] R. W. G. Hunt. *Measuring Colour*. Ellis Horwood Series in Applied Science and Industrial Technology, 2nd Ed., 1992.

[Jack97] W. B. Jackson, M. R. Said, D. A. Jared, J. O. Larimer, J. L. Gille, and J. Lubin. Evaluation of human vision models for predicting human observer performance. In H. L. Kundel, Ed., *Proceedings of SPIE Medical Imaging 1997: Image Perception*, pp. 64–73, Apr. 1997.

[Jans01] R. Janssen. *Computational Image Quality*. Society of Photo-Optical Instrumentation Engineers (SPIE), 2001.

[Jobs02] D. Jobson, Z. Rahman, and G. Woodell. The statistics of visual representation. In Z. Rahman, R. A. Schowengerdt, and S. E. Reichenbach, Eds., *Visual Information Processing XI*, pp. 25–35, 2002.

[Kend90] M. Kendall and J. D. Gibbons. *Rank Correlation Methods.* Oxford University Press, New York, NY, 5th Ed., 1990.

[Kraw05] G. Krawczyk, K. Myszkowski, and H.-P. Seidel. Perceptual effects in real-time tone mapping. In *SCCG '05: Proceedings of the 21st spring conference on Computer graphics*, pp. 195–202, ACM, New York, NY, USA, 2005.

[Kraw06] G. Krawczyk, K. Myszkowski, and H.-P. Seidel. Computational Model of Lightness Perception in High Dynamic Range Imaging. In B. E. Rogowitz, T. N. Pappas, and S. J. Daly, Eds., *Human Vision and Electronic Imaging XI, IS&T/SPIE's 18th Annual Symposium on Electronic Imaging (2006)*, pp. 1–12, 2006.

[Krug99] V. Kruger and G. Sommer. Affine Real-Time Face Tracking Using a Wavelet Network. In *RATFG-RTS '99: Proceedings of the International Workshop on Recognition, Analysis, and Tracking of Faces and Gestures in Real-Time Systems*, p. 141, IEEE Computer Society, Washington, DC, USA, 1999.

[Kuan04] J. Kuang, H. Yamaguchi, G. M. Johnson, and M. D. Fairchild. Testing HDR Image Rendering Algorithms. In *IS&T/SID 12th Color Imaging Conference*, pp. 315–320, 2004.

[Kuan05] J. Kuang, G. M. Johnson, and M. D. Fairchild. Image Preference Scaling for HDR Image Rendering. In *IS&T/SID 13th Color Imaging Conference*, pp. 8–13, Scottsdale, Arizona, 2005.

[Kuan06a] J. Kuang, G. M. Johnson, and M. D. Fairchild. iCAM for high-dynamic-range image rendering. In *APGV '06: Proceedings of the 3rd symposium on Applied perception in graphics and visualization*, pp. 151–151, ACM, New York, NY, USA, 2006.

[Kuan06b] J. Kuang, C. Liu, G. M. Johnson, and M. D. Fairchild. Evaluation of HDR Image Rendering Algorithms Using Real-World Scenes. In *International congress of imaging science, ICIS'06*, Rochester, NY, USA, 2006.

[Kuan07] J. Kuang, H. Yamaguchi, C. Liu, G. M. Johnson, and M. D. Fairchild. Evaluating HDR rendering algorithms. *ACM Transactions on Applied Perception*, Vol. 4, No. 2, p. 9, 2007.

[Lai00] Y. K. Lai and C.-C. J. Kuo. A Haar wavelet approach to compressed image quality measurement. *Journal of Visual Communication and Image Understanding*, Vol. 11, pp. 17– 40, 2000.

[Lars97] G. Ward Larson, H. Rushmeier, and C. Piatko. A Visibility Matching Tone Reproduction Operator for High Dynamic Range Scenes. *IEEE Transactions on Visualization and Computer Graphics*, Vol. 3, No. 4, pp. 291–306, 1997.

[Ledd04] P. Ledda, L. P. Santos, and A. Chalmers. A local model of eye adaptation for high dynamic range images. In *AFRIGRAPH '04: Proceedings of the 3rd international conference on Computer graphics, virtual reality, visualisation and interaction in Africa*, pp. 151–160, ACM Press, New York, NY, USA, 2004.

[Ledd05] P. Ledda, A. Chalmers, T. Troscianko, and H. Seetzen. Evaluation of Tone Mapping Operators using a High Dynamic Range Display. In *SIGGRAPH '05: Proceedings of the 32nd annual conference on Computer graphics and interactive techniques*, pp. 640–648, ACM Press, August 2005.

[Lehm78] W. Lehmacher and K. D. Wall. A new nonparametric approach to the comparison of K independent samples of response curves. *Biometrical Journal*, Vol. 20, pp. 261–273, 1978.

[Li98] B. Li, G. Meyer, and R. Klassen. A Comparison of Two Image Quality Models. In *SPIE Conference on Human Vision and Electronic Imaging III*, pp. 98–109, 1998.

[Lisc06] D. Lischinski, Z. Farbman, M. Uyttendaele, and R. Szeliski. Interactive local adjustment of tonal values. *SIGGRAPH '06: Proceedings of the 33rd annual conference on Computer graphics and interactive techniques*, Vol. 25, No. 3, pp. 646–653, 2006.

[Livi88] M. S. Livingstone and D. Hubel. Segregation of form, color, movement, and depth: Anatomy, physiology, and perception. In *Science*, pp. 740– 749, 1988.

[Lubi95] J. Lubin. A visual discrimination model for imaging system design and evaluation. In E. Peli, Ed., *Vision Models for Target Detection and Recognition*, pp. 245– 283, World Scientific, Singapore, 1995.

[Mann74] J. L. Mannos and D. J. Sakrison. The Effects of a visual fidelity criterion on the encoding of images. In *IEEE Transactions on Information Theory*, pp. 525– 536, 1974.

[Mant05] R. Mantiuk, S. Daly, K. Myszkowski, and H.-P. Seidel. Predicting Visible Differences in High Dynamic Range Images – Model and its Calibration. In B. E. Rogowitz, T. N. Pappas, and S. J. Daly, Eds., *Human Vision and Electronic Imaging X, IS&T/SPIE's 17th Annual Symposium on Electronic Imaging (2005)*, pp. 204–214, 2005.

[Mant06] R. Mantiuk, K. Myszkowski, and H.-P. Seidel. A Perceptual Framework for Contrast Processing of High Dynamic Range Images. *ACM Transactions on Applied Perception*, Vol. 3, No. 3, pp. 286–308, 2006.

[Mant08] R. Mantiuk and H.-P. Seidel. Modeling a Generic Tone-Mapping Operator. *To appear in: Computer Graphics Forum*, Vol. 27, No. 3, 2008.

[Mart98] W. L. Martens and K. Myszkowski. Psychophysical Validation of the Visible Differences Predictor for Global Illumination Applications. *Proceedings of IEEE Visualization '98*, pp. 49– 52, 1998.

[Marz02] P. Marziliano, F. D. S. Winkler, and T. Ebrahimi. A no-reference perceptual blur metric. In *International Conference on Image Processing*, pp. III–57– III–60, 2002.

[Matk05] K. Matkovic, L. Neumann, A. Neumann, T. Psik, and W. Purgathofer. Global Contrast Factor - a New Approach to Image Contrast. In L. Neumann, M. Sbert, B. Gooch, and W. Purgathofer, Eds., *Computational Aesthetics in Graphics, Visualization and Imaging 2005*, pp. 159–168, Eurographics Association, May 2005.

[Mont91] D. C. Montgomery. *Design and Analysis of Experiments.* John Wiley and Sons, Inc., New York, 1991.

[Mysz98] K. Myszkowski. The Visible Differences Predictor: applications to global illumination problems. In *Proceedings of the Ninth Eurographics Workshop on Rendering*, pp. 223– 236, Wien, Austria, 1998.

[Nemc04] A. Nemcsics. *Colour Dynamics, Environmental Colour Design.* Akadmiai Kiad, Budapest, 2nd Ed., 2004.

[Nemc80] A. Nemcsics. Coloroid Color System. *Color Research and Application*, Vol. 5, pp. 113–120, 1980.

[Nemc87] A. Nemcsics. Color space of the Coloroid color system. *Color Research and Application*, Vol. 12, pp. 135–146, 1987.

[Neum05] L. Neumann, A. Nemcsics, and A. Neumann. Computational Color Harmony based on Coloroid System. In L. Neumann, M. Sbert, B. Gooch, and W. Purgathofer, Eds., *Computational Aesthetics in Graphics, Visualization and Imaging 2005*, pp. 231–240, 2005.

[Neum98a] L. Neumann, K. Matkovic, and W. Purgathofer. Automatic Exposure in Computer Graphics Based on the Minimum Information Loss Principle. In *Proceedings of Computer Graphics International (CGI '98)*, pp. 666– 679, IEEE Computer Society Press, 1998.

[Neum98b] L. Neumann, K. Matkovic, and W. Purgathofer. Perception Based Color Image Difference. *Computer Graphics Forum*, Vol. 17, No. 3, pp. 233–242, 1998.

[Oh01] J. Oh, S. I. Wooley, T. N. Arvanitis, and J. N. Townend. A multistage perceptual quality assessment for compressed digital angiogram images. *IEEE Transactions on medical imaging*, Vol. 20, No. 12, 2001.

[Osbe98] W. Osberger, N. Bergmann, and A. Maeder. An Automatic Image Quality Assessment Technique Incorporating Higher Level Perceptual Factors. In *Proceedings ICIP-98*, pp. 414–418, USA, 1998.

[Palm02] S. E. Palmer. *Vision science – photons to phenomenology.* The MIT Press, Cambridge, 3rd Ed., 2002.

[Patt00] S. Pattanaik, J. Tumblin, H. Yee, and D. Greenberg. Time-Dependent Visual Adaptation for Fast, Realistic Image Display. In *SIGGRAPH '00: Proceedings*

of the 27th annual conference on Computer graphics and interactive techniques, pp. 47– 54, ACM SIGGRAPH, Addison Wesley, 2000.

[Patt02] S. Pattanaik and H. Yee. Adaptive gain control for high dynamic range image display. In *SCCG '02: Proceedings of 18th spring conference on Computer Graphics*, pp. 83–87, ACM Press, 2002.

[Patt98] S. N. Pattanaik, J. A. Ferwerda, M. D. Fairchild, and D. P. Greenberg. A multiscale model of adaptation and spatial vision for realistic image display. In *SIGGRAPH '98: Proceedings of the 25th annual conference on Computer graphics and interactive techniques*, pp. 287–298, ACM Press, 1998.

[Peli90] E. Peli. Contrast in Complex Images. *Journal of the Optical Society of America A*, Vol. 7, No. 10, pp. 2032–2040, October 1990.

[Prik01] J. Přikryl. *Radiosity Methods Driven by Human Perception.* PhD thesis, Technische Universitat Wien, 2001.

[Rasc05] K. Rasche, R. Geist, and J. Westall. Re-coloring Images for Gamuts of Lower Dimension. *Computer Graphics Forum*, Vol. 24, No. 3, pp. 423–432, 2005.

[Rein02] E. Reinhard, M. Stark, P. Shirley, and J. Ferwerda. Photographic tone reproduction for digital images. In *SIGGRAPH '02: Proceedings of the 29th annual conference on Computer graphics and interactive techniques*, pp. 267–276, ACM Press, 2002.

[Rein05a] E. Reinhard and K. Devlin. Dynamic Range Reduction Inspired by Photoreceptor Physiology. *IEEE Transactions on Visualization and Computer Graphics*, Vol. 11, No. 3, pp. 13–24, 2005.

[Rein05b] E. Reinhard, G. Ward, S. Pattanaik, and P. Debevec. *High Dynamic Range Imaging: Acquisition, Display, and Image-Based Lighting.* Morgan Kaufmann, 2005.

[Renc02] A. C. Rencher. *Methods of Multivariate Analysis.* Wiley series in probability and statistics, 2nd Ed., 2002.

[Rogo98] B. E. Rogowitz, T. Frese, J. R. Smith, C. A. Bouman, and E. B. Kalin. Perceptual image similarity experiments. In B. E. Rogowitz and T. N. Pappas, Eds., *Proceedings of SPIE Human Vision and Electronic Imaging III*, pp. 576–590, July 1998.

[Rush95] H. Rushmeier, G. Ward, C. Piatko, and P. Sanders. Comparing Real and Synthetic Images: Some Ideas About Metrics. In *Eurographics Rendering Workshop 1995*, pp. 82–91, 1995.

[Sava00] A. E. Savakis, S. P. Etz, and A. C. P. Loui. Evaluation of image appeal in consumer photography. In B. E. Rogowitz and T. N. Pappas, Eds., *Human Vision and Electronic Imaging V*, pp. 111–120, SPIE, 2000.

[Sche00] A. Scheel, M. Stamminger, and H.-P. Seidel. Tone Reproduction for Interactive Walkthroughs. *Computer Graphics Forum*, Vol. 19, No. 3, pp. 301–311, 2000.

[Schl94] C. Schlick. An Adaptive Sampling Technique for Multidimensional Ray Tracing. In P. Brunet and F. Jansen, Eds., *Photorealistic Rendering in Computer Graphics*, pp. pp. 21–29, 1994.

[Shei02] H. R. Sheikh, Z. Wang, L. Cormack, and A. C. Bovik. Blind Quality Assessment of JPEG2000 Compressed Images. In *IEEE Asilomar Conference on Signals, Systems and Computers*, 2002.

[Shna04] A. Shnayderman, A. Gusev, and A. M. Eskicioglu. A Multidimensional Image Quality Measure Using Singular Value Decomposition. In *Proceedings of the SPIE Image Quality and System Performance Conference*, pp. 82–92, San Jose, CA, 2004.

[Shna06] A. Shnayderman, A. Gusev, and A. M. Eskicioglu. An SVD-Based Gray-Scale Image Quality Measure for Local and Global Assessment. *IEEE Transactions on Image Processing*, Vol. 15, No. 2, pp. 422–9, 2006.

[Sieg88] S. Siegel and N. J. Castellan. *Nonparametric statistics for the behavioral sciences.* McGraw-Hill, London, 2nd Ed., 1988.

[Smit08] K. Smith, P.-E. Landes, and K. M. Jöelle Thollot. Apparent Greyscale: A Simple and Fast Conversion to Perceptually Accurate Images and Video. *To appear in: Computer Graphics Forum*, Vol. 27, No. 3, 2008.

[Smit99] S. W. Smith. *The scientist and engineer's guide to digital signal processing.* California Technical Publishing, San Diego, CA, USA, 2nd Ed., 1999.

[Spen95] G. Spencer, P. Shirley, K. Zimmerman, and D. P. Greenberg. Physically-based glare effects for digital images. In *SIGGRAPH '95: Proceedings of the 22nd annual conference on Computer graphics and interactive techniques*, pp. 325–334, ACM Press, 1995.

[Stat07] StatSoft, Inc. *Electronic Statistics Textbook.* Tulsa, OK: StatSoft, `http://www.statsoft.com/textbook/stathome.html`, 2007.

[Szel06] R. Szeliski. Locally adapted hierarchical basis preconditioning. In *SIGGRAPH '06: Proceedings of the 33nd annual conference on Computer graphics and interactive techniques*, pp. 1135–1143, ACM Press, 2006.

[Taba07] B. G. Tabachnick and L. S. Fidell. *Using multivariate statistics.* Pearson Education, Inc., 5th Ed., 2007.

[Taub02] D. S. Taubman and M. W. Marcellin. *JPEG2000: Image Compression Fundamentals, Standards and Practice.* Kluwer Academic Publishers, 2002.

[Tayl97] C. Taylor, Z. Pizlo, J. Allebach, and C. Bouman. Image Quality Assessment with a Gabor Pyramid Model of the Human Visual System. *Proceedings of the 1997 IS&T/SPIE International Symposium on Electronic Imaging Science and Technology*, Vol. 3016, pp. 58– 69, 1997.

[Teo94] P. C. Teo and D. J. Heeger. Perceptual image distortion. In *Proceedings ICIP-94: IEEE International Conference on Image Processing*, pp. 982–986, 1994.

[Thur27] L. L. Thurstone. A law of comparative judgement. *Psychological Review*, Vol. 34, pp. 278–286, 1927.

[Tong99] X. Tong, D. Heeger, and C. J. van den Branden Lambrecht. Video quality evaluation using ST-CIELAB. In *Proceedings of SPIE Human Vision and Electronic Imaging*, pp. 185– 196, San Jose, CA, 1999.

[Torg58] W. S. Torgerson. *Theory and methods of scaling*. John Wiley & Sons, Inc., New York, NY, USA, 1958.

[Tumb93] J. Tumblin and H. Rushmeier. Tone Reproduction for Realistic Images. *IEEE Computer Graphics and Applications*, Vol. 13, No. 6, pp. 42–48, 1993.

[Tumb99a] J. Tumblin and G. Turk. Low Curvature Image Simplifiers (LCIS). In *SIGGRAPH '99: Proceedings of the 26th annual conference on Computer graphics and interactive techniques*, pp. 83–90, Addison Wesley, 1999.

[Tumb99b] J. Tumblin, J. K. Hodgins, and B. K. Guenter. Two methods for display of high contrast images. *ACM Transactions on Graphics*, Vol. 18, No. 1, pp. 56–94, 1999.

[VQEG00] VQEG. Final report from the video quality experts group on the validation of objective models of video quality assessment. Available from `http://www.vqeg.org`, March 2000.

[Wang02a] Z. Wang and A. C. Bovik. A Universal Image Quality Index. *IEEE Signal Processing Letters*, Vol. 9, No. 3, pp. 81– 84, 2002.

[Wang02b] Z. Wang, A. C. Bovik, and L. Lu. Why is image quality assessment so difficult. In *Proceedings of IEEE International Conference Acoustics, Speech, and Signal Processing*, pp. 3313– 3316, 2002.

[Wang03] Z. Wang, H. R. Sheikh, and A. C. Bovik. Objective video quality assessment. In B. Furht and O. Marqure, Eds., *The Handbook of video databases: design and applications*, Chap. 41, pp. 1041– 1078, CRC Press, 2003.

[Wang04] Z. Wang, A. C. Bovik, H. R. Seikh, and E. P. Simoncelli. Image quality assessment: From error visibility to structural similarity. In *IEEE Transactions on Image Processing*, pp. 600–612, 2004.

[Ward94] G. Ward. A contrast-based scalefactor for luminance display. *Graphics Gems IV*, pp. 415– 421, 1994.

[Wats00] B. Watson, A. Friedman, and A. McGaffey. Using Naming Time to Evaluate Quality Predictors for Model Simplification. In *CHI Letters*, pp. 113– 120, 2000.

[Wats87] A. B. Watson. The cortex transform: Rapid computation of simulated neural images. *Computer Vision Graphics and Image Processing*, Vol. 39, No. 3, pp. 311–327, 1987.

[Weib03] Weibull.com. Analysis of Variance. Web Pages, 2003. `http://www.weibull.com/LifeDataWeb/analysis_of_variance.htm`.

[Wink00] S. Winkler. *Vision Models and Quality Metrics for Image Processing Applications.* PhD thesis, EPFL, Switzerland, 2000.

[Wysz82] G. Wyszecki and W. S. Stiles. *Color Science: Concepts and Methods, Quantitative Data and Formulae.* John Wiley & Sons, Inc., 2nd Ed., 1982.

[Yosh05] A. Yoshida, V. Blanz, K. Myszkowski, and H.-P. Seidel. Perceptual evaluation of tone mapping operators with real-world scenes. *Human Vision & Electronic Imaging X, SPIE*, pp. 192–203, 2005.

[Yosh06] A. Yoshida, R. Mantiuk, K. Myszkowski, and H.-P. Seidel. Analysis of Reproducing Real-World Appearance on Displays of Varying Dynamic Range. *Computer Graphics Forum*, Vol. 25, No. 3, pp. 415–426, September 2006.

[Zhan96] X. Zhang and B. A. Wandell. A spatial extension of CIELAB to predict the discriminability of colored patterns. In *Society for Information Display Symposium Digest*, pp. 731– 735, 1996.

A Statistical Tests

"Graeca sunt, non leguntur."

(Principle of medieval monks)

In this appendix we describe statistical tests that are useful and necessary to process data obtained in subjective experiments. Generally, we can divide the tests into the *parametric* and *non-parametric statistics*. Parametric statistics assume that the population fits a parametrized distribution (e.g. the normal distribution), while non-parametric statistics do not make any assumption about the underlying distribution. Non-parametric statistics are less statistically sensitive than their parametric counterparts, however they are appropriate when the sample sizes are small[1] and non-parametric (e.g. rank ordered data).

A.1 ANOVA

Analysis of variance (ANOVA) is one of the most frequently used statistical methods. It is used to uncover the main and interaction effects of independent variables (called "treatments" or "factors") on dependent variables. We distinguish several types of ANOVA [Stat07]:

- *Non-parametric and Parametric ANOVA*
  ANOVA is available for scores or interval data as a parametric ANOVA, refer to [Taba07, Page 37]. This is the type of ANOVA available from the standard menu options in statistical packages. Non-parametric versions of ANOVA [Sieg88, Ande03] are used when we have rank or ordered data, we cannot use parametric ANOVA when the data is below interval measurement. Where we have categorical data we do not have an ANOVA method – we would have to use χ^2 (Chi-square) [Taba07, Page 58] which is about interaction rather than about differences between groups.

- *One-way and Multi-Factor ANOVA*
  One-way (or one factor) ANOVA [Taba07, Page 37] is the simplest version of ANOVA. We are looking at the differences between the groups. There is only one grouping factor which we are using to define the groups. This type of ANOVA can also be used to compare variables between different groups.

  Multi-factor ANOVA [Taba07, Page 43] allows us to ask questions about more than one factor in the same experiment and it allows also to detect interaction effects between variables. Interaction means that the effect that one independent variable has on dependent variables depends on the level of other independent variable(s). Therefore, it allows to test more complex hypotheses about the data. A general way to express all interactions is to say that an effect is modified (qualified) by another effect [Stat07].

[1] When the data set is large (e.g. $N > 100$) it often makes sense to use parametric statistics [Kend90].

- *Univariate and Multivariate ANOVA (MANOVA)*
  An ordinary (univariate) ANOVA assumes just one dependent variable. When there is more than one dependent variable, one can utilize multivariate ANOVA (MANOVA). MANOVA [Taba07, Page 243] evaluates differences among centroids (composite means) for a set of dependent variables when there are two or more groups of an independent variable [Taba07]. Once statistically significant differences are found, techniques are available to assess which dependent variables are influenced by which independent variables.

- *Repeated measures ANOVA*
  A repeated measures ANOVA [Taba07, Page 43] is used when we have groups on which we have measured something a few times. If we have repeated measurements of the same variable (under different conditions or at different points in time) on the same subjects, then the factor is a repeated measures factor (also called a within-subjects factor). If we compare different groups of subjects (e.g. males and females; three strains of bacteria) then we refer to the factor as a between-groups factor [Stat07].

To show the principle, we will explain the *one-way ANOVA* in more detail. The one-way ANOVA performs a comparison of the means of a number of replications of experiments performed where a single input factor is varied at different settings or levels [Weib03]. The object of this comparison is to determine the proportion of the variability of the data that is due to the different treatment levels or factors as opposed to variability due to random error. The model deals with specific treatment levels and is involved with testing the *null hypothesis* $H_0 : \mu_1 = \mu_2 = \cdots = \mu_a$ where μ_i represents the level mean. Basically, rejection of the null hypothesis indicates that variation in the output is due to variation between the treatment levels, and not due to random error. If the null hypothesis is rejected, there is a difference in the output of the different levels at a significance, and it remains to be determined between which treatment levels the actual differences lie.

Required inputs to perform a one-way ANOVA are the number of levels being compared, a, and the number of replications at each level, n_i. Typically, the user is asked to provide a value for the risk factor, α, which represents the Type 1 error probability the user is willing to accept. Next, the actual results of the experiments need to be entered. The experiment result data will be denoted as $y_{i,j}$, where $i = 1, 2, \cdots, a$ is the number of levels being tested and $j = 1, 2, \cdots, n_i$ is the number of replicates at each level.

In analysis of variance, the output of each experiment, or observation, is thought to consist of variations of an overall mean value. These variations can have two sources: variation due to the factor or level and variation due to random error. The model used for the data in ANOVA analysis follows the form $y_{ij} = \mu + \tau_i + \epsilon_{ij}$, where μ is the overall mean, τ_i is the level effect and ϵ_{ij} is the random error component. Since the purpose of this analysis is to determine if there is a significant difference in the effects of the factors or levels, the null hypothesis can also be written as $H_0 : \tau_1 = \tau_2 = \cdots = \tau_a = 0$. The mean of the responses over a level is denoted as $\bar{y_i} = \frac{\sum_{j=1}^{n_i} y_{ij}}{n_i}$, while the overall mean of the data is $\bar{y_{..}} = \frac{\sum_{i=1}^{a} \sum_{j=1}^{n_i} y_{ij}}{N}$. The analysis is broken down into "sums of squares" that measure the variability due to the levels and due to the errors. The general form is $SS_T = SS_L + SS_E$, where SS_T is the total sum of squares, which is a measure of the overall variability of the data: $SS_T = \sum_{i=1}^{a} \sum_{j=1}^{n_i} (y_{ij})^2 - \frac{(y_{..})^2}{N}$, SS_L is the sum of squares for the levels, which measures the variability due to the levels or

factors: $SS_L = \sum_{i=1}^{a} \frac{(y_i)^2}{n_i} - \frac{(y_..)^2}{N}$. With SS_T and SS_L known we can compute the sum of squares due to the errors: $SS_E = SS_T - SS_L$. There are degrees of freedom terms associated with each of the sums of squares. The degrees of freedom are given by: $df_L = a - 1$, $df_E = N - a$ and $df_T = N - 1$. Mean square values are calculated by dividing the sum of square terms for the level and error by their respective degrees of freedom values. These values represent the variance of the level and error components of the data. Mean squares values for levels and errors are $MS_L = \frac{SS_L}{a-1}$, $MS_E = \frac{SS_E}{N-a}$. Cochran's theory [Mont91] states that the ratio $F_0 = \frac{MS_L}{MS_E}$ is F-distributed with degrees of freedom $a-1$ and $N-a$. Therefore, if $F_0 > F_{crit}$, where $F_{crit} = F_{\alpha,a-1,N-a}$, we reject the null hypothesis and conclude that some of the variability of the data is due to differences in the treatment levels.

The generally used format for output for this type of analysis is an ANOVA table, which contains basic informations about the analysis results:

Source of Variation	**Sum of Squares**	**Degrees of freedom**	**Mean Square**	**F Values**
Between levels	SS_L	df_L	MS_L	F_0
Error (within levels)	SS_E	df_E	MS_E	F_{crit}
Total	SS_T	df_T		

A.2 Correlation Coefficients

Correlation coefficient is a measure of the size and direction of a linear relationship between two random variables [Taba07].

A.2.1 Pearson Product-Moment Correlation Coefficient (r)

Pearson product-moment correlation coefficient r is the most common parametric measure of correlation between variables X and Y. The coefficient is defined as follows:

$$r = \frac{\mathrm{cov}(X,Y)}{\sqrt{(\mathrm{var}X\,\mathrm{var}Y)}},$$

where cov denotes covariance and var denotes variance.

A.2.2 Kendall Tau Coefficient (t)

Kendall tau (t) is a non-parametric measure of concordance between two sets of n rankings, i.e. it is a coefficient of rank correlation ranging in [-1, 1]. Kendall tau is calculated as follows:

$$t = \frac{2(P-Q)}{n(n-1)},$$

where P is the number of concordant pairs, Q is the number of discordant pairs, $(P-Q) = S$ is the excess of concordant pairs over discordant pairs, and n is the number of pairs of values.

A.2.3 Spearman's Rho (r_s)

Spearman's Rho is a non-parametric coefficient of rank correlation ranging in [-1, 1]. Spearman's Rho is calculated as follows:

$$r_s = 1 - \frac{6 \sum d^2}{n^3 - n},$$

where d is the difference between each rank of corresponding values, and n is the number of pairs of values.

For conjugate rankings (X, X') when correlated with Y they give values of t (stands for r_s as well) which are equal in magnitude but opposite in sign - the scales are symmetrical about zero, unbiased. In practice (when neither coefficient is too close to unity) r_s is about 50% greater than t in absolute value, but this is not an invariable rule [Kend90].

B Multiscale Transforms

This appendix presents a brief introduction to a series of multiscale transforms that are widely applied to signal and image processing [Font01]. In general, signal analysis by using a multiscale transform is characterized by the following elements. A signal $u(t)$ presents a set of features and structures occuring at different spatial scales. This signal is to be analyzed by a multiscale transform $U(b, a)$ involving two parameters: b, associated with the *time* variable t of $u(t)$, and a, associated with the *analyzing scale.* The scale parameter a is usually related to the inverse of the *frequency* f, i.e., $\frac{1}{a} \cong f$, leading to a dual interpretation of these transformations and suggesting the terms *time-scale* and *time-frequency.*

Many different multiscale-based signal analysis frameworks can be summarized by the following Algorithm:

Algorithm 3: Typical Multiscale Analysis
(1) Obtain the signal $u(t)$ to be analyzed
(2) Calculate the multiscale transform $U(b, a)$ of $u(t)$
(3) Extract the important scale characteristics of $u(t)$ from $U(b, a)$

The *scale-space* approach to multiscale signal analysis is one of the most popular multiscale methods. The scale-space of a signal allows tracking its singularities (or of one of its derivatives) through the scale dimension. This fits to conjecture by Marr that the perceptually important dominant signal points correspond to singularities remaining along longer scale intervals. The scale space is defined as follows.

Definition B.0.1 *Let $u(t)$ be the signal to be analyzed, $g_a^{(1)}$ be the first derivative of the Gaussian function $g_a(t)$ and $U^{(1)}(t, a) = u(t) * g_a^{(1)}(t)$. Let $\left\{U^{(1)}(t, a_0)\right\}_{zc}$ denote the zero-crossings set of $U^{(1)}(t, a_0)$. The* ***scale-space*** *of $u(t)$ is defined as the set of zero-crossings of $U^{(1)}(t, a)$, i.e.*

$$\left\{(b_0, a_0) | a_0, b_0 \in R, a_0 > 0, and\ b_0 \in \left\{U^{(1)}(t, a_0)\right\}_{zc}\right\}$$

The term scale-space is sometimes also used for $U(t, a)$, obtained by the convolution of the signal with a series of Gaussians. In general, the extrema of $U^{(n)}(t, a)$ can be defined from the zero-crossings of the scale-space generated by $g_a^{n+1}(t)$.

The *time-frequency* transforms have originated as an alternative to Fourier analysis capable of *signal local analysis.* Fourier transform has engaging properties, but it operates on the whole signal. The short-time Fourier transform has been defined as an attempt to circumvent this problem by introducing an *observation window* aimed at selecting a signal portion during the transform. The following transform was defined from the Fourier equation:

$$U(b, f) = \int_{-\infty}^{\infty} g^*(t - b) u(t) \exp(-j2\pi f t) dt,$$

where $g(t)$ is the window function that *slides* over the signal $u(t)$. Time-frequency analysis is based on the above considerations, and one of its most important tools is the *Gabor transform,* i.e., the short-time Fourier transform where $g(t)$ is a Gaussian window.

The *time-scale* transform (or more often referred to as the *wavelet* transform) faces the following problem: high-frequency events frequently occure along short time intervals while low-frequency components remain longer during the signal evolution. The application of the short-time Fourier transform has the drawback that the analyzing window size is the same for all frequencies. Therefore, in the wavelet transform the kernel size varies with the frequency, allowing high-frequency events to be localized with better time resolution, while low-frequency components are analyzed with better frequency resolution. This property is known as *relative bandwidth* or *constant-Q.* The continuous wavelet transform is defined as:

$$U[\psi, u](b, a) = U_\psi(b, a) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} \psi^* \Big(\frac{t-b}{a} \Big) u(t) dt.$$

In the case of the Morlet wavelet, the Gabor transform and so-called Gabor wavelets [Krug99], the transform presents a strong and interesting biological inspiration (the receptive fields of neural cells involved in visual processing).