

Photometric Image Processing for High Dynamic Range Displays

by

Matthew Trentacoste

B.Sc., Carnegie Mellon University, 2003

A THESIS SUBMITTED IN PARTIAL FULFILMENT OF
THE REQUIREMENTS FOR THE DEGREE OF

Master of Science

in

The Faculty of Graduate Studies

(Computer Science)

The University Of British Columbia

January, 2006

© Matthew Trentacoste 2006

Abstract

Many real-world scenes contain a dynamic range that exceeds conventional display technology by several orders of magnitude. Through the combination of several existing technologies, new high dynamic range displays, capable of reproducing a range of intensities much closer to that of real environments, have been constructed. These benefits come at the cost of more optically complex devices; involving two image modulators, controlled in unison, to display images. We present several methods of rendering images to this new class of devices for reproducing photometrically accurate images. We discuss the process of calibrating a display, matching the response of the device with our ideal model. We then derive series of methods for efficiently displaying images, optimized for different criteria and evaluate them in a perceptual framework.

Contents

| | |
|---|-----|
| Abstract | ii |
| Contents | iii |
| List of Tables | vi |
| List of Figures | vii |
| Acknowledgements | ix |
| 1 Introduction | 1 |
| 1.1 Image Processing for HDR Displays | 4 |
| 1.2 Photometric Imaging | 5 |
| 1.3 Terminology | 6 |
| 2 Related Work | 8 |
| 2.1 Perception and Psychophysics | 8 |
| 2.1.1 Local Contrast Perception | 9 |
| 2.1.2 Luminance Quantization | 11 |
| 2.1.3 Visual Difference Prediction | 15 |
| 2.2 Tonemapping Operators | 18 |
| 2.2.1 Taxonomy of Operators | 19 |
| 2.2.2 Validation | 22 |
| 2.2.3 Shortcomings of Tonemapping | 23 |
| 2.3 HDR Technology | 24 |
| 2.3.1 Projector-based Display | 25 |

| | |
|---|-----------|
| 2.3.2 LED-based Display | 29 |
| 2.4 Display Calibration | 30 |
| 2.4.1 Gamma | 30 |
| 2.4.2 Implications for HDR displays | 35 |
| 3 Processing Algorithms | 37 |
| 3.1 Reference Algorithm | 38 |
| 3.1.1 Nonlinear System | 39 |
| 3.1.2 Observations | 42 |
| 3.2 Performance-Related Modifications | 43 |
| 3.2.1 Simplification of Simulation | 44 |
| 3.2.2 Problem Decomposition | 46 |
| 3.2.3 Approximate Solution | 48 |
| 3.3 Implementation | 50 |
| 3.3.1 Target Backlight | 52 |
| 3.3.2 Deriving LED Intensities | 54 |
| 3.3.3 Backlight Simulation | 57 |
| 3.3.4 Blur Correction | 59 |
| 3.4 Error Diffusion | 60 |
| 3.4.1 Rationale | 60 |
| 3.4.2 Backlight Update Process | 62 |
| 3.4.3 Corrective Image Filter | 63 |
| 4 Measurement and Calibration | 66 |
| 4.1 LED Array | 66 |
| 4.2 LCD Panel Response | 68 |
| 4.3 Diffuser Pointspread Function | 69 |
| 5 Evaluation | 74 |
| 5.1 Preliminaries | 74 |
| 5.2 Algorithm Evaluation | 77 |
| 5.3 Discussion | 84 |

| | |
|--------------------------------|----|
| 6 Conclusions | 87 |
| 6.1 Contributions | 87 |
| 6.2 Future Work | 88 |
| 6.3 Closing Remarks | 89 |
| Bibliography | 91 |

List of Tables

| | | |
|-----|--|----|
| 2.1 | Table of tonemapping operators cited in this thesis | 19 |
| 5.1 | Table of percent of total pixels at or above a detection level | 84 |

List of Figures

| | |
|--|----|
| 2.1 Modulated transfer function of the ocular medium | 10 |
| 2.2 Contrast versus intensity curve | 12 |
| 2.3 CIE L^* curve | 13 |
| 2.4 Derived just noticeable differences (JND) curves | 14 |
| 2.5 Inputs and resulting output from each stage of the VDP process | 16 |
| 2.6 Selection of tonemapping operators applied to images | 20 |
| 2.7 Internal schematic of projector display | 25 |
| 2.8 Photograph of BrightSide DR37-P | 29 |
| 2.9 Comparison on L^* , Rec. 709, and sRGB OETFs | 31 |
| 2.10 Tone scale curves for different intensity surroundings | 33 |
| 2.11 Flowchart of the different aspects of gamma | 34 |
| | |
| 3.1 Two primary challenges in image presentation | 38 |
| 3.2 Flowchart of nonlinear optimization | 39 |
| 3.3 Input image and resulting backlight and LCD panel images | 43 |
| 3.4 Sparsity pattern of simulation matrices | 45 |
| 3.5 Blur correction steps | 47 |
| 3.6 Veiling glare restricting neighborhood | 49 |
| 3.7 Flowchart of stages of the implementation | 51 |
| 3.8 Tonemapped original HDR image for reference | 52 |
| 3.9 Output of target backlight pass | 53 |
| 3.10 Output of pass to determine LEDs | 56 |
| 3.11 Remapping of hex grid to regular grid | 57 |
| 3.12 Output of backlight simulation pass | 58 |

| | |
|---|----|
| 3.13 Output of blur correction pass | 59 |
| 3.14 Difference between desired and actual p depending on c_B | 61 |
| 3.15 Comparison of error diffusion to original method | 64 |
| 4.1 LCD panel response | 69 |
| 4.2 Pointspread function of diffuser | 71 |
| 4.3 Spatial response of display | 72 |
| 5.1 Comparison of veiling glares | 75 |
| 5.2 Example of HDR VDP output | 76 |
| 5.3 TestPattern | 79 |
| 5.4 FrequencyRamp | 80 |
| 5.5 Apartment | 82 |
| 5.6 Moraine | 83 |
| 5.7 TestPattern distance comparison | 85 |
| 5.8 FrequencyRamp distance comparison | 85 |

Acknowledgements

Firstly, I'd like to thank everyone that contributed ideas, discussion, and proof-reading to this work. In particular, I'd like to thank Wolfgang Heidrich, Lorne Whitehead, Abhijeet Ghosh, Helge Seetzen, Bob Woodham, Ciaran Llachlan Leavitt, Rafal Mantiuk, Erik Reinhard, and Greg Ward.

I want to thank my family for their support through this: my mother Kathy, my father Michael, and my sisters Angela and Emily.

Also, I owe a debt of gratitude to everyone at BrightSide Technologies and the Structured Surface Physics Lab for helping me in more ways than I can count. I'd like to thank Neil McPhail, Vincent Kwong, Michelle Mossman, Pete Longhurst, Jason Harrison, Thomas Wan, Henry Ip, Gary Yurkovich, Richard MacKellar and everyone else for their efforts in getting me the information, time, and materials I needed.

Finally, a big shout-out to all the people who, through administering equal measures of sanity and insanity, helped me make it to the end. The members of Imager Small, past, present, or in spirit: Dave Burke, Vladislav Kraevoy, Fred Kimberly, Ritchie Argue, James Slack, Kristian Hildebrand, Abhijeet Ghosh, Peter Macallan, Tyson Brochu, Dinos Tsinkis, Ciaran Llachlan Leavitt, Derek Bradley, and Chen Yang. And, everyone of Pod6 and EastVan, especially: Drew Smith, Erin Caton, Kira Lorber, Tom Schulz, Ross Kakuschke, Carrie Murdoch, Nicole Sanches, and Rich Hamakawa.

Chapter 1

Introduction

The high dynamic range (HDR) rendering pipeline has been the subject of considerable interest from the computer graphics community in recent years. The intensities and dynamic ranges found in many scenes and applications vastly exceed those of conventional imaging techniques, and the established practices and methods of addressing those images are insufficient.

Existing digital cameras can faithfully record images over a wide range of intensities, but are significantly limited in the dynamic range. Dynamic range is the ratio of brightest to darkest value that they can record simultaneously. Even given a means of generating or acquiring such data, conventional file formats cannot accurately store it. The same is true with monitors; conventional display technologies can give a correct impression of relative luminance over a limited luminance range, but they are limited in their ability to reproduce values that are bright or dark enough to accurately represent anything more than a fraction of the luminances encountered in ordinary scenes. A standard display does not have nearly the level of contrast, or the dynamic range, to directly reproduce many real-world scenes.

Researchers have developed additions and modifications to existing methods of acquiring, processing, and displaying images to accommodate contrasts which exceed the limitations of conventional, low dynamic range (LDR) techniques and devices. Methods exist for acquiring HDR images and video from multiple LDR images. First investigated by Mann & Picard [44], these techniques were introduced to graphics by Debevec & Malik [16]. Additional work has been done by Mitsunaga & Nayar [50] and Robertson et al [63] on still images, while Kang et al [34] have applied the methods to video.

File formats have been designed to accommodate the additional data storage re-

quuirements. The OpenEXR [33] format efficiently stores data without complicated encoding for real-time applications, and JPEG-HDR [78] encodes the additional information while maintaining transparent, backwards-compatible support for existing applications. Additional formats have been developed for efficiently encoding images by storing only perceptually-relevant information. The work on images by Larson [37] and on video by Mantiuk et al [46] are examples of formats that can significantly compress image data while remaining perceptually lossless. Techniques have been developed to compress the dynamic range of images while preserving features of the original. These techniques, known collectively as tonemapping operators, allow the display of HDR images on conventional monitors with contrast ratios of about 300 : 1, including conventional Cathode Ray Tube (CRT), Liquid Crystal Display (LCD), and projector-based displays.

While digital image manipulation and storage technologies can adequately address HDR images, complications in the acquisition and display stages remain. Neither commodity cameras nor commodity display devices fully support HDR imaging. The traditional HDR acquisition methods mentioned above make assumptions about the content, require multiple exposures, as well as a static scene. Likewise, tonemapping operators can map scene details into a range displayable on a conventional display, but cannot fully reproduce the scene. There is an inevitable loss of information in the process, and the lower intensities of a conventional display cannot completely reproduce the same sensation as the original scenes.

Recent advances in image sensor technology are providing a direct means of accomplishing these goals without those restrictions. New cameras are capable of capturing larger dynamic ranges in a single exposure than can existing models. While still specialty items, these devices, such as the HDRC VGAx [30] camera, the Thompson Viper Film-Stream [72] video camera, and the SpheronVR SpheroCamHDR [68] panoramic camera, are becoming more common. Most relevant to this thesis, high dynamic range display systems have been developed to accurately reproduce a much wider range of luminances. The work done by Ward [77] and Seetzen et al [64, 65] has provided devices that vastly exceed the dynamic range of conventional displays. These devices are capable of higher intensity whites, and lower intensity blacks while maintaining adequately low quantization across entire luminance range.

No single known material is capable of reproducing the luminances and bit depths in the resolutions and form factors required for displaying HDR images, and a fundamental change in how output device display images is required. A conventional display uses a single high resolution LCD panel as an optical filter in front of a uniform light source, like a fluorescent lamp. The limited contrast of a LCD panel requires an additional optical modulator to be added, and the design of an HDR display accomplishes this by replacing the uniform light with a second low resolution, high contrast display. There are many ways to create the display, but in practice either a projector or a grid of ultra-bright LEDs is used. By simultaneously controlling the LCD panel and the second display, the two work in tandem to produce the final image.

This new configuration offers many benefits over conventional displays, but presents several additional challenges. If one desires to alter the luminance of a pixel by using the low resolution backlight, the surrounding pixels are altered as well. Fundamentally, this limitation implies that HDR displays cannot exactly reproduce the luminances of a real scene. However, since the display is intended to be viewed by human subjects, exact reproduction is not necessary. As long as the display introduces less distortion than the human visual system, the original image and the displayed image will appear the same.

Unlike conventional displays, the pixels in the HDR display are no longer completely independent of one another. It is therefore necessary to employ image-processing algorithms to factor an HDR into values to send to the LCD panel, and to the low resolution back plane, respectively. This thesis addresses the challenge of: given an image as input, compute a matching set of front and back images such that the optics of the display combine to produce the same observed image as the original.

Many real-world scenes contain a dynamic range that exceeds conventional display technology by several orders of magnitude. Through the combination of several existing technologies, new high dynamic range displays have been constructed. While these displays are capable of reproducing intensity ranges comparable to some real environments, their benefits come at a cost. The hardware setup requires a more optically complex device; reproducing pictures involves two sets of controllable image elements to be operated simultaneously.

1.1 Image Processing for HDR Displays

The goal of the work presented here is to overcome the challenges of the HDR display hardware design and accurately reproduce photometric images. Achieving this goal entails designing efficient algorithms to produce the best images possible, characterizing the monitor, and calibrating it to reproduce the same appearance it is given as input. The full realization of this goal is a monumental challenge, drawing upon work from numerous areas of research, and cannot be resolved within the scope of a single thesis. In order to verify the accuracy of the reproduction, the results would need to be validated to match to a human viewer, and much of the required perceptual foundation has not been completely explored.

We will only address the challenge of accurately reproducing perceived luminances. Due to the vast scope, other areas such as color, motion, and spatial frequency will not be fully resolved. We touch on the topics of motion and color, but our coverage is not comprehensive. We will present methods of processing images that address the inherent challenges of the HDR display within the set of constraints that the hardware configuration places. We will calibrate those methods to accurately reproduce luminance, and draw upon psychophysical studies to verify the results. The remainder of this thesis is structured as follows:

Related Work: Chapter 2 covers the topics related to the work presented. This collection of topics provides key insights into understanding our methods and their evaluation. The four areas discussed are: aspects of perception and psychophysics, the tonemapping operators conventionally used to view HDR images, the physical construction of the HDR display systems, and calibration methods used for LDR displays.

Rendering Algorithms: Chapter 3 describes the task of rendering images and details the difficulties faced in doing so. We will present the idealized model of the display hardware that will be the foundation of our work, and will discuss the general high-level view of the problem and the areas of optimization. Finally, several efficient algorithms for achieving these goals will be detailed.

Characterization and Calibration: Chapter 4 begins with an enumeration of the differences between the real display hardware and the idealized model assumed in the algorithms. Paying specific attention to the complexities introduced by the low resolution backlight, we detail the measurements required to correct for those disparities and calibrate the output, and how those measurements are incorporated into the image processing methods. We present the measurements taken in addition to the calibration process.

Evaluation: Chapter 5 presents the results of the work and evaluates them using a perceptually-based metric. Due to the limitations of the hardware design, the HDR display is not capable of reproducing the luminance of the original scene at every pixel. Because the stated goal is to reproduce the appearance of the original to a human, instead of an exact photometric representation, the output of the display is only required to be sufficiently close so that an observer cannot discern any differences. Improvements are unnecessary if they cannot be discerned by a human observer, and metrics derived from human perceptual studies provide a meaningful bound. Movies do not exceed 30 Hz and interactive applications do not exceed 72 Hz because viewers cannot discern the difference. Similarly, as displays approach the simultaneous contrast perception of the human visual system (HVS), it becomes necessary to analyze them in terms of the observer's abilities.

1.2 Photometric Imaging

While high dynamic range images are not subject to the limitations of intensity and dynamic range associated with conventional images, they share many of the same ambiguities. A survey of HDR images quickly reveals that there is no consensus on what the pixel values mean in terms of real luminance values. The data is still effectively relative, and often scaled arbitrarily. It is not uncommon to find an image of a nighttime scene with pixel values orders of magnitude greater than the pixel values in an image representing a sunny scene.

Ideally, in addition to faithfully representing ratios comparable to the original scene,

the image should contain enough information to determine the luminance values of that scene from the pixels in the image. In order to properly record luminance, pixel intensities must be linearly stored in absolute units of light, such as candela-per-meter-squared (cd/m^2). In order to properly record color, additional information must be included to describe the gamut in which the colors exist. This accuracy implies measuring the acquisition device to quantify its characteristics, and providing a mapping of pixel values back to the recorded luminance. This extra set of constraints is commonly termed photometric imaging, as it directly relates pixel values to the measured photons of light in the original scene.

Work has already been done by Krawczyk et al [35] on calibrating HDR acquisition devices to encode photometric data. A natural extension of this type of acquisition is to accurately reproduce photometric images, performing the same calibration on display devices, which was not possible in the past. Real scenes have an average dynamic range of 3 orders of magnitude. Both daytime and nighttime scenes have roughly the same contrast, but vastly different mean luminances. Intensity and dynamic range limitations make this level of calibration impossible on conventional displays, however the HDR display can represent contrasts of this magnitude and has a peak intensity comparable to that of indoor scenes. It is the first display able to reproduce original scene luminances, thus providing strong motivation to photometrically calibrate its output.

1.3 Terminology

Unfortunately, there is often confusion about the terminology used to describe the quantities of light in the real scene, the values of the imaging pipeline, and the image perceived by the viewer. We are discussing an imaging system that differs from tradition systems, requiring shifting between multiple representations of light, and the implications of those to human perception. It is critical to describe exactly what is meant by each term.

Luminance – radiance weighted by the spectral sensitivity associated with the bright-

ness sensation of vision. It is the result of weighting a spectrum by the function $Y(\lambda)$ ¹ and represents the relative intensities of wavelengths visible to human observers. It directly corresponds to scene intensities and is also referred to as *linear light*. In the case of LDR images it has been used to mean values proportional to intensity, or *relative intensity*, as opposed to photometric images which record *absolute intensity*.

Lightness – the nonlinear quantization of luminance that expresses it in perceptually uniform units. Due to the nature of the visual system, for different absolute luminances, the same change in relative luminance appears different in magnitude. Equal sized changes of lightness appear the same, invariant of the value they are relative to. This is often referred to as *just-noticeable-difference* (JND) space, and is detailed in Section 2.1.2. Some literature uses lightness to specifically denote the standard LDR approximation, CIE L^* ; a metric not appropriate for photometric images.

¹Also written as $V(\lambda)$.

Chapter 2

Related Work

There is a wide variety of research related to the topic of processing images for HDR displays. Section 2.1 describes several aspects of perception and psychophysics, and provides background on the attributes of the human visual system. The focus is on aspects related to HDR displays and their evaluation, and discusses considerations for extending methods to photometric imaging. Section 2.2 presents some of the principal work on the conventional method of viewing HDR images: tonemapping operators. In particular, we highlight operators that share similarity to the image processing presented later. Section 2.3 describes the physical construction of the HDR display systems. We provide a concrete foundation from which to understand the considerations made in the rendering methods. Section 2.4 examines calibration methods used for LDR displays to provide a context for what is required to calibrate HDR displays.

2.1 Perception and Psychophysics

Any analysis of the display of images includes an inherent discussion about the viewer: the perceptual makeup of the human observer. The human visual system (HVS) is powerful, capable of accommodating a wide range of different conditions. Alternate means of representing visual information have evolved to overcome biological limitations, and have modified the perception of imagery encountered. One example is that the appearance of scenes is dependent on the intensities and contrast ranges they contain [23], with numerous everyday examples such as, “bright colors look more vivid” and “things appear blueish at night.” An immense body of research on the study and characterization of the various aspects of the HVS exists; far larger in the literature than what could

be covered here. We will discuss several aspects of human perception important to the differences in displaying images on LDR and HDR displays.

2.1.1 Local Contrast Perception

While we can see a vast dynamic range across a scene, we are unable to see more than a small portion of it within a small angle subtended by the eye. This inherent limitation can be explained by scattering properties of the cornea, lens, and vitreous fluid, and by inter-reflection from the retina. It reduces the visibility of low contrast features in the neighborhood of bright light sources. One example is the difficulty encountered when trying to discern the license plate numbers of an oncoming car at night if the headlights are on. Typical LDR display settings cannot produce the contrast ranges for this to have an effect on the perception of the displayed image. However, it has significant influence on perception of real world scenes, and of images on HDR displays.

Ocular scattering, a well documented phenomenon, depends on a large number of parameters including spatial frequency, wavelength, pupil size as a function of adaptation luminance [51], and age of the subject. This scattering of light has been the topic of numerous studies and is conventionally modeled as an Optical Transfer Function (OTF) in the angular frequency domain and as a Point Spread Function (PSF) in the angular domain. Different researchers [55, 56] have derived models based on various sets of the aforementioned parameters and Vos [75] attempted to unify a number of the existing models. Much of the subsequent work [17, 48] has either validated or built upon his model, largely by considering additional parameters of the model or by optimizing for specific applications.

While different values are reported for the threshold past which we cannot make out high contrast boundaries, most agree that the maximum perceivable contrast is somewhere around 150 : 1. Scene contrast boundaries above this threshold appear blurry and indistinct, and the eye is unable to judge the relative magnitudes of the adjacent regions. From Moon & Spencer's original work on glare [52], we know that any high contrast boundary will scatter at least 4% of its energy on the retina to the darker side of the boundary, obscuring the visibility of the edge and details within a few degrees of

it. If the contrast of an edge is 25 : 1, then details on the darker side will be competing with an equal amount of light scattered from the brighter side, reducing visible contrast by a factor of 2 in the darker region. When the edge contrast reaches a value of 150 : 1, the visible contrast on the dark side is reduced by a factor of 12, rendering details indistinct or invisible. Figure 2.1 shows the model by Deeley et al [17] at several adaptation luminances.

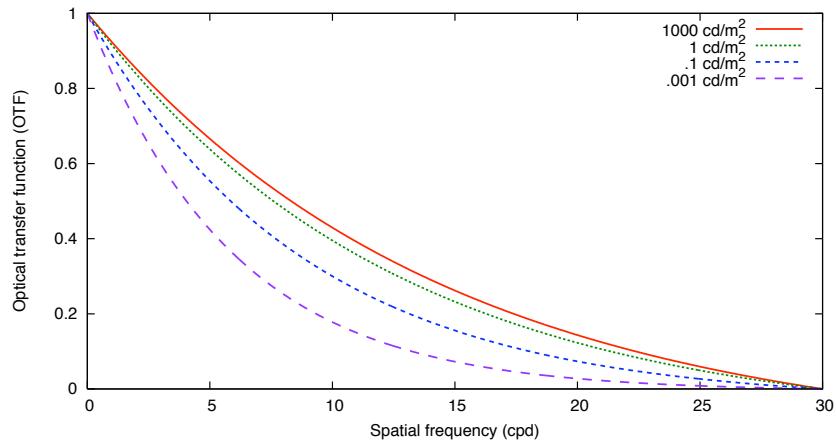


Figure 2.1: Modulated transfer function of the ocular medium at several adaptation luminances.

Just because human observers cannot perceive all details in the presence of high contrast features, one cannot claim high contrast content has no effect – clearly it does. An observer will notice when one region is much brighter than another, both by the challenge it creates in viewing the boundary, and by the accommodation that goes on when shifting from side to side. When the threshold is very large, observers notice a sensation and may even experience discomfort as they attempt to see detail near a bright source. A familiar example for any driver is that a photographic print of a nighttime scene with an oncoming car and headlights is merely an allusion to the real experience – it cannot duplicate the visceral experience of glare, or reproduce the effect it has on a human observer. It is exactly this kind of experience that an HDR display can uniquely reproduce.

HDR display technology described in Section 2.3 only exploits the inability of humans

to see detail in the immediate vicinity of a high-contrast boundary; it makes no assumptions about our overall response to varying brightnesses. Relative (and even absolute) luminances are maintained, and edges will be reproduced exactly when they are below the maximum contrast of the front display of about 250 : 1 in the current production model. Only when this range is exceeded is some fidelity lost near high contrast boundaries, but this effect is well below the detectable threshold, and has not been visible in any experiments [65].

2.1.2 Luminance Quantization

It has long been known that the human visual system does not respond linearly to the luminance of a scene. Stated another way: lightness, the perceptually uniform measure of light, is a nonlinear function of luminance. The human visual system is much more sensitive to changes of low luminance. Given a low intensity Y_d corresponding to a dark scene and a high intensity Y_b corresponding to a bright scene, and some change ΔY , the perceived change in lightness between Y_d and $Y_d + \Delta Y$ will be greater than the perceived change in lightness between Y_b and $Y_b + \Delta Y$.

The psychophysical studies measuring the perception of lightness employ the same design, and focus on the difference, ΔY . The procedure measures the smallest value of ΔY where $Y + \Delta Y$ can be differentiated from Y . This is repeated for different intensities Y and the relation is known as *threshold-versus-intensity* [29] (TVI). It is also commonly referred to as just-noticeable-differences (JND), the unit of lightness, the perceptually uniform function of luminance. A JND is the smallest detectable luminance difference at a given luminance level; adding a JND to a particular luminance level defines the next perceptually relevant step on the luminance scale.

Visual psychologists have studied this phenomenon in depth and have proposed numerous models describing the relationship. Much of the work addresses the more complete relation of contrast perception as a function of lightness and spatial frequency. The work most familiar to computer graphics is from Blackwell for the CIE [13] used by Ward [76], as well as the work by Ferwerda et al [26] in their model of visual adaptation. The Ferwerda curve is shown in Figure 2.2, and includes separate

measurements of the response of the cones and the rods. From the figure, it can be seen that threshold perception of luminance resembles a logarithmic function, but decreases in sensitivity at very low light levels.

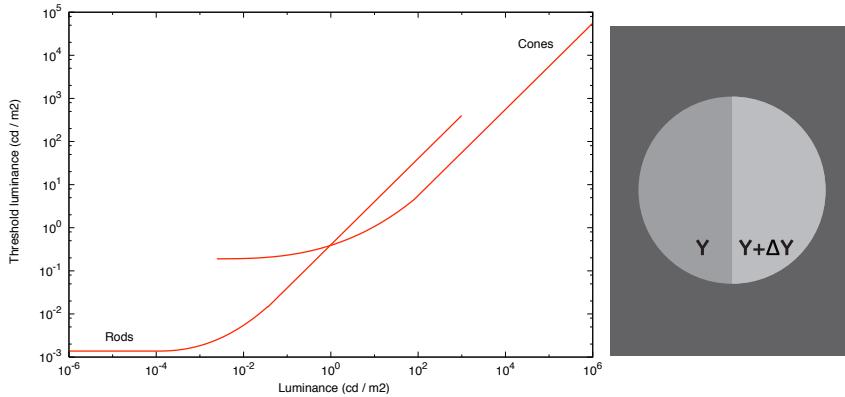


Figure 2.2: The left figure is a plot of the contrast versus intensity curve of the Ferwerda measurements. The left right figure is an example of the test used to determine the threshold differences that can be detected.

A threshold-versus-intensity function describes the quantization sensitivity for different intensities. However, it does not provide a mapping from scene luminances to perceived lightness, and since it represents differential values, it does not provide a mapping function from luminance to perceptually uniform JNDs. Integrating the TVI function provides the function of lightness in terms of luminance relative to some base luminance Y_0 .

There have been many attempts to define perceptually uniform intensity metrics over the years. The most commonly used metric in LDR applications is the CIE 1976 standardization of lightness L^* . It is a nonlinear function of luminance Y relative to a reference white Y_n , where Y and Y_n are, both defined in terms of CIE luminance (CIE 1931 XYZ tristimulus [14] color-matching functions). L^* is used in both the CIELAB and CIELUV [14] color spaces, which target print and video respectively, and L^* models contrasts approximately¹ 100 : 1 and a peak luminance of somewhere around 200 cd/m².

¹We were unable to ascertain the exact method that inspired the formulation of L^* and have been forced to make an educated guess based on targeted applications and indirect evidence. Regardless of the exact function, it is apparent that it is not an accurate fit for larger dynamic ranges.

The equation for L^* is

$$L^* = \begin{cases} 903.3 \frac{Y}{Y_n}, & \frac{Y}{Y_n} \leq 0.008856 \\ 116 \left(\frac{Y}{Y_n} \right)^{\frac{1}{3}} - 16, & 0.008856 < \frac{Y}{Y_n} \end{cases} \quad (2.1)$$

and approximates the response of a 0.4-power function, mapping from a normalized luminance to a value between 0 and 100. The response is plotted in Figure 2.3. A linear segment is included for practical reasons and the break occurs where the function equals an L^* value of 8, corresponding to a contrast ratio of 100 : 1. Obtaining values below 8 is rare in practice and the break is considered the effective limit for video applications, reinforcing the fact that L^* is only applicable to LDR images.

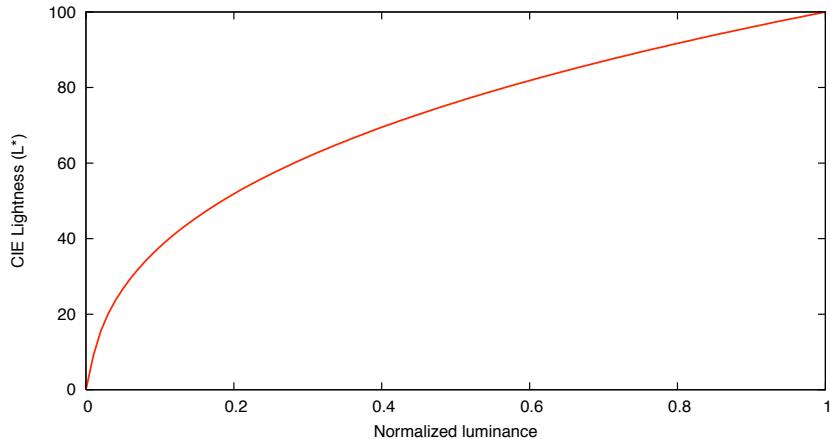


Figure 2.3: CIE L^* curve.

In addition to the numerous LDR lightness functions, several HDR luminance quantizations have been proposed. The two we are aware of are the DICOM standard grayscale display curve [19] and Mantiuk et al's [46] derivation. Figure 2.4 contains plots of both functions compared to a log function.

The DICOM standard is based on work by Barten [9] on deriving an analytic formula for the contrast sensitivity of the human visual system. Barten's original work [7] addressed creating a complete model of the sensitivity of the human visual system as a function of all attributes, including observer-specific values of the eye, luminance level, spatial frequency, temporal change [8], and orientation. From this Barten arrived at a

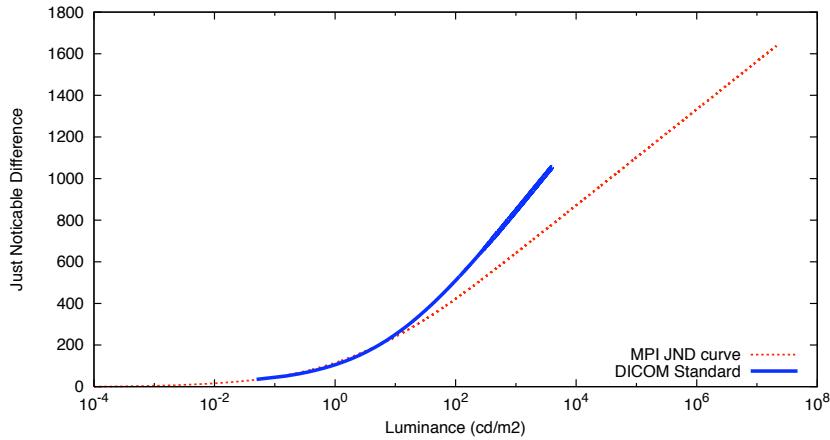


Figure 2.4: Plot of the MPI and DICOM just noticeable differences (JND) curves. The DICOM curve covers a smaller range of luminances but grows significantly faster, consistent with its intended use with LDR devices of difference intensities.

simplified form [10] related to a determined standard observer, which the DICOM standard simplifies further to derive a function of only luminance. The DICOM standard grayscale display curve is defined over a fairly wide range of $0.5 - 4000 \text{ cd/m}^2$, which encompasses a range from the black level of CRTs and up to the reference white of lightboxes, and has been validated in perceptual experiments. Regrettably, the DICOM standard was designed to address LDR output devices operating at different luminance levels. The HVS perceives contrast differently at different intensities. To ensure that radiological images were viewed properly and doctors did not draw different conclusions based on the brightness the display device, DICOM added a modification similar to the tone scale alteration described in Section 2.4.1.

In their SIGGRAPH 2004 paper, Mantiuk et al [46] describe a different luminance quantization. Working directly from the threshold-versus-intensity results [13, 26] described above, they solve the differential equation mapping the TVI measurements to luminance values and numerically invert it to yield a lookup table mapping luminances to JNDs. Their results cover the full range of the TVI curves, covering luminances from 10^{-4} cd/m^2 to 10^8 cd/m^2 . While their work addresses real scenes, and does not include any modification to make image appearances luminance-invariant it has not had any formal perceptual validation.

Perceptual luminance quantization has important implications for the design of imaging systems. Because the smallest change an observer can detect is 1 JND, it is redundant to provide additional display driving levels in the space of 1. In case of LDR displays, which have a limited number of driving levels, the display response is adjusted to match perceptual quantization, as discussed in Section 2.4.1. HDR displays do not suffer from the same problems, as described in Section 2.3.

2.1.3 Visual Difference Prediction

Many fields, such as video editing and print design, require the accurate portrayal of images. When researching and designing systems for these fields, creators desire the ability to simulate the characteristics of their designs prior to production, and verify them afterwards. Traditional metrics, such as least squares error between images, are exceedingly poor metrics of perceived difference. Human perceptual sensitivity is a very complicated process, and has components that greatly vary, depending on the feature in question. There is a desire to have methods that can model the differences that a human observer can perceive between the original image and the image reproduced by an imaging system.

The solution for accurate modeling of the HVS comes from a combination of two separate areas of research. On one end, work, such as the research of Barten [7], has been conducted on modeling of contrast and spatial sensitivity of the human visual system, building upon the aspects of perception presented above. On the other end, work has been conducted on defining color appearance models that describe how we perceive color. The basic CIELAB and CIELUV [14] attempts at perceptually uniform color, have given way to CIECAM97 [2] and CIECAM02 [53], which consider such effects as background, surround effects for simple environments. The combination of these two areas of research are full image appearance models such as Fairchild et al's iCAM [23] and Pattanaik et al's multiscale observer model [59], which describe effects of both areas and are designed to address high dynamic range images.

While image appearance models can render images in a similar manner to the HVS, they aren't sufficient for comparing differences. Even though images are transformed to

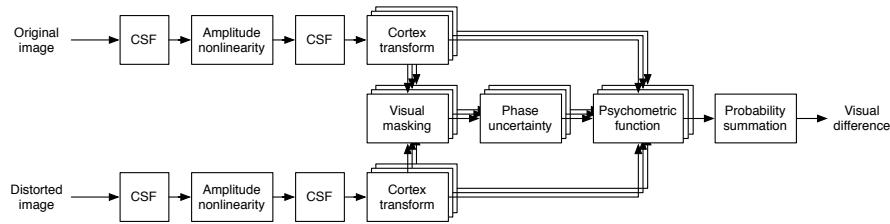


Figure 2.5: Inputs and resulting output from each stage of the VDP process.

model the visual system, we cannot simply compare pixels. The probability of detecting differences in perceived images is equally complex as the simulation of the perceived images. This motivates the development of metrics that can account for complexity. The method then, is to take the original image and the distorted image to be compared, process both with some form of image appearance model to transform them into something the observer would perceive. Then, it compares them using a function that mimics human detection mechanisms, usually based on some form of spatial frequency hierarchy such as Gabor pyramids [41], to obtain the perceived difference. The combination of an image appearance model with a set of perceptually-based detection mechanisms forms a *visual difference*.

While several such models exist, two of the most popular are the Visible Differences Predictor (VDP) by Daly [15] and the Sarnoff Visual Discrimination Model [42]. However, to our knowledge, only one visual difference metric exists for HDR images: the work by Mantiuk et al [47] extending the VDP to HDR images, and its subsequent calibration [45]. We use their high dynamic range visible differences predictor (HDR VDP) as the basis of our validation in Chapter 5. There we describe how we apply the HDR VDP to verify our results.

The HDR VDP consists of the two parts described above. In the case here, the first part has 3 phases. It first applies an optical transfer function (OTF), then applies nonlinear luminance quantization to express the image in JND units, and finally filters each image with a contrast sensitivity function (CSF) such as the ones described by Virsu et al [74] or Barten [7]. The second part has 4 phases. First, it applies the cortex transform [79], then it adjusts the images to account for visual masking and phase

uncertainty, weights the inputs based on a psychometric function, and finally combines the probabilities to get the visual differences. We introduce these components in the following paragraphs and discuss them further in Chapter 5.

The majority of the HDR VDP modifications occur in the image appearance modeling phase. The original VDP only targeted images of limited contrast, and as a result it did not address the scattering of light due to the ocular medium. It does not include a model the optical transfer function described in Section 2.1.1, which the HDR VDP performs as a first step. While the original VDP did model luminance quantization, and accounted for the change in the function due to different adaptation luminances, it still operated on relative luminances and assumed a maximum dynamic range of the images. The HDR VDP replaces this with the absolute, JND-scaled luminance quantization described in Section 2.1.2. The HDR VDP derives a quantization from the contrast sensitivity function similar to the method of Daly. The CSF of the original VDP varied with adaptation intensity, but due to the limited dynamic range the function was only evaluated for one luminance. The HDR VDP must account for multiple CSFs in the same image due to the range of luminances present. As an optimization, they prefilter the images by the CSF of many luminance levels, then blend the values based on the adaptation at a given pixel.

Because the input to the detection mechanisms is a perceptually linearized image, no change is necessary; the only difference is a scale factor to change from a normalized relative unit to a JND-scaled unit. The cortex transform, which models the orientation-sensitive cells in the visual cortex, decomposes each image into a spatial frequency hierarchy which is further filtered by orientation. The modeling of masking and phase uncertainty accounts for the contrast scaling of the difference between the two images relative to the original signal it modifies, while the the psychometric weights all of the inputs based on a psychophysical model of contrast sensitivity. Finally, a product series of all of the images of the spatial frequency hierarchy computes the probability of detecting the distortion for each pixel.

2.2 Tonemapping Operators

For any image representation to seem realistic, it needs to evoke the same response in the visual system as did the original. This challenge goes beyond computer graphics, it is a problem fundamental to any media and familiar to both artists and photographers. The intensity and contrast of real scenes vastly exceed the range that can be produced by canvas, photographic print, or by conventional computer display. A simple linear rescaling of the luminance values is insufficient, and more complicated mappings of luminance, collectively known as tonemapping operators, are required.

Tonemapping operators have been the traditional method of displaying high dynamic range images, and the only available means prior to HDR displays. The first research was done by Oppenheim et al [57] in 1968, while the first operator to explicitly address HDR images was that of Miller et al [49] in 1984, who attempted to introduce topics in computer graphics to the field of lighting design. Operators were first introduced to computer graphics by Tumblin et al [73] in 1993.

Numerous fields have been confronted with this problem and derived different methods tailored to their needs to, and address issues, as a result the means in which tonemapping operators reduce the dynamic range varies. The various methods draw inspiration from many different aspects of images, such as making assumptions on reflection, paying attention to how artists have overcome the challenge, or by emulating portions of the HVS. Even with this variation, two basic classes of methods exist: global operators and local operators. Global operators consider overall properties of the image (such as using histogram data) and apply the same function to every pixel of the image. Local operators consider properties of a local neighborhood, and vary the function for a given pixel accordingly. This implies that global operators will map two pixels of a given luminance to the same intensity regardless of their location, while local operators could map the same 2 pixels to different intensities depending on neighboring pixels. Table 2.1 contains a list of all tonemapping operators we cite in this thesis, while Figure 2.6 compares the results of several.

| | | |
|-------------------------|--|---|
| Global Operators | Miller et al. Tumblin et al. (1993) Ward (1994) Ferwerda et al. (1996) | [49] [73] [76] [26] |
| Local Operators | Oppenheim et al. (1968) Chiu et al. (1993) Pattanaik et al. (1998) Ashikhmin (2002) Durand & Dorsey (2002) Fattal et al. (2002) Reinhard et al. (2002) Fairchild & Johnson (2002) | [57] [12] [59] [4] [21] [25] [61] [22] |

Table 2.1: Table of tonemapping operators cited in this thesis

2.2.1 Taxonomy of Operators

In this section, we discuss the two classes and describe some of the operators that provide insight into processing images for HDR displays. A full overview is beyond the scope of this document, but other resources provide excellent coverage. Devlin [18] provides a comprehensive overview of techniques up to 2002, and as Reinhard et al's book [62] covers the majority of tonemapping operators and provides source code of the implementations. We focus on operators that make use of aspects of human perception or that share similarities with the processing of HDR images for display.

Global operators, as stated, modify each pixel based on global characteristics of the image. They are the faster of the two classes of operators because the amount of information they consider is fundamentally limited. The core idea is to create some mapping from HDR to LDR that roughly corresponds to how our visual system responds to luminance, hopefully preserving the same details. They can only handle limited dynamic ranges because they are effectively forced to be monotonic. Since they cannot smoothly alter the surrounding area like local operators, any global reverse of gradients would introduce undesirable discontinuities.

Miller et al [49] employ a function to map scene intensities to preserve perceived



Figure 2.6: Selection of tonemapping operators applied to images. The Ward [76] (top), Reinhard et al [61] (middle), Durand & Dorsey [21] (bottom) tonemapping operators applied to two sample images. (Left image courtesy of Greg Ward.)

brightness ratios. The function, intended for displaying images of indoor scenes, was derived from work by Stevens & Stevens [69] and was only defined correctly up to values of about 1000 cd/m^2 . Tumblin et al [73] took the same brightness function and modified it to preserve the brightness values directly, as opposed to ratios thereof, resulting in a more usable operator.

Ward [76] and Ferwerda et al [26] take a different approach. They use threshold-vs-intensity (TVI) measurements, discussed in Section 2.1.2, to derive luminance quantizations which model the perception of lightness in terms of just-noticeable-differences ($JNDs$). Ward bases his operator on the contrast sensitivity data collected by Blackwell [13] on photopic viewing conditions, while Ferwerda et al use different data, but extend their operator to model both photopic and scotopic viewing conditions. These

JND values are used to perceptually linearize the input image, quantizing it so only the information regarding perceivable changes is retained. Differences are preserved but none of the limited display steps are wasted on details which are undetectable by the HVS.

Local operators, on the other hand, preserve local contrast while still reducing it globally. In addition to considering global image characteristics, these tonemapping operators take the local neighborhood of a pixel into consideration when determining its value. The result is often a more effective reduction of dynamic range, especially for images with extreme contrasts, but this effectiveness comes at a higher computational cost.

Reinhard et al [61] note that photographers have overcome the dynamic range limitations in the photoprinting process and mimic many of the conventional photographic techniques. Their photographic tonemapping operator makes use of the Zone System [1] to map a range of intensities into a lower dynamic range while preserving texture detail across the entire range, then mimics the dodging and burning² of developing photographic prints with different sized blurred versions of the image to further decrease contrast around bright and dark areas.

Many perceptually-based local operators are derivatives of the image appearance models [23, 59], discussed in Section 2.1.3, opposed to operators that purely address dynamic range reduction. Compared to pure tonemapping operators, image appearance models include elements of the human visual system, such as modeling ocular scatter to add blooming around bright light sources, that may degrade the resulting images. This seemingly undesirable decision can be explained by Spencer et al's [67] observation that viewers get a better impression of the luminances and dynamic range in tonemapped image if it includes their own perceptual shortcomings. To use image appearance models for tonemapping, the appearance model is first applied to the original image containing scene luminances and results in the simulation of perceived image. To display the image, the parameters of the output device are input to the inverse of the model which is then

²Dodging and burning refer to the selective over- and under-exposure of areas of the print relative to some base exposure. These techniques serve to reduce global contrast in the image produced.

applied to the result of the first step. Pattanaik et al's [59] multiscale observer model is often referred to as the most complete example, containing all elements of human vision understood well enough to be modeled, while Ashikhmin's [4] operator is similar but only considers portions relevant to dynamic range reduction.

Finally, there are two operators which share important features with the image processing for HDR displays. Durand & Dorsey's [21] bilateral filter is an edge-preserving smoothing filter that removes large-scale luminance differences, but preserves details by separating the image into a base luminance layer and detail layer. This separation of base and detail layers is the same general methodology discussed in Section 3.2, but differs in what it does with those layers. Chiu et al's [12] work divides the original by a blurred version of the image, discarding large-scale luminance differences but retaining details. The HDR display performs a similar operation optically, so the image processing needs to account for this. We modify the LCD panel image to correct for intensity discrepancies with the low-resolution backlight. In both Chiu's operator and in our work, this results in reverse gradients around areas of high luminance, where the dimmer side of the high-contrast boundary is further darkened. While this effect is undesirable in a tonemapped image, it is beneficial when processing images for display.

2.2.2 Validation

While tonemapping operators have been in use for a considerable period of time, work has only recently begun on verifying how accurately they perform the task of replicating the visual representation of images. The first work on the subject was by Drago et al [20], who performed a study where users assigned a value to the similarity of two tonemapped images and rated the images on how natural they appeared by preference. Park and Montag [58] evaluate tonemapping operators for use on HDR scientific images, asking users to rank operators on their opinion of scientific usefulness in addition to preference, and measured the effectiveness of different operators for various tasks. Kuang et al [36] and Fairchild et al [24] both studied user preference between operators to create rankings of their accuracy. They made the important observation that users prefer images that are more colorful and contain more contrast than is natural, implying

that studies based on preference have limited ability to determine the accuracy of tested operators.

More recent studies have moved away from judging user preference of operators. Yoshida et al [80] asked users to rank the accuracy of operators by comparing tonemapped images to the real scenes. Ledda et al [40] concluded that tonemapped images might not be similar enough to real scenes to obtain meaningful relations from operator comparisons. Instead, they ask users to compare the results of tonemapping operators to images on an HDR display, which they previously demonstrated [39] was an accurate depiction of real scenes. Validation is still being actively investigated, and has recently gotten attention outside of academia, which in turn resulted in the formation of the CIE technical committee TC8-08 to study tonemapping operator validation [32].

2.2.3 Shortcomings of Tonemapping

The goal of tonemapping operators is to faithfully reproduce the visual representation of an image in an output medium that is not capable of directly representing the intensities or dynamic range of the original. While they succeed in depicting more visual information than by not using one at all, they cannot completely realize the goal. Conventional displays are too limited to convey images of real scenes with complete accuracy. For the range of luminances found in indoor scenes, the same range covered by HDR displays, JND metrics predict over 1000 discernible values. Conventional output mediums can only reproduce about 25% of those values, resulting in a significant loss of information.

Furthermore, there are perceptual and psychophysical effects that depend on intensity alone. The sensation one feels when the pupil contracts in the presence of a bright light cannot be mimicked through any image processing. While a tonemapping operator could show details in all areas of an image of a car and headlights at night, no one would confuse it with the original. Tonemapping operators can mimic processes of the HVS to deliver more information, but cannot reproduce the visceral experiences of the original scene luminances.

2.3 HDR Technology

In a conventional LCD, two polarizers and a liquid crystal are used to modulate the light coming from a uniform backlight, typically a fluorescent tube assembly. The light is polarized by the first polarizer and transmitted through the liquid crystal where the polarization of the light is rotated in accordance with the control voltages applied to each pixel of liquid crystal. Finally, the light exits the LCD by transmission through the second polarizer. The luminance level of the light emitted at each pixel is controlled by the polarization state of the liquid crystal. It is important to point out that LCDs cannot completely prevent light transmission - even at the darkest state of a pixel, light is emitted and as such the dynamic range of an LCD is defined by the ratio between the light emitted at the brightest state and the light emitted in the darkest state. For a high end LCD, this ratio is usually around 300 : 1, with monochromatic specialty LCDs (e.g. those for medical imaging) going up to 700 : 1. The luminance level of the display can be easily adjusted by controlling the brightness of the backlight, but the dynamic range ratio will remain the limiting factor. In order to maintain a reasonable 'black' level of about 1 cd/m^2 , the LCD is thus limited to a maximum brightness of about 300 cd/m^2 .

The fundamental idea of the HDR display is to use an LCD panel as an optical filter of programmable transparency to modulate a high intensity but low resolution image from a second display. For example, assume we have any display with a contrast range of $c_1 : 1$ between the darkest and the brightest intensity producible by that display. If we now put an LCD panel with a contrast ratio of $c_2 : 1$ in front of the first one, then the (theoretical) contrast of the combined system is $(c_1 \cdot c_2) : 1$. Two different versions of HDR displays have been constructed around this principle: one using a projector as the rear display, and one using a diffused grid of LEDS as the rear displays.

In practice, the first display needs to be able to produce a very high intensity image, because color LCD panels only have a transparency of about 3-8%, even when switched to 'white', so that most energy is actually absorbed. Another reason for using a display with a very high base intensity is that a lot of the HDR images we would like to show have, very bright regions in them.

For reasons discussed below, the projector version of the HDR display is mostly a

prototype and there are no plans for a production model. However, it is slightly simpler in its design and provides an excellent introduction to the LED-based design that is used in the HDR display.

2.3.1 Projector-based Display

For the projector-based HDR display [64], the backlight and the first modulator are combined into a single DLP using a Digital Mirror Device with a dynamic range of about 800 : 1. The three central components of the HDR display are then the projector, the LCD and the optics that couple the two. Using these components, each image on the HDR display is the result of modulated light coming from the projector which is directed onto the rear of the transmissive LCD by the optics system, modulated a second time by the LCD, and properly diffused for viewing. Figure 2.7 contains a photograph and diagram of the internal construction.

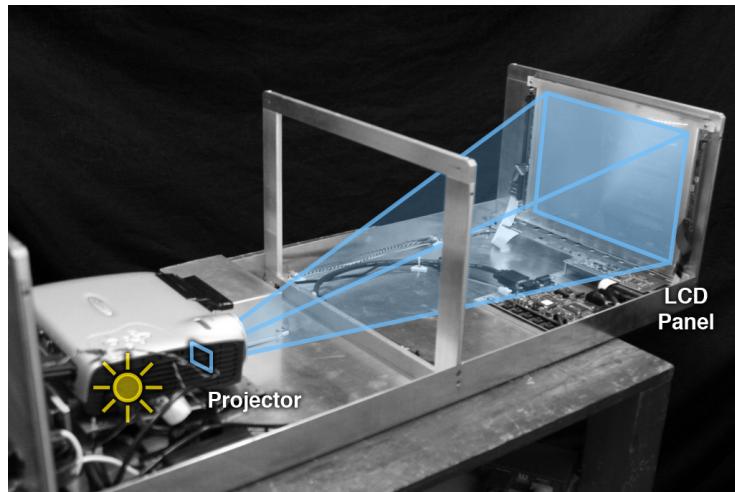


Figure 2.7: Internal schematic of projector display.

To reduce unnecessary light loss, the color wheel of the projector has been removed, resulting in a monochrome display system with a roughly threefold increase in brightness due to the absence of the color filters. New control electronics have been integrated into the commercially available projector to re-synchronize it in absence of this color wheel.

The LCD panel has been separated from the conventional backlight and all of the optical layers behind the display have been removed to create a transmissive image modulator.

The optics used in the HDR display include the conventional projection lens of the projector, and a Fresnel lens directly behind the LCD display to collimate the projected light into a narrow viewing angle for maximum brightness of the HDR display and to avoid color distortion due to diverging light passing through the color filters of the LCD. Finally, a standard LCD diffuser was used to redistribute the collimated light into a reasonable viewing angle.

All three components have been installed in a single housing with appropriate alignment mechanisms to create a close matching of the DLP and LCD pixels. The alignment can be fine-tuned through the controls of the DLP projector. However, a perfect match is impractical as alignment at the sub-pixel level is exceedingly hard to achieve and maintain. To avoid moiré patterns and alignment artifacts associated with even a minor misalignment, the projector image has been deliberately blurred. As described in the following section, compensating for that blur in the LCD image is a key component of processing images.

Using this configuration, the light output of each pixel of the HDR display is effectively the result of two modulations, first by the DLP and then by the LCD pixel, along the same optical path. The upper boundary of the dynamic range results from full transmission of both pixels (i.e. the 255th level on both modulators), and the lowest boundary from the lowest possible transmission of both modulators (i.e. the 0th level on both modulators). Since the DLP has a dynamic range of 800 : 1 and the LCD a dynamic range of 300 : 1, the theoretical dynamic range of the HDR display is 240,000 : 1. Imperfections in the optical path introduce noise that reduces the dynamic range to a measured 54,000 : 1. The luminance values matching these boundaries are a result of the brightness of the projector and the transmission of the LCD.

In this case, the projector is rated at 1200 Lumens, or approximately 3600 Lumens once the RGB color filters are removed (since each filter for red, green and blue eliminates approximately 2/3 of the incoming light). The particular LCD panel used has a measured transmission of approximately 7.6% in the white state (this is quite high for an LCD since even the theoretical maximum for a color LCD without any losses is only 16% due to

the light reduction of 50% at the polarizer and another 66% due to the RGB color filter). Assuming that the light emitted by the HDR display is diffused across a solid angle ω , the maximum luminance is then given by:

$$L_{\max} = \frac{\Phi_{\max}}{A\omega}, \quad (2.2)$$

where A is the area of the LCD and Φ_{\max} is the maximum outgoing flux. In the HDR display prototype, the flux is approximately 182 Lumens (2400 Lumens \times 7.6%). The area A is the area of the 15 in LCD (697 cm^2) and the solid angle of diffusion ω is approximately 0.66 sr (40° diffusion horizontally, 15° vertically). The maximum luminance for this particular configuration is then approximately 3956 cd/m^2 . The actual measured peak luminance was 2700 cd/m^2 Lumens. The theoretical minimum luminance is less than 0.01 cd/m^2 , while measurements yielded a value of 0.05 cd/m^2 . Clearly, a shift of this range toward even higher luminance values would be possible with a brighter projector or with a more transmissive LCD. Unlike a standard low dynamic range display, even an order of magnitude increase of the maximum luminance would not significantly reduce the quality of the ‘black’ state since 1 cd/m^2 is still a very satisfying ‘black’, especially if other parts of the image contain very high luminance values.

Within that luminance range, a very large number of different combinations of output settings for the DLP and LCD can be achieved. If both systems were linear 8-bit devices then the total number of combinations would be 256^2 , over 17 000 of which are distinct. Due to the nonlinear gamma of each system, the actual range of distinct addressable steps is different, but still significantly larger than what is needed to display the 962 JND steps necessary to provide all visible and distinguishable luminance steps in the measured luminance range of the system (including all losses) of 0.05 cd/m^2 to 2700 cd/m^2 .

High power consumption and the resulting thermal management requirements are a consequence of the image creation mechanism inside the projector. Unlike a cathode ray tube (CRT) display, where light is created only in the regions of the image that are supposed to be bright, an LCD or DLP projector creates a uniform light distribution that is then modulated by the LCD or DLP mirror chip. The power consumption of an LCD or DLP projector is thus independent of the image and always very high as there has

to be enough light produced by the lamp such that a full screen ‘white’ can be shown. Combined with the low modulation efficiency of the LCD or DLP this causes the high power consumption. In the HDR display the situation is worse than in a conventional, single-modulator display. The lamp of the projector has to emit enough light to allow a full screen image at the highest possible brightness of the HDR display. To achieve $10\,000\,cd/m^2$ on a 15in screen we would need an outgoing flux of approximately 500 Lumens (see Section 2.3.1). Even with a very high transmission LCD this requires at least 5000 Lumens to be emitted from the projector. In the prototype presented in Section 2.3.1 the color wheel/filter of the projector has already been removed to reduce the losses in the projector but even so the modulation efficiency of the projector is slightly less than 50%. The lamp thus has to produce in the order of 1000 Lumens. Yet, in almost all HDR images the area that is actually at such a high brightness of $10\,000\,cd/m^2$ is very small.

In fact, a random selection of 100 HDR images indicated that average HDR images have less than 10% of the image content in the high luminance range (above $3000\,cd/m^2$) and that the average luminance over all images was less than $800\,cd/m^2$ for indoor scenes and $2100\,cd/m^2$ for outdoor scenes. The projector HDR display consequently creates a factor of between 12.5 and 4.75 too much light at any given time. As seen in the discussion of the projector-based HDR display in Section 2.3.1 there are significant obstacles to overcome.

To realize the dream of television or computer displays presenting images that look indistinguishable from the real world, it is not sufficient to merely show images with the appropriate luminance range and resolution; it is also necessary to make a commercially viable system that achieves these higher quality images within the hardware and software infrastructure and market price points of today. The version of the HDR display described in this section retains the high image quality of the projector display and overcomes the commercialization barriers: power, thermal, cost and form factor.

2.3.2 LED-based Display

As mentioned in Section 2.3.1 and discussed in Section 3, it is possible to compensate for a low resolution of the rear image of the HDR display. It is important to realize that this correction works properly, as long as the local image contrast does not exceed the dynamic range of the front modulator. From the psychophysical theory presented in Section 2.1.1 we can establish the largest size of a rear image pixel. A second version [11, 71] of the HDR display uses light emitting diodes (LED) at the largest possible size allowed by the veiling luminance effect that has been validated previously through experimental tests [65]. Figure 2.8 shows the current generation of LED-based HDR display, the BrightSide DR-37P.



Figure 2.8: Photograph of BrightSide DR37-P.

The production version has been constructed using Seoul Semiconductor 2.5 Watt white LEDs (PN W10290) on a 18.8mm hexagonal close-packing matrix where each LED is individually controlled over its entire dynamic range with 256 addressable steps. 1380 LEDs have been mounted behind a 37in Chi Mei Optoelectronics V370H1-L01 LCD panel with a 250 : 1 simultaneous contrast ratio³ and 1920 × 1080 resolution. For a full white box occupying the center third of the screen, the maximum luminance is

³Display manufacturers often employ various methods of distorting the calculation of dynamic range, such as altering room illumination between measurements. The ANSI 9 checkerboard provides a standard measure of the usable display dynamic range, which we use to determine this number.

measured as 4760 cd/m^2 . For a black image, the minimum luminance is zero, since all LEDs are off. The minimum luminance is less than 6 cd/m^2 on a ANSI 9 checkerboard (the VESA contrast standard). And like the projector display, while not every pair of driving values of the LEDs and LCD panel results in a unique luminance, the approximately 17 000 unique luminances that can be produced is significantly larger than the 875 JNDs predicted.

2.4 Display Calibration

While there are many area-specific display calibration requirements, such as those made by medical imaging and film production, they all share some common traits. For LDR display devices, the simplest approach is to alter the image values to compensate for a device with a nonlinear response, adjusting the input so the output is linearized. However, this alone is insufficient and there are other factors that must be considered in calibrating displays. The properties of human perception make it a much more subtle problem. We will analyze traditional calibration practices, explain their motivation, and discuss which portions still apply to HDR displays.

2.4.1 Gamma

Any discussion of display calibration eventually involves a discussion about gamma, one of the most misunderstood topics in electronic imaging. It has been adapted to serve many roles simultaneously, obscuring its original purpose. The primary considerations in the creation and calibration of displays are to minimize quantization on a lossy (8 bit) channel, to linearize the display response, and to account for the change in perception of the observer to maintain rendering intent.

Minimize Quantization. A key question to ask in designing any display system is how many distinct input/output levels are necessary to cover the desired range without banding or similar quantization artifacts? As described in Section 2.1.2, human perception of lightness is nonlinear, and for practical purposes in imaging, it is stated that we can detect

1% differences in luminance. Covering a range of 100 : 1 (near the maximum effective contrast of a conventional LDR display) with an increment of 0.01, as required to avoid quantization in the areas with highest sensitivity, requires 10 000 values, or roughly a 14-bit representation. If covering the range with a ratio of 1.01, it takes roughly 460 values, or 9 bits. Based on other factors affecting our perception, 8 bits are used in practice, and this quantization is the primary factor in the design of LDR imaging systems. This fact is incorporated into the design of all optoelectric transfer functions (OETFs), such as the television standard Rec. 709 [31] and the computer standard sRGB [70], which are similar to L^* described in Section 2.1.2. All three curves are plotted in Figure 2.9.

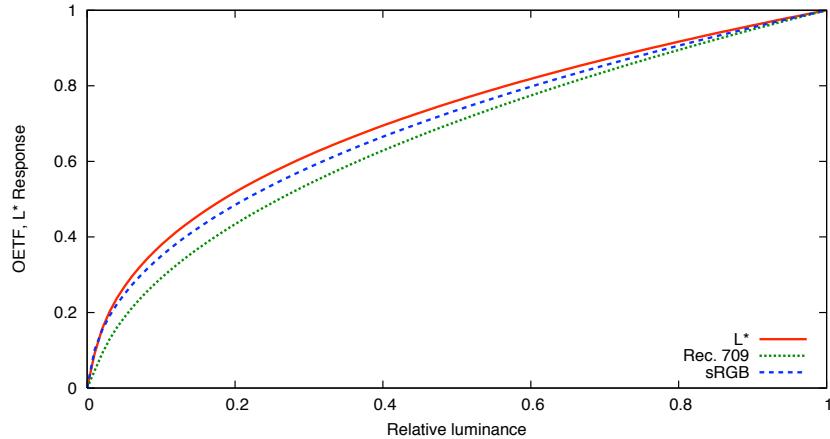


Figure 2.9: Comparison on L^* , Rec. 709, and sRGB OETFs.

Recently, Muka & Reiker [54] have argued that, for conventional displays with a typical dynamic range of 300 : 1 or so, an 8-bit representation of images is sufficient for medical diagnosis. They argue that the difference between an 8-bit digital display and a 10-bit or higher bit depth is minimal, and perhaps not noticeable at all. However, as the range of displayable luminances increases, so does the number of JND steps required to cover that range, which is reflected in the numbers presented in Section 2.1.2. If the original medical data was of a bit depth of 10-bit or greater, an HDR display would be able to display the additional data, if combined with proper image processing techniques, such as work by Ghosh et al [27]. They process volume data to preserve the additional HDR information, and subsequently use it to tune image presentation to extract key

features.

Linearize Response. Regardless of whether images are encoded linearly or not, in order to display images without artifacts, the output device must have quantization characteristics paired with human lightness sensitivity. We could make a display with a linear response capable of outputting 14 bits of driving levels. However, this is very difficult, if not impossible, to manufacture and most of those driving values would be wasted because of the sensitivity of our visual system. Ideally, we want a display that produces relative luminances with a response that is the inverse of the lightness perception of the HVS. For example, on the low end of human perception, small changes in luminance cause a relatively large changes in lightness, so we would need a display such that a large change in driving value on the low end would cause a relatively small change in luminance.

This relationship is almost exactly the case with conventional displays [60]. In the case of CRTs, this is the result of the physics of the electron gun. CRTs have a response proportional to a power between 2.35 and 2.55, roughly the inverse of the 0.4-power of our lightness perception of LDR images. In LCDs, plasma display panels (PLPs), and digital light processors (DLPs) this is accomplished by a lookup table in the display controller that mimics the inverse response of human vision. The important thing is not that the display response is the inverse of the power relation of the OETFs, but that it has a response that is roughly the inverse of human lightness perception, because that is what the OETFs model. The inverse signal distributes values in a way that minimizes the quantization of the signal in perceptual terms, having equal spacing of the lightness of the values, not the luminance.

Rendering Intent. From Section 2.1.2, it can be observed that the perception of lightness is a function of the luminances and the contrast range being addressed. For the luminances and contrasts of conventional displays, the closest match was found to be a power function of $Y^{1/3}$, while for real scenes it was found to be closer to a $\log(Y)$ function. Our perception of contrast is dependent on the scene intensity, to the point that a lower contrast image shown at a higher luminance can appear to have more contrast

than another image having more contrast but shown at a lower luminance.

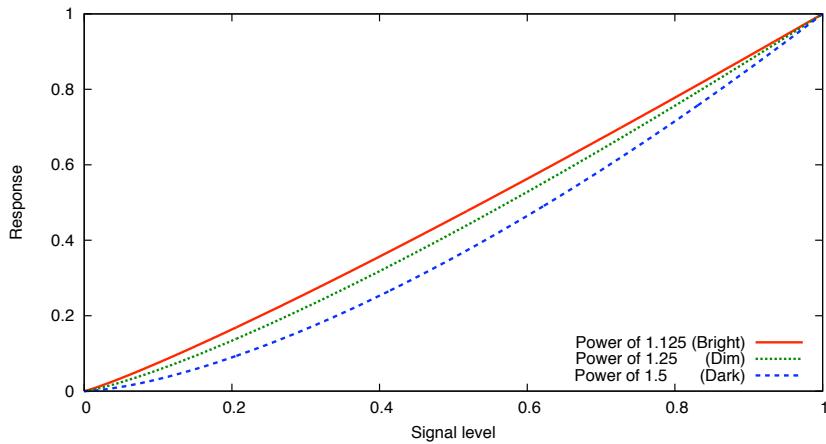


Figure 2.10: Tone scale curves for different intensity surroundings.

Another important attribute of human perception is the effect of the surround, the ambient level of light in the environment on image perception when viewing a display. From the work of Bartelson and Breneman [6], it is known that the light level of the surround has a significant impact on adaptation, and thus lightness perception. The surround luminance has a direct impact on viewer adaptation and the contrast they perceive. Decreasing the surround luminance reduces the contrasts perceived by an observer.

These two observations have implications for reproducing images at different intensities. The designer of the OETF must compensate for a change in the perceived contrast of the reproduced image, known as tone scale alteration. For conventional displays, along with a specified peak display luminance for an OETF (such as the SMPTE standardization [66] of 103 cd/m^2 for studio video monitors), there is also a specified ambient luminance. Together these two values determine the required amount of tone scale alteration, and is one of the many reasons for the existence of multiple standards for OETFs such as REC 709 [31] and sRGB [70].

The OETF has to be matched to the ambient luminance. In order to do so, imaging systems incorporate an additional power term into the OETF to accomplish the tone scale alteration. This additional term purposely mismatches the previously paired display

response with the appropriate OETF. The combination of the OETF encoding and the display decoding exponent causes nonlinear outputted luminances, and is referred to as the end-to-end power. Work has been done by Antwerp [3] and Barbier [5] on sensing the ambient luminance level and adapting the tone scale to match, but the fact remains that some form of alteration is always required for conventional displays.

As an example, consider viewing the same image in two different environments: a theater and an office. The theater has an ambient luminance of around 5 cd/m^2 and is considered a dark viewing environment. The office environment easily has an ambient luminance of 200 cd/m^2 and is considered a light viewing environment. In the case of the film, the encoding step is the recording of scene luminances to the negative film when shooting, and the decoding step is the copying of values to the slide film used in the projector. Projector film has a decoding exponent of 2.5 while negative film has an encoding exponent of 0.6. This results in an end-to-end exponent of 1.5, suitable for the dark viewing environment. In an office environment, the CRT or LCD has the same decoding exponent of 2.5, but the sRGB [70] specification states an encoding gamma of 0.45. As a result, the office display has an end-to-end exponent of 1.125, suitable for the light viewing environment.

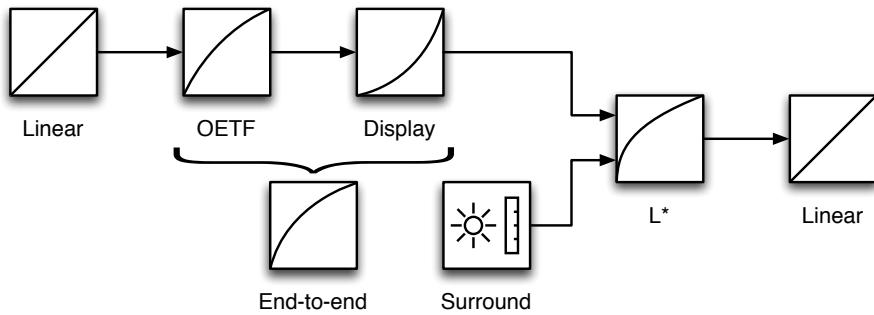


Figure 2.11: Flowchart of the different aspects considered in the design of a gamma curve. Starting with linear scene luminances, an OETF encodes the luminance values. This OETF is paired to the display response, and the difference between these is known as the end-to-end encoding power. The end-to-end power is chosen to match the current perceptual response to luminance, determined by the surround luminance. All of these components work together to produce a perceived image that is perceptually linear and without artifacts.

OETFs serve many purposes simultaneously, and because the math is associative, many different corrective measures can be collapsed into a single formulation. Additionally, since most OETFs contain a linear segment near zero for various practical purposes, the offset means that the exponent one sees in the equation is not the exponent of the curve that most closely approximates the OETF. Combined, these factors lead to the confusion seen surrounding the concept.

2.4.2 Implications for HDR displays

Together, the three concepts detailed above form the basis of calibration curves for LDR display devices. The different portions can be measured and calibrated more exactly than what we have described, but the basic functions do not change. With an understanding of the considerations involved in designing LDR imaging systems and the formulation of OETFs to match the display device and the viewing environment, we can now consider the portions pertinent to HDR image processing.

Nonlinearly encoding data is only necessary when the number of addressable driving levels of the display device do not provide a luminance quantization equal to or less than the smallest luminance quantization perceivable by the observer. The number of JNDs in the range of intensities of a conventional display is roughly equal to the number of driving levels of the display device, but the quantization is not uniform and has more resolution at low luminances. This requires modifying the output system to accommodate the luminance quantization of the viewer, motivating the 2.5 exponent in conventional display responses. In the case of the HDR display, the number of driving levels is significantly larger than the number of JNDs. We can address luminances with linear encoding, as long as the quantization level of the display is the same or less than human lightness perception. Compared to conventional displays, this is more difficult to show mathematically due to the dual-modulator configuration, but studies show that it works in practice [65].

There is still a need to invert LCD response to produce linear light. The LCD panel and LED are inherently tied together in their representation of the image, and they must work in the same linear space. This calibration proceeds the same way as the standard

measurement method and produces a lookup table (LUT) to invert the values. It is also useful because the LCD displays the fine changes in detail, and these are output with a quantization suited to human vision by virtue of being sent through the LCD panel. No additional calibration is necessary for the LEDs since the display hardware linearizes their output.

In our work, we assume the effect of the surround on lightness and contrast perception is negligible and do not address it. As stated, we only concern ourselves with luminances within the range of the directly reproducible values (see Chapter 3). The display has a high enough peak luminance that it can reproduce real scenes without scaling down the values, and does not require tone scaling to adjust for the loss of contrast perception at lower luminances.

Additionally, the HDR display has a high enough peak luminance and a large enough dynamic range that contrast and color perception should not significantly differ since the luminances are roughly the same. The display is bright enough to drive viewer adaptation in dim viewing conditions, lessening any effect of the surround. The higher peak intensity of the display means that the viewing conditions play a relatively smaller role in the perceived image. Much of the studies into accurate image reproduction on LDR displays have been concerned with altering images and display characteristics to accommodate the present viewing conditions. This is not a complete solution, as the surround still does contribute some to the viewer's perception. In some cases, it may be necessary to take active control of the room illumination, similar to Ghosh et al [28], to ensure that the images are perceived as intended.

Chapter 3

Processing Algorithms

This chapter details the primary contribution of the thesis: methods of processing images to drive HDR displays. We first discuss the overall challenge and formulate a high-level approach. Working from that method, we modify the algorithm to introduce a number of optimizations. Given the optimized method, we then discuss the implementation of the methods used in practice.

It is worth noting that while HDR displays are more capable than conventional monitors, they are still fundamentally limited. While the HDR displays have a higher peak luminance and larger dynamic range, they cannot represent arbitrarily high luminances. For example, the peak intensity is only a fraction of the intensity of direct sunlight. The space of directly displayable images is much larger, but the fundamental constraints still apply. Considering that, there are two separate challenges faced in presenting HDR images for display:

1. How to map an image containing luminances or colors that exceed the capabilities of the monitor into the color space of display.
2. How to process image data for display, taking image intensities and a color gamut within that of the display and producing the best possible image.

The first portion concerns performing the tasks of tonemapping operators and color appearance transformations to preserve impression. The second task concerns the actions like applying the gamma curve for LDR displays, or the work presented here used with HDR displays. Accomplishing these two tasks will transform an image into a displayable space defined by a device, then handle the intricacies of how that device maps values into that space. These two stages are illustrated in Figure 3.1.

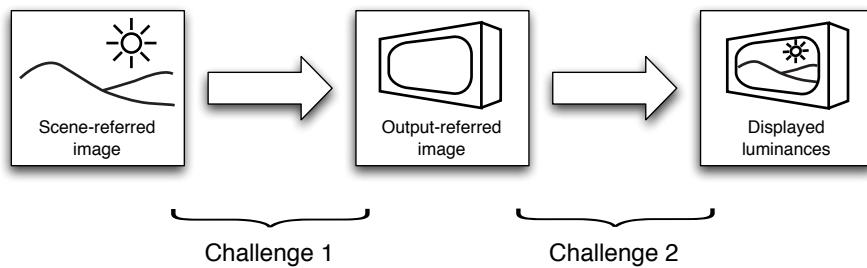


Figure 3.1: Two primary challenges in image presentation. The first challenge involves the accurate capture and transformation of images taken with a calibrated camera to a calibrated display. The second challenge involves the accurate display of those images on a given hardware device.

As discussed in Section 1.1, we will only be addressing the second challenge, that of providing a solid foundation for displaying an image within the color space of the display, that can be built upon to address larger challenges. For clarity, all the work presented in this chapter assumes an idealized display and only addresses algorithmic challenges. In this chapter we assume hardware that responds linearly, and contains no variation from its specification. Chapter 4 deals with calibration methods for nonlinear components, and manufacturing variations.

3.1 Reference Algorithm

Given an image within the displayable color space, we must determine the LED driving values and LCD panel image, that when combined by the optics of a given HDR display, minimize the perceived error between the original and the reconstruction. Not only must the pair of images accomplish that goal, but those images must be displayable by the monitor hardware. The hardware constraints force us to search for two LDR images that can be combined to approximate an HDR image. The same general approach applies to both form factors of HDR displays discussed in Section 2.3, but we focus on LED displays. All implementations were done for the BrightSide DR-37P in particular.

3.1.1 Nonlinear System

The challenge of producing accurate images may be framed in many different ways. We start with a simple approach that makes as few assumptions as possible. We describe a nonlinear optimization problem that compares the displayed image to the desired image using a perceptually-based objective function. The goal is to find the front and back images such that, when simulated, results in an image with the minimum perceived difference with the original. Figure 3.2 contains a diagram of the inputs and outputs of such a system.

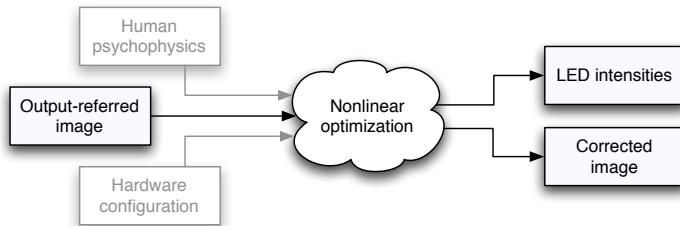


Figure 3.2: Flowchart of nonlinear optimization. Given an image, hardware configuration and perceptual makeup, the nonlinear solver produces the correct LED intensities and LCD image.

This optimization is multi-component problem. In order to have access to the image produced by the display, a simulator of the display hardware and optics is required. In order to perform a valid comparison between the displayed image and the desired image, we need to transform both to a space where meaningful comparisons can be made. A model of portions of the human visual system is also required. The resulting images define the system of equations, objective function, and constraints are required, and this influences the choice of solver.

Simulation of the Display Hardware. The first component addresses the requirements for the conversion of driving values into displayed luminances by simulating the image produced by the display hardware given LED and LCD values. The simulator takes in hardware driving values in the range $[0, 1]$ and maps them to displayed luminances measured in absolute photometric units. First we address how to model the combination

of LEDs and the diffuser that form the backlight, which requires knowledge of the positions of each of the LEDs on the hexagon grid, and the pointspread function (PSF) of the diffuser in front of them. The measured PSF includes both the scattering of the diffuser and the effects of the optics included in the LED package. We model the operation as a 2D convolution of a set of Dirac delta functions at the position of the LEDs, by the diffuser PSF. In the set of Dirac delta functions δ_D , each LED δ_j is modulated by a driving value d_j . With that, the entire simulation for the panel can be formulated.

$$I(p, d) = p \cdot (\text{PSF}_D * \delta_D) \quad (3.1)$$

where I is the simulated image, p represents the values of LCD panel, and PSF_D is the function fit to the measured PSF. The convolution of PSF_D and δ_D simulates the blurring of the physical LEDs, resulting in the backlight for the LCD. This backlight is then multiplied by the pixel transparencies p of the LCD panel to form the final image.

Perceptual Transformation. Once we have the simulated image, the next task is to compare it with the given desired image. We employ a perceptual objective function similar to the visible differences predictor (VDP) described in Section 2.1.3. For the purposes of this formulation we use a simplified model and only include the most important effects for this application: ocular scatter and perceived lightness. To simplify further, we ignore the detection mechanisms, and thus produce more conservative results since the probability of perceiving all differences is 1. We describe the perceptually uniform function ψ as

$$\psi(I) = L \left(\text{PSF}_e(Y_{avg}) * I \right), \quad (3.2)$$

where PSF_e is the pointspread function of the human eye at a given adaptation luminance Y_{avg} , and L is the luminance quantization in JND units.

Objective Function and Constraints. The objective function for our nonlinear optimization is then the difference between the two perceptually uniform image representations. Taking the least squares error to compare the simulated image of a set of LCD and LED values (p, d) and the desired image \bar{I} we have

$$\min_{p, d} \left\| \psi(I(p, d)) - \psi(\bar{I}) \right\|. \quad (3.3)$$

This objective function is then subject to a set of constraints on the physical system: the driving values of the LCD and LEDs must be physically feasible. For our model $p, d \in [0, 1]$. The display must remain within the power limits of the wall outlet. The DR37-P display would pull 4000 W if driven at full power. In addition to being undesirable from a cost standpoint the cooling system cannot sustain this power draw, as a standard breaker is only 1500 W. The majority of the subsystems in the display draw constant power regardless of image content, but the power consumption of the LED array can vary greatly, and we have the inequality constraint

$$e \sum_j d_j \leq e_{tot}, \quad (3.4)$$

where the maximum available power is e_{tot} , and the power cost per LED at full intensity is e .

While this approach generates acceptable images, it does not resolve all of the ambiguity in the system. The equations representing the simulation contains redundant parameters, since both the backlight and LCD panel can be controlled independently at every pixel. Given a panel value p_i and the corresponding pixel of the backlight B_i for the set of light emitting diode values d such that $\bar{I}_i = p_i B_i$, then $\hat{p}_i = 2p_i$ and $\hat{B}_i = B_i/2$ will produce the same value. All combinations of \hat{p}_i, \hat{B}_i are valid provided that $\bar{I}_i = \hat{p}_i \hat{B}_i$. The perceptual metric does not resolve this ambiguity of what the exact solution should be. While a useful guide in the minimization problem, the metric is an insufficient to act as a constraint. It effectively defines a measure where all solutions of the same least squares error are considered the same.

In the space of valid solutions there is room for differentiation. Different applications might sacrifice quality defined by the objective function for other image features such as peak luminance and contrast. The power feasibility constraints do not provide meaningful controls over the space of possible solutions. We need additional definitions to obtain unique solution, and define what would be the best pairing.

When determining the best pairing of images, there are multiple factors to consider. Numerous attributes, such as dynamic range and reconstruction error, are related to the inherent tradeoffs between range and quantization of the system. The set of LCD panel driving values can be divided between increasing the dynamic range and compensating

for the low frequency of the rear panel. The panel can distribute image values across its entire range of driving values and contribute to the dynamic range, or it can be driven at some median value and the values above and below that used to correct for discrepancies between the backlight and the desired image.

Numerical Optimization. The numerical solution to this optimization problem is straightforward to obtain. Since the objective function is a nonlinear least-squares problem, there are many specialized solvers for this specific system of equations from highly-optimized software packages to `lsqnonlin` function in Matlab. `lsqnonlin` employs a trust-region method to obtain the solution and requires the Jacobian of the objective function, and the analytic representation of our system allows us to efficiently evaluate the necessary derivatives. While the system is large, the majority of the unknowns are the LCD pixels p , a sparse system. This sparsity, and the fact we can evaluate the derivatives allows us to solve the system efficiently enough to store in memory.

As seen in Figure 3.3, the numerical optimization method produces accurate results. However, we do not focus our attention on it. More effective means of obtaining solutions exist, which we describe in detail in Section 3.2. While we employ the solver method to validate our formulation of the problem, in practice we do not make use of it for processing images.

3.1.2 Observations

Figure 3.3 shows a sample of the LED and LCD outputs of the algorithm for a given input. The backlight image is a low-frequency version of the original and contains the major features of the original image. The LCD panel contains the remaining image content adjusted for the backlight. The difference between light and dark regions is more uniform, since the backlight represents a portion of the luminance difference. Similar to the work of Chiu et al [12] described in Section 2.2.1, the panel has reverse gradients around light sources to compensate for the light leaking across the edge in the backlight. Recall that the input image is HDR, and must be tone mapped to be printed, while the two resulting images are LDR and can be shown directly.

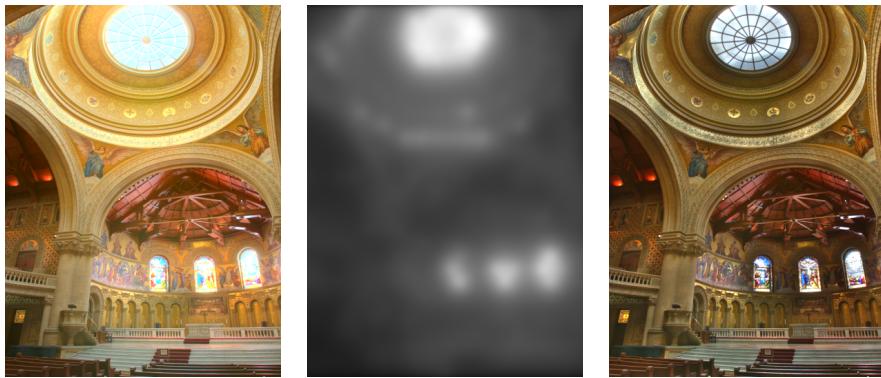


Figure 3.3: The left image represents the original HDR image tonemapped for print. The center image represents the low-frequency luminance image of the backlight, and the right image represents the LCD image compensated for the backlight. (Image courtesy of Greg Ward.)

Obtaining a larger dynamic range implies less ability to correct for the back panel; the limited number of bits in the LCD panel can either be used to extend the dynamic range or to correct the low frequency of the backlight. While the resulting image may be less accurate according to the objective function, the larger dynamic range can subjectively be preferable. This tradeoff is application dependent; casual viewers and professionals have different requirements. Unlike conventional displays, HDR display algorithms require some additional information of what attributes are most important to process it best.

3.2 Performance-Related Modifications

The above discussed optimization method has one major disadvantage, the amount of time it takes to obtain a solution. The system of equations is very large; $m \times m + n$ where m is the number of LCD pixels, and n is the number of LEDs, approximately 2 million by 2 million for DR-37P. The functions of the simulation and perceptually uniform transformation are complex, and the system must iterate. A full solution can take hours per image, restricting the above formulation to precomputed applications.

Precomputation is infeasible in most real-world applications. A monitor has to display images in real-time, and the base requirement of 60 Hz implies that the algorithm

must complete its work in under 12.5 *ms*, using computational resources that could be included in the display, such as a graphics processing unit (GPU) or a field-programmable gate array (FPGA). Additionally, each display model has a different set of intrinsic parameters determined by its construction, and while most of the variation between different versions of the same model can be eliminated by calibration, it might not be possible to completely remove the variation. Images would have to be specifically processes for each model and type of display. This solution would not be forward-compatible or practical in a mixed hardware environment, and would require each application to have knowledge of and to support the intrinsic parameters of each hardware revision.

It is necessary to find new ways to accomplish the same tasks more effectively. Equations outlined above are highly structured, and we can take advantage of this structure to find ways to reduce the complexity of functions involved, reduce the size of the system of equations, and reduce the number of iterations.

The first major optimization is to discard the perceptually uniform transformation. It is too computationally intensive to feasibly model in real-time. Instead of including it in the algorithm, we use it to verify algorithms using test sets of images, to check that our methods produce acceptable results. We test the algorithms and parameters chosen to ensure they come as close to adhering to the bounds as possible. Beyond that, we have identified three major areas of optimization: reformulating the system to better match the hardware construction, breaking that system down into several more tractable sub-problems, and obtaining approximate solutions for those problems.

3.2.1 Simplification of Simulation

The constraints enforced by the HDR display hardware configuration are of particular importance; an examination of the display hardware and its effects upon the optimization function can give some insight into what is being performed. The dual-modulator hardware setup forces all algorithms to share some fundamental characteristics which can be potential targets of optimization. The two main considerations are that the original image content must be distributed between the LCD panel and LED backlight, and that the LCD panel must account for the low frequency of the backlight.

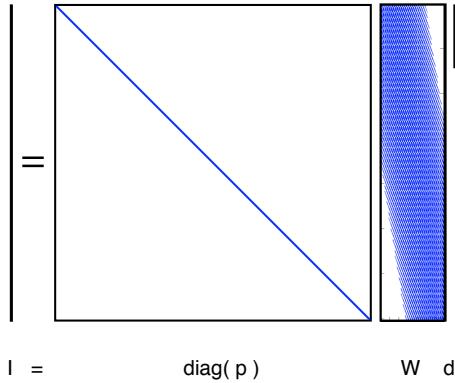


Figure 3.4: Sparsity pattern of simulation matrices.

The formulation of the simulation above is not the most effective means of producing displayed luminances. While described as a functional mapping in Equation 3.1, it can be shown to be a linear system. The structure of this system is apparent in a sparsity pattern of the matrices, as seen in Figure 3.4. From the formulation of convolution of δ_D by the PSF of the diffuser we have

$$I = p \int_{-\infty}^{\infty} \text{PSF}_D(\tau) \cdot \delta_D d\tau \quad (3.5)$$

$$p \int_{-\infty}^{\infty} \text{PSF}_D(\tau) \sum_j d_j \delta(t_j - \tau) d\tau \quad (3.6)$$

$$p \sum_j d_j \int_{-\infty}^{\infty} \text{PSF}_D(\tau) \cdot \delta(t_j - \tau) d\tau \quad (3.7)$$

$$p \sum_j d_j \text{PSF}_D(t_j) \quad (3.8)$$

where t_j represents the difference in positions of the pixel under consideration and the LED j . Since $\text{PSF}_D(t_j)$ is constant for a given LED layout and diffuser, we can precompute the values, and this is equivalent to the linear system

$$I = \text{diag}(p) W d, \quad (3.9)$$

tied together by the $m \times n$ weighting matrix W , where m is the number of pixels and n is the number of LEDs. This matrix W accounts for the layout of the LEDs and PSF

of the diffuser, where each column contains the intensity of LED j at each of the LCD pixels.

3.2.2 Problem Decomposition

The result of observation is that we do not need to solve for ideal LCD pixels simultaneously to solving for ideal LEDs, and instead we can break down the large problem into two sequential steps. Solve for the LED values, and create the matching LCD image. However, this alteration changes the formulation of the problem being solved.

The naive nonlinear optimization does not exploit this fact to the full extent, and solves for both LEDs and LCD pixels simultaneously. Without perceptual transforms, the pixels of the LCD panel are linearly independent since $\text{diag}(p)$ is a diagonal matrix by definition. Since the LCD panel is a modulator of the backlight and because the simulated image I should match the desired image \bar{I} as closely as possible, it is simple to choose p for a given B . For any given backlight image $B = Wd$, we set a given LCD pixel to

$$p = \frac{\bar{I}}{Wd}. \quad (3.10)$$

to be equal to the respective pixel of the original image divided by that of backlight. Figure 3.5 demonstrates this relation visually. The backlight image in the center is a low-frequency approximation of the original image on the left. The approximation is unable to reproduce the high-contrast boundary; producing less light than desired on the brighter side and more light than desired on the darker side. The LCD image in Figure 3.5 right compensates for this blur by letting more light through on the brighter side and less light through on the darker side, so that the end result is a high-contrast boundary between two uniform regions of luminance.

Disregarding the LCD panel means that we are attempting to solve a different system. The separation removes all influence of LCD panel, and its correcting effect is no longer present when solving. Instead of solving for both the LCD pixels and LED values that yield the simulated image with the minimum difference to the original, we are now solving for set of LED values that minimize the difference

$$\min_d \| Wd - \bar{B} \|_2 \quad (3.11)$$

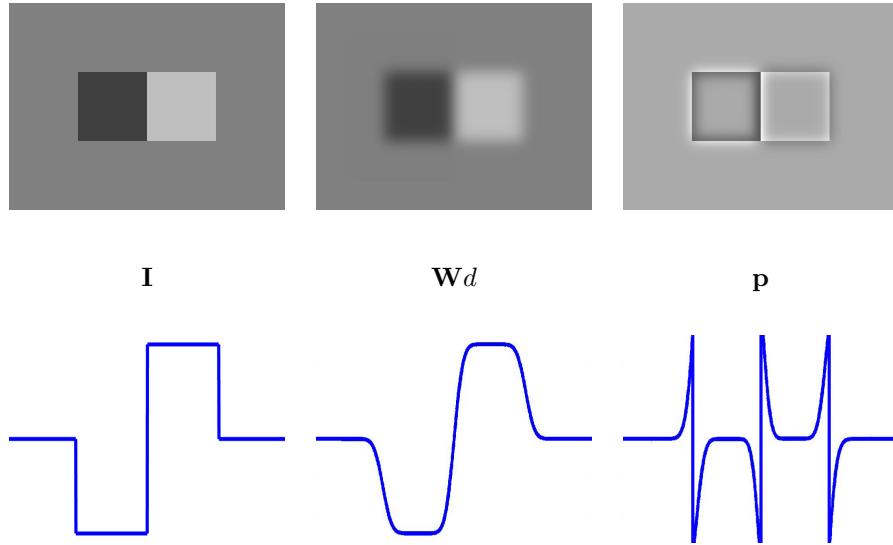


Figure 3.5: Blur correction steps. From the input image (left), the low frequency backlight (center) is computed and simulated, which is used to compensate the original to produce the LCD panel image (right).

from the target backlight image \bar{B} . Because W represents the convolution of the δ_D by the PSF_d in the original Equation 3.1, determining the ideal LED values is essentially a de-convolution problem of \bar{B} to find d .

In order to take advantage of this separation, we need to be able to determine the target backlight \bar{B} from desired image \bar{I} . We begin by considering the idealized projector version, where we initially assume that both the projector and the LCD panel are perfectly linear, and have the same dynamic range. For now, let us also assume perfect alignment and neglect the blurring of the projector image. Recall that the front panel is an optical filter of the rear image. Under these assumptions, the target luminance can be achieved by normalizing the intensity range of the display and the image to be presented to $[0, 1]$, and using the square root of this normalized intensity to drive both the projector and the LCD panel. The even split between pixel values on the projector and the LCD panel is preferable to a scenario in which one value is very large and the other is very small, since quantization artifacts are relatively large for small values. Also, if different combinations of values are used for adjacent pixels of the same intensity, the

imperfect alignment present in real hardware systems would cause significant artifacts. The same basic principle applies with the LED version, but instead of the projected image behind the panel, there is diffused grid of LEDs.

Given that we can determine the target backlight, we only need to solve for the LEDs values to produce that backlight. It is possible to decouple the de-convolution that determines LED values and the simulation that is required to determine the matching LCD image. In the case where there are m pixels and n LEDs, this reduces the original system of size $m \times m + n$ to 2 systems of size $m \times n$, where the de-convolution must solve an $m \times n$ system and the simulation must evaluate an $m \times n$ system. This decoupling yields an immense performance increase, since there are roughly 1 000 times fewer LEDs than LCD pixels.

3.2.3 Approximate Solution

Additionally, the constraint differences between the two stages allows us to optimize each individual stage in ways that would be impossible if we solved them as a combined problem. Thus far the operations we have presented have not altered the resulting value. The remaining changes to be introduced cause the solution obtained to differ from the exact solution. The error introduced is acceptable if it is not detectable by the HVS, and is a necessity for increasing the performance to achieve the required the real-time rates.

In the case of the projector-based display, because the support of PSF is sufficiently small, we do not need to solve for the influence of adjacent values, and instead just simulate the effect of the diffuser. We first choose a simple estimate of what the projector intensity should be as described above, then simulate the effect of blurring, and choose pixel values of the LCD panel that compensate for these effects. In the case of the LEDs, the support of the PSF is much wider, so many LEDs influence the value at a given pixel, and the full de-convolution is necessary.

In both the projector and LED configurations, the low spatial frequency of the backlight implies that the target backlight would also be low-frequency. In the case of the LED-based display, there are comparatively few elements of the backlight to solve for, and the PSF is much lower frequency. We can downsample to a lower resolution and

solve that system without significant artifacts, reducing the m number of equations in the system to the number of LEDs, n .

Additionally, instead of addressing all LEDs, we consider a smaller neighborhood when computing values through de-convolution. Considering LEDs that are more distant has diminishing returns; they are less able to contribute light to the point being considered, but have the same computational cost per LED. Also, the properties of the HVS and dynamic range of the LCD panel limit the distance at which LEDs can still be adjusted to change a specific pixel value. The shape of the ocular PSF defines an area known as veiling glare where changes in luminance cannot be detected because the light scattering obscures the details. Outside of this area, increasing the intensity of the LCD beyond the locally desired value would be detected. This relation can be seen in Figure 3.6.

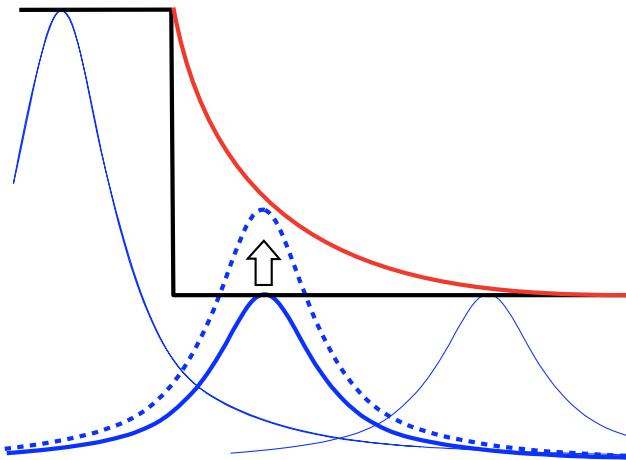


Figure 3.6: This demonstrates the restriction veiling glare places on LEDs to be considered. If more light is needed at LED 1 (left), the intensity of LED 2 (center) can be increased as long as it is less than the veiling glare. However, LED 3 (right) is not covered by the veiling glare, and cannot be adjusted without detection.

While the weighting matrix is dense, the number of LEDs available to be freely altered with respect to a given LED is quite limited and the resulting matrix is a relatively sparse, banded matrix. Unlike de-convolution, veiling glare cannot be used to lessen the computational complexity of simulating the backlight B for a given d , and we have

to consider all LEDs at each pixel to accurately represent the full extent of the PSF. However, the PSF is low frequency and the simulation can be done at a lower resolution and upsampled to the full resolution without significant perceived quality difference. This reduces the evaluation of $m \times n$ system by a large constant factor, and gives a separate tunable quality parameter.

Before decoupling de-convolution and simulation, the solver had to be run until convergence to ensure the LEDs and LCD matched. Not only must the methods produce the desired image, a pairing of LCD pixels and LED values must be matched. Any choice of p and d that do not approximate the correct value will produce highly objectionable artifacts. While this is mostly addressed in the discussion of calibration in Chapter 4, some of the issues are algorithmic. Because the reformulation ensures they are always matched, iterating the solver for the de-convolution acts as an additional means of improving image quality, as opposed to being a necessary step in generating a usable set of p and d . Convergence is no longer required, and we have the option to perform fewer iterations. This substantial performance improvement is a necessity for supporting interactive applications.

3.3 Implementation

Considering the formulation of Section 3.1 and the optimizations of Section 3.2, we now detail how in practice we process images for display. Based upon the previously detailed arguments our approach is decomposed into several stages, with the corresponding flowchart and images in Figure 3.7:

1. Given the desired image \bar{I} , determine target backlight \bar{B} .
2. Determine the LED driving levels d that most closely approximate \bar{B} .
3. Given d , simulate the resulting backlight B .
4. Determine the LCD panel p that corrects for the low resolution of the backlight B , and when combined with B by the display optics, approximates \bar{I} .

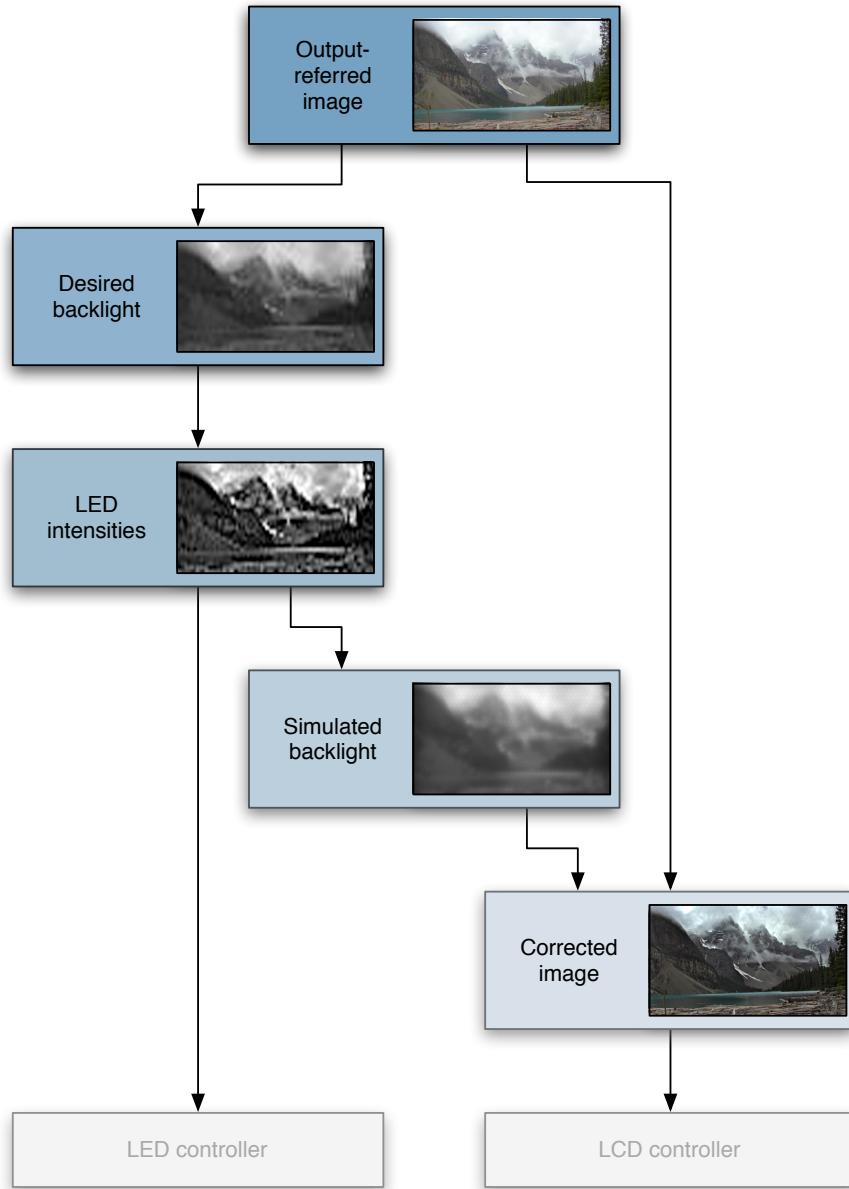


Figure 3.7: Flowchart of stages of the implementation. The scene-referred input image is used to determine the desired backlight, which in turn is used to determine the LED driving values. The backlight is simulated from these driving values, and this simulation is used to compensate the original image.

We address the details of the algorithm on the two hardware platforms currently used in production, a graphics processing unit (GPU) and a field-programmable gate array (FGPA) located in the HDR display. We also describe the methods used in the software testbed to provide high-quality comparisons by which we can judge the chosen optimizations. In Figure 3.7, we show images depicting the output of each stage of the process, and for comparison show a tone mapped version of the original HDR image in Figure 3.8.



Figure 3.8: Tonemapped original HDR image for reference. (Image courtesy of Greg Ward.)

3.3.1 Target Backlight

The first stage takes the desired image \bar{I} and produces the target backlight \bar{B} . The input \bar{I} should be in photometric units, and can be in color but should have the same chromaticity, white point, and primaries as the HDR display. The output \bar{B} will be a monochromatic image in photometric units.

In order for the subsequent steps to proceed correctly, the luminances of the image must be within the range displayable by the monitor. Our task is to clamp the values of \bar{I} to the range $[0, I_{max}]$. The definition of I_{max} is more complicated than a conventional display and is the topic of Chapter 4, but for the purpose of this discussion, we can assume that the value is known.

If it is a color image, we need to convert it to a single-channel luminance¹ representation Y because the backlight is monochromatic. This conversion is to take the maximum of the 3 channels of a given pixel, $Y_i = \max\{R_i, G_i, B_i\}$. One would be inclined to think that Y should be the average or, given the perceptual focus of the discussion, CIE Y . However, because the LEDs are the only way to add light to the system, the \bar{B} must contain luminances therefore at least as high as \bar{I} at every pixel. The only way to ensure this is to take the maximum value, any weighted average will not provide enough light in some cases.

The next step is to divide the dynamic range between the two displays by taking the square root of the clamped luminance image. The actual exponent depends on the ratio of dynamic ranges between the LCD panel and the LEDs², ρ . Because the desired properties are only defined on the range $[0, 1]$, we first normalize the image, raise to it to the appropriate power, and then scale back to photometric units.



Figure 3.9: Output of target backlight pass.

The final step is to take advantage of the low frequency backlight, and to down-sample to the resolution of the LED grid. In software, this can be implemented by any

¹In this case Y is not the same as the CIE tristimulus value Y . We only intend it as a grey-scale representation.

²Since a single LED can be turned off entirely, it could be considered to have an infinite dynamic range. However, for any collection of LEDs, their dynamic ranges are determined by the PSF shape and the values of their neighbors. In the current configuration, this is approximately 1/2.

properly filtered resize function. On the display FPGA, this is implemented as the average of neighborhoods of pixels around LED positions. On the GPU, this is implemented by recursively taking block averages to work within the finite number of texture accesses available. Figure 3.9 shows the output of this stage: a monochrome, low-resolution sampling of square root of the original image.

3.3.2 Deriving LED Intensities

This stage approximates the solution the de-convolution problem and efficiently determines the LED driving values. The process takes in a target backlight \bar{B} in photometric units, and produces driving values $d \in [0, 1]$ that minimizes the difference with the target backlight. As stated, the full solution to this problem would require minimizing a system with as many equations as there are LEDs, subject to the constraints of feasible values and power consumption. We discard the constraints and clamp the output to feasible values $[0, 1]$, and scale the final result to adhere to power limits if necessary. Without constraints, the problem reduces solving the linear system of equations from Equation 3.11

$$Wd = \bar{B}$$

where \bar{B} is the image of the target backlight. While generally less computationally intensive than a minimization problem, the cost of this process is still prohibitive, and thus we desire an approximate solution for real-time applications.

Direct methods of solving systems of linear equations do not decrease the norm with each iterate, and have to be run to convergence to get useful information about the solution. Iterative solvers, on the other hand, make incremental improvements; a few loops will give some meaningful progress and we can stop if the intermediate result is sufficient. Additionally, because the de-convolution is numerically unstable, we do not want to iterate too much; running more iterations may actually decrease the quality of the solution. We chose one of the simplest iterative solvers, the Gauss-Seidel method, on which to base our implementation. The basic Gauss-Seidel iteration

$$d_j^{(k)} = \frac{\bar{B}_j - \sum_{i < j} w_{ji} d_i^{(k)} - \sum_{i > j} w_{ji} d_i^{(k-1)}}{w_{jj}} \quad (3.12)$$

is the result of the reordering of the system of equations and solving for the unknowns d_j . Every step, a new estimate $d^{(k)}$ of the solution is chosen by comparing the current value of the system to the desired value. The new solution estimate is used to update the value of the system.

Modifications to Solver. We make several modifications to this formulation to suit our purposes. Instead of considering all other LEDs for each LED, we use a smaller neighborhood $\mathcal{N}(\delta_j)$, and only perform a single iteration. The resulting computation is a weighted average of the neighborhood of LEDs. Given a desired backlight image \bar{B} , it tries to account for light contributions from other LEDs weighted according to PSF_D . By choosing $d^{(0)} = \bar{B}$, it collapses to

$$d_j = \frac{\bar{B}_j - \sum_i^{\mathcal{N}(\delta_j)} w_{ji} \bar{B}_i}{w_{jj}} \quad (3.13)$$

for a given LED j , where w_{jj} is the value of the pointspread for that LED, or simply the $\max(\text{PSF}_H)$. Then, for a given LED j , the desired luminance value of the backlight at its position is compared to the luminance coming from the surrounding LEDs. The value of LED j is chosen to compensate for any disparity between the desired backlight and the light present. The results are clamped to $[0, 1]$ and passed to the subsequent simulation stage and the LED controller hardware in the display.

While the method draws inspiration from iterative solvers, iteration is not feasible with the current set of optimizations. The pointspread has wider support than the radius of $\mathcal{N}(\delta_j)$, the set of LEDs we choose to alter is based partly on efficiency and partly on veiling glare. The result is that the full extent of the PSF is not represented in the de-convolution. Some amount of the light contributed by the tails of the PSF of each LED is not within the radius that is accounted for. In a single iteration, as in the weighted average case, this is not concern. The only input is the original backlight, and we are not relying on any incomplete intermediate results. With multiple iterations, the intermediate backlight representations $B^{(k)}$ do not match the result of simulating the corresponding $Wd^{(k)}$. The limited range of the update causes the iterates to diverge and the artifacts in the resulting d become progressively more significant. In order to prevent this, we would have to simulate the backlight properly at every iterate; an operation that is too

costly to be practicable.

Figure 3.10 shows the output of this stage. While low resolution, this image appears similar to applying a de-blurring filter to the target backlight of the previous stage. The operation cannot restore the original high-frequency content of the image, but it can adjust the LEDs to account for blurring. Compared to the ideal backlight \bar{B} , the LED values have darker respective blacks and brighter respective whites to account for the scattering that the diffuser will introduce.

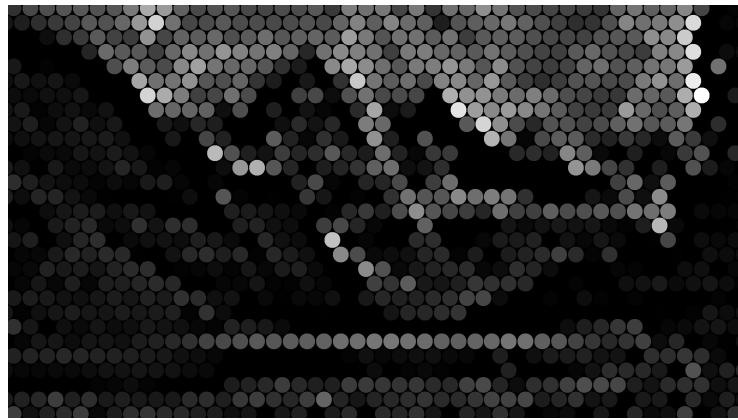


Figure 3.10: Output of pass to determine LED intensities.

LED Grid Representation. This operation is implemented differently in software, on GPUs, and on the FPGA. All three platforms can implement the math in a straightforward manner, and the software and FPGA implementations can make use of data structures designed to represent the hex grid. GPUs, however, can only work in terms of texture operations, and must figure out a way to represent the hex grid using a regular grid.

Current hardware can perform dependent reads of the form $A[B[i]]$ to provide pointer-like indirection, and the positions of each neighbor of the LED could be accessed via a lookup table (LUT). However, GPU architectures are designed to accelerate imaging operations and assume that memory access is coherent. Disregarding this coherency assumption comes at a high performance cost, and significantly impacts the efficiency of any algorithm. It is possible to map the hexagon grid to a subset of a regular grid.

If anisotropically scaled, the LEDs fall on a regular grid, and occupy every other cell in the form of a checkerboard as seen in Figure 3.11. We take this scaling difference into consideration when downsampling, so that the position of each real and fake LED maps to a pixel, resulting the real LEDs being distributed as if on a checkerboard. We then blur the image to distribute information from the pixels that do not represent LEDs to the pixels that do. Neighboring LEDs are obtained using an image filter with modified sample positions that map to adjacent real LEDs. The result of this is read out from the positions of the real LEDs into a vector that stores the final result.

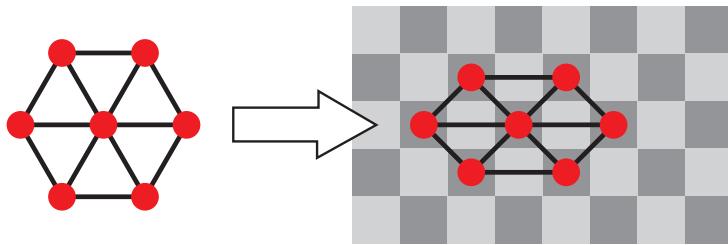


Figure 3.11: Remapping of hex grid to regular grid necessary for GPU computation.

3.3.3 Backlight Simulation

The backlight simulation stage takes the LED values $d \in [0, 1]$ from the previous stage and produces a simulation of the backlight in photometric units. We need to forward-simulate the low-frequency image of the LEDs generated by the diffuser in order to derive the LCD pixel values, and then implement the reconstruction by convolving the LED intensities by the PSF_D . Since the PSF_D is low frequency, all of the following methods can be implemented at a lower resolution and upsampled. If performing at a lower resolution, we need to ensure the pixels of the simulation image are aligned with the LEDs to avoid rounding error.

We use different approaches for software and hardware implementations. In software, we can implement the simulation directly as a convolution. An all-black image with single pixels at the positions of the LEDs set to the the respective driving values is convolved by the PSF_D scaled in photometric units. On the FPGA, we directly evaluate

each pixel by reading the value of the PSF_D for the distance to the current pixel from a lookup table (LUT) and modulate it by the current driving value. On GPUs, we use a splatting approach, and simply draw screen aligned quadrilaterals with textures of the PSF into the framebuffer. Each texture is modulated by its driving value and we use alpha blending to accumulate the results.

As we discuss in Section 4.3, the tail of the PSF can be very long. While we simulate the pointspread function with wide support, we truncate at some distance for efficiency. However, the discontinuity at the point of truncation can cause artifacts in the final image, so we must smoothly transition the PSF to zero at the point we truncated. This difference in PSF intensity along with the missing intensity outside the area used lead to differences in the final pixel luminances in the simulation and the pixel luminances of the desired image. While insignificant when compared to the peak luminance of the display, this disparity can contribute to a perceivable mismatch in dark regions. Because the spatial frequency of the remaining portion of the PSF is very low, we compensate by adding a term u to each pixel of the backlight image to represent the light not accounted for. u is chosen to be a fraction of the set of LED driving values d , where the exact amount is determined by the difference in energy between the actual PSF and the truncated simulation.

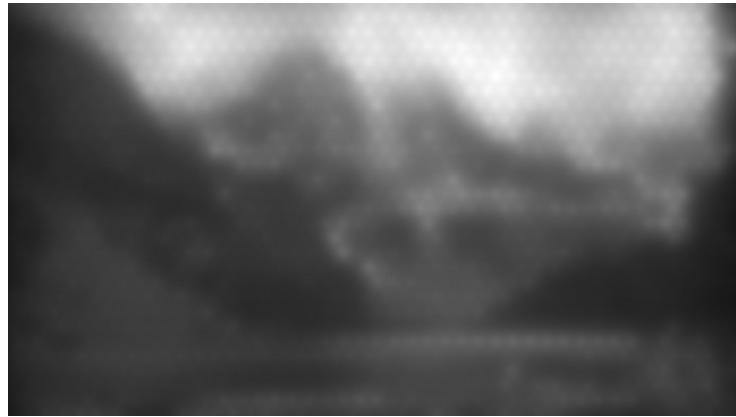


Figure 3.12: Output of backlight simulation pass. In this image, features corresponding to the LED positions are visible because the image is linearly scaled. These features are not visible when the physical display is viewed because of human lightness sensitivity.

Figure 3.12 shows the output of this stage. As much as it is possible, it should resemble the image in Figure 3.9 before downsampling. Though it rarely will match perfectly, it should be close. More importantly than exactly matching the target backlight, the output should match the actual backlight produced by the optics of the hardware. This consistency is essential to obtain the proper output from the next stage.

3.3.4 Blur Correction

Given the simulated backlight \bar{B} , we need to produce the matching LED image p . We correct the original image \bar{I} for the difference due to the blurriness of the backlight. Since the LCD panel modulates backlight, recall Equation 3.10 and the fact that

$$p = \frac{\bar{I}}{\bar{B}},$$

we divide the original image by the backlight simulation to get blur corrected image. This operation is simply an element-wise division, applied to each channel of \bar{I} if it is color, with the addition of an ϵ to compensate for zero values in the simulation. The result is clamped to the range $[0, 1]$ and sent to the LCD controller hardware.



Figure 3.13: Output of blur correction pass.

Figure 3.13 shows the output of this stage. It displays the same characteristics as the LCD panel image in Figure 3.3. Since the result was obtained by dividing by a low-frequency version of the original image, the LCD panel contains the same reverse

gradients as the work of Chiu et al [12]. The LCD image still contains all of the high-frequency and color information as the original image, but it has a more uniform set of low-frequencies to provide a match the backlight which contains them.

3.4 Error Diffusion

The 4 stages outlined in the previous section are capable of producing high quality images, but can be insufficient in certain circumstances. The numerous computations at lower resolutions, approximations, and other shortcuts accumulate with each pass and the LCD panel is left with the task of correcting for all of the inaccuracies.

3.4.1 Rationale

It is often not sufficient that simply p be in the range $[0, 1]$. In the ideal case, with a high-frequency backlight, discussed above both $\bar{B} = \sqrt{\bar{I}}$ and $p = \sqrt{\bar{I}}$. When the backlight is low-frequency, the formulation of p changes and includes a term c_B to account for the difference between \bar{B} and $\sqrt{\bar{I}}$. In the real system with a low-frequency backlight B that is only an approximation of the desired \bar{B} , then p is defined as

$$p \equiv c_B \sqrt{\bar{I}}, \quad (3.14)$$

where $c_B = B/\bar{B}$. If c_B becomes too large or too small then p is either rounded to 0 due to the 8-bit quantization or clamped to 1 since the LCD panel is LDR. Either case means the loss of color and high frequency information that only the LCD can represent. The limited bit depth and dynamic range of the LCD panel means that it cannot completely compensate for the low frequency backlight and still represent all of the image information, but this inherent loss is acceptable as long as its below perceptual limits.

Consider the pathological case of a sinusoidal curve with an amplitude of roughly half the dynamic range of the LCD panel, with some frequency too high for the backlight to represent. All of the dynamic range of the LCD panel is needed to represent the high-frequency signal. If c_B is any value other than 1, there will be information loss as seen

in Figure 3.14. Some of the loss on the dark end can be dismissed with veiling glare, it is still a big concern in bright areas. If the backlight is not sufficiently bright, then the LCD panel goes to the driving value 255 and loses all texture detail. Many professional users, such as in the medical imaging field, require stronger guarantees of the quality of the content on the LCD panel.

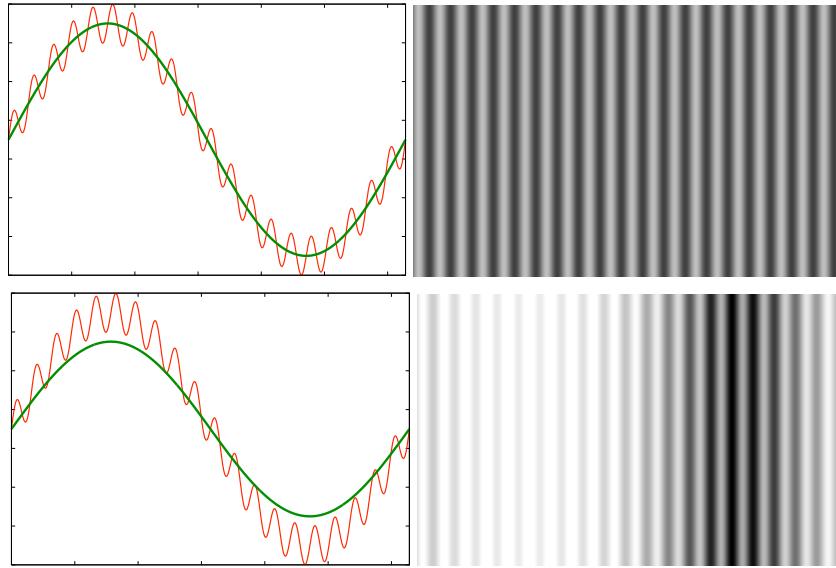


Figure 3.14: Difference between desired and actual p depending on c_B . Note clamping of values over $p = 1$. In the top left image, the low frequency component (green) matches that of the intended image (red), and the high frequency component is accurately reproduced on the LCD panel (top right). In the bottom left image, the low frequency component does not match the intended image, and the resulting LCD panel image is clamped (lower right).

We must perform some additional operations to ensure that the LED values d better approximate the optimal backlight \bar{B} , so that c_B is closer to 1 and the LCD has more of its bit depth available for representing high-frequency detail and color. One solution is to fully simulate the iterates of the de-convolution stage in Section 3.3.2, but this does not directly address or make any strong guarantees about the particular problem presented here.

Suppose that we wish the average LCD panel p_{avg} to equal some value α and we choose the LED values d so this is the case as often as possible. This task is the

minimization of a slightly different linear system to that of Equation 3.11. This operation would be something too computationally intensive for all the reasons outlined above if done from scratch. As we already have a reasonable approximation of \bar{B} , we can just pick changes to d to accomplish the goal. After de-convolving to get the LEDs and simulate the backlight, we then modify both to improve the value of c_B .

3.4.2 Backlight Update Process

If we assume the current d are close to forming \bar{B} , then these changes Δd_j should be small. Instead of attempting to solve for all LEDs simultaneously, we take a greedy approach and iterate over the LEDs 1-by-1. We pick the best value for that LED, update d and the simulation B , and then the subsequent LEDs correct for the new backlight. For each LED, we choose the value that minimizes the difference between c_B and 1 at that point, yielding the relation

$$\frac{\bar{I} - \Delta d_j W_j}{Wd} = \alpha \quad (3.15)$$

which can be formulated as the least squares problem of d_j . Substituting B for Wd the problem becomes

$$\left\| \bar{I} - \Delta d_j W_j - \alpha B^{(j)} \right\| = 0, \quad (3.16)$$

where $B^{(j)}$ is the backlight after the first j LEDs have been updated. Solving this, we get the following formulation in terms of image manipulation over the set of pixels under consideration (x, y)

$$\sum_{x,y} \left(\bar{I}_{(x,y)} - \Delta d_j S_{j,(x,y)} - \alpha B_{(x,y)}^{(j)} \right)^2 M_{j,(x,y)} = 0 \quad (3.17)$$

where S_j is the texture splat of the pointspread image for the LED location associated with W_j . Similar to the constrained radius used in the de-convolution to find the LED driving values in Equation 3.13, we add a masking term M_j to define a portion of the image in a neighborhood around the current LED j , such that $M_j(x, y) = 1$ if $\text{pixel}(x, y)$ is within some radius of LED j , and $M_j(x, y) = 0$ otherwise. Solving that equation for Δd_j , we get

$$\Delta d_j = \frac{\sum_{x,y} S_j \bar{I} M_j - \alpha \sum_{x,y} S_j B^{(j)} M_j}{\sum_{x,y} S_j^2 M_j}. \quad (3.18)$$

This algorithm proceeds in an identical manner to the PSF splatting method for simulating the backlight, and operates as a post-process to the output of the simulation, and iterates over the LEDs in scanline order. For the current pointspread image S_j , the corresponding sections of \bar{I} and B are picked and their respective elements are multiplied and then summed together. The resulting $d_j + \Delta d_j$ is written to the LED driving values, and the backlight is modified accordingly by accumulating $\Delta d_j S_j$ into B . The new d can be passed onto the LED controller as normal, and the resulting B is used as the input for the blur correction in Section 3.3.4. Unlike the problems with iterating in Section 3.3.2, this operation can be repeated multiple times. Because the update fully simulates the change in pointspread contribution, it avoids the problem of the backlight representation diverging from the actual values. However, it is too computationally intensive to be run more than once per frame on currently available hardware.

3.4.3 Corrective Image Filter

This design makes one incorrect assumption: that all of the image data being considered is, in fact, correct. This is not true; half the LEDs in the area under consideration have already been modified to the new value, and half have not. If the Δd for the current LED was determined from the current backlight, then any modification to a subsequent LED would reduce the accuracy of the current estimate. Even though Δd is usually small, the accumulated error can be significant. We need to make the algorithm properly account for the changes that will occur after that iteration.

We can utilize the way in which the algorithm iterates over the LEDs to accomplish this goal. Because it iterates in scanline order, we know that the LEDs above and to the left of the current LED have already been updated and are accurate, while the LEDs below and to the right have not yet been updated and are inaccurate. Even if we have not computed what the Δd_k is for some $k > j$, we have the value of the image I and the backlight B at that point. We can assume that the value of LED k will change by the difference between I and B , and weight will change by the opposite amount, and can design an image filter that performs the corrective measure. Adding that filter as a

term to Equation 3.17, we get

$$\sum_{x,y} \left(\bar{I}_{(x,y)} - \Delta d_j S_{j,(x,y)} - \alpha B_{(x,y)}^{(j)} \right)^2 M_{j,(x,y)} F_{(x,y)} = 0, \quad (3.19)$$

which, when solved for Δd like Equation 3.18 yields

$$\Delta d_j = \frac{\sum_{x,y} S_j \bar{I} M_j F - \alpha \sum_{x,y} S_j B^{(j)} M_j F}{\sum_{x,y} S_j^2 M_j F} \quad (3.20)$$

and operates in the same manner as the original process. Since the problem is already in terms of differences Δd , all that is required is to make a filter relative to the LED position that is positive for accurate LEDs and negative for inaccurate LEDs. Figure 3.15 shows the difference in the results of the algorithm for filter inclusion. The left image is the result of the process without any error diffusion; because it does not consider the full extent of the simulation, the LEDs immediately outside the bright area do not account for the light spilling over from the bright feature. The area is too bright and will be made darker by the error diffusion pass. Without the masking term, the updates do not consider that the subsequent LEDs around it will be darkened as well and it over-compensates, as seen in the center image. With the masking term, the LED correctly accounts for the final light, as seen in the right image.

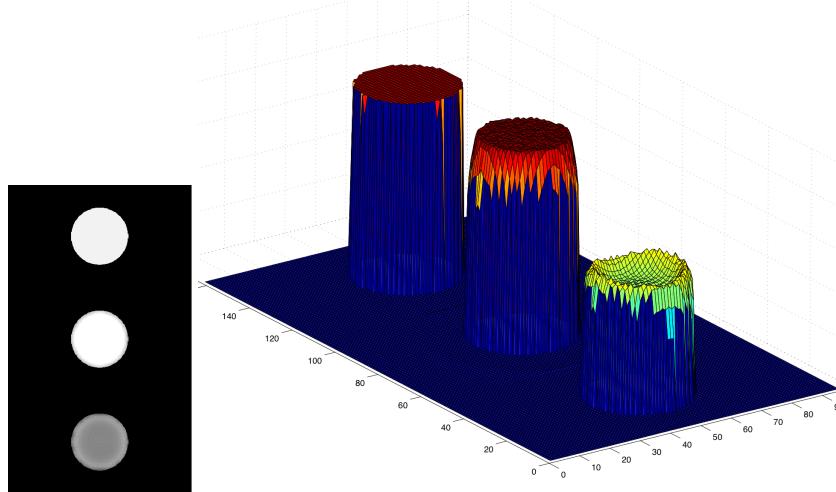


Figure 3.15: Comparison of error diffusion to original method.

Finally, the value of α can be freely set to achieve the desired displayed image. A

value of $\alpha = 1$ will cause the backlight to be the same intensity as \bar{I} , causing the operation to match the target backlight \bar{B} as best it can. Error diffusion acts as a second iteration to achieve the desired image. A value of $\alpha = 0.5$ causes the backlight to be twice as bright as \bar{I} , resulting in $p_{avg} = 0.5$ and there being as many bits available for correction as possible, minimizing quantization artifacts in the LCD panel. More complicated schemes, such as choosing the value of α depending on the local neighborhood, are also possible and can be employed to provide feature-specific tone scaling of the backlight.

The result is a method for adjusting the backlight closer towards the ideal solution of the system without a significant change in computational overhead. It can addresses issues that other parts of the algorithm do not, including details with high-spatial frequency. Additionally, it provides extra parameters to tune the image processing to the needs of the user.

Chapter 4

Measurement and Calibration

The previous chapter describes how to process images for display on an idealized representation of the display. The focus of this chapter is to describe the transformations necessary to make that process work in practice, to correct for artifacts of the hardware setup and provide a device that linearly responds over a given range of intensities.

Images sent to the LED array and the LCD panel are optically combined to form the final image. When the software simulates the backlight and compensates the LCD image for that backlight, it assumes that the images accurately represent the luminances being produced. Inaccuracies in one can interact with inaccuracies in the other, and otherwise small errors can be significantly amplified and therefore become detectable. In fact, a full solution to the nonlinear optimization of the LEDs and LCD pixels using approximate calibration data almost always looks worse than the approximate solution using accurate calibration data.

This level of calibration implies accurate measurement of the display characteristics. Many attributes of the display must be measured to ensure that the simulation results are correct. These include the LED array alignment, the luminances and response of each LED, the LCD panel response, and the pointspread of the backlight diffuser. All attributes related to light intensities are measured in calibrated units, which provide the necessary means of comparing the original image to the simulated result.

4.1 LED Array

Two main features of the LED grid need to be considered. Without accurate information, the simulated backlight will not match the actual backlight. In order to achieve this,

the position and response function of each LED must be known. Position information determines the correct placement of the PSFs used in the simulation, and response information determines the correct luminance for a driving level.

LED Array Alignment. The hardware schematics serve as a starting point for the calibration data. The simulation already knows where the LEDs should be positioned, and needs to be provided with a difference between the actual and desired positions. These differences can result from an individual LED being positioned incorrectly on the circuit board, or the circuit board being positioned incorrectly relative to the LCD panel. In a production scenario, these sources of error are considered manufacturing challenges to meet quality control tolerances. The majority of the remaining misalignment is between the LCD panel and the circuit boards, which are generally accurate within 3 pixels. Humans have poor sensitivity to the low-frequency backlight, and further calibration is not normally necessary. If desired alignment can be further calibrated by examining the difference between the location of several LED PSFs and the corresponding LCD pixel positions.

LED Response. Due to the variance in LED construction and the circuitry that supplies power, the response of the LEDs is neither linear nor is it the same for each LED. Without calibration, they do not respond linearly to driving values and they have different peak intensities. Additionally, the LEDs do not power on at a driving value of 1, and reach peak intensity at driving values less than 255 due to power supply issues. LEDs also significantly vary in response with the ambient temperature of the board.

Some of these issues, such as variance with temperature, must be addressed in the hardware design. These adjustments include, sensors to actively change the driving values to account for the change. Others can be measured after construction, and compensated for with a calibration file. The scope of this calibration task is beyond the breadth of this work, but work on efficiently measuring and calibrating the LED array is underway by Lau et al [38]. We assume that the calibration procedure has been performed and that we can make use of the results. In practice, the hardware controller takes care of the response linearization, and the only parameter we have to consider is

the LED peak intensity, which is provided in absolute units of luminance.

4.2 LCD Panel Response

As discussed in Section 2.4, the LCD panel has a nonlinear response and the input signal must be modified to account for the response. The process of adjusting the LCD panel image to compensate for the backlight in Section 3.3.4 takes place in linear space, and the LCD panel modulation of the actual LED luminances. We need to create a new version of the image from the inverse of the LCD response that will cause the LCD panel to respond linearly.

As discussed in Section 2.4, the LCD panel controller circuitry approximates a power function of roughly 2.5. The production of correct images requires compensating for this nonlinearity, and using that to produce a new image sent to the LCD controller. To obtain the inverse, we follow the same procedure as for LDR display calibration: we measure the luminance of each of the LCD panel's driving values, and represent the inverse as a fitted function or by using a lookup table (LUT). Since the LCD panel acts as a modulator, we do not need to capture any absolute measurement of its response, and use a normalized function. The response of the DR-37P LCD panel is shown in Figure 4.1.

Even though the LCD panel has the same number of bits as an LDR display, the computed inverse must be stored at a higher resolution than that used for an LDR display. The standard practice when implementing an LUT for an OETF like Rec 709 or sRGB is to use a 12-bit lookup table based on the lightness sensitivity discussed in Section 2.4.1 and assumed contrast range of 100 : 1. That contrast range is rarely achieved in actuality since ambient light washes out detail in the dark regions and obscures the error resulting from the 12-bit representation.

In the case of the HDR display, the backlight can be significantly brighter than for a conventional display. The error from the 12-bit representation is still present and can become visible with increases with the backlight luminance. A different method is required to accurately represent the inverse of the response function and we employ a two-part representation to efficiently store the necessary level of detail. We fit an

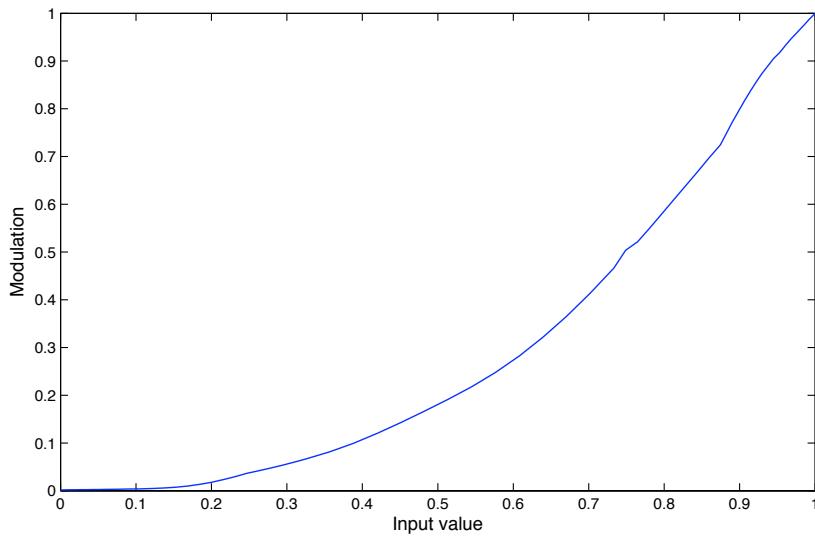


Figure 4.1: LCD panel response.

analytic function such as a *gain-offset-gamma* model to the inverse of the response and evaluate this continuous representation at runtime. To account for fact that the model will not exactly match the data, we store the ratio of the fitted value and the actual value into a LUT.

4.3 Diffuser Pointspread Function

The pointspread function (PSF) of the diffuser calibration is the most critical to accurately rendering images. Unlike the previous attributes mentioned, the pointspread is not a simple artifact that can be corrected for in a post-process; it is tightly coupled with the image processing algorithm. Numerous intrinsic properties of the display are related to it, such as the spatial response and peak intensity discussed in next section. Additionally, its shape determines the weighting matrix W and the values used in the Gauss-Seidel step.

The shape of the pointspread depends on more things than the diffuser. There is a complex optical path from the time a photon leaves an LED to the time it exits the LCD panel. The first segment of the path is the focusing element on each LED,

followed by an empty cavity between the LEDs and the diffuser, then the diffuser itself, which backscatters and causes inter-reflection in the cavity with the circuit boards, into brightness enhancing films to columnate the light, and finally the LCD panel itself.

The actual PSF depends on viewing angle, but because we cannot make any assumptions on the location of viewer, we do not model this in software. Similarly, because the viewer could be anywhere, the design of the optical package must ensure the correct results for all angles. Ensuring uniformity of the PSF between LEDs is more of an engineering and quality control issue, and we assume that we can ignore the angle dependence of the PSF. We also assume that the shape of the PSF is the same for each LED. With the exception of the focusing elements of the LEDs, there is little variation in the materials used, and we find the focusing element variation to be negligible. Beyond that, the practicalities of storing an individual PSF for each LED is too costly to be feasible.

The measurement procedure is straightforward. We turn on a single LED behind the diffuser and place a camera directly in front of that LED, and capture an HDR image of the pointspread of that LED. The HDR image can be generated using an HDR camera such as the Lumetrix IQCam [43], or with an LDR camera using the algorithm of Robertson et al [63]. Because of the variation in peak intensity of LEDs, we normalize the measured data, and later multiply it by the peak value computed from calibrating the individual LED intensities and responses. Figure 4.2 depicts the shape of the PSF.

Several sources of measurement error can affect the quality of the image PSF. Artifacts can appear due to the LCD pixel spacing and camera photosite spacing, and noise present in the HDR image. For these and other reasons, we do not use the measured image data directly, but instead fit a function to it. For now, we assume that the pointspread is radially symmetric, and we fit a function to a cross section of the captured HDR image, a technique that can be extended to more complex shapes if necessary. The PSF is Gaussian-like, but has a wider tail, so we model it as the sum of several Gaussians of varying scales and widths. We recover these values by solving a minimization problem for the relative scales and widths of the component gaussians, using the least-squares error between the fitted function and measured data as the objective function. Finally, we compute the radial PSF from the fitted cross-section.

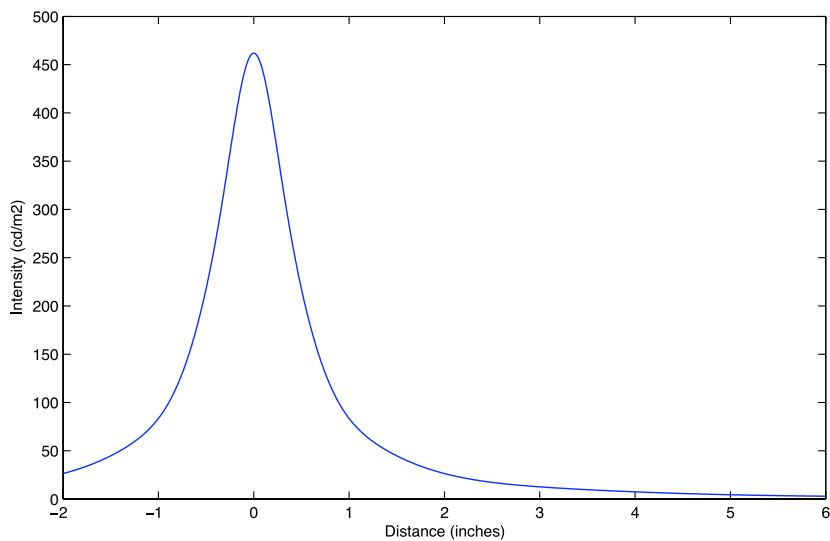


Figure 4.2: Points spread function of diffuser.

Spatially-Variant Response. We have already stated that the backlight at any given pixel is the sum of contributions from many other LEDs, so all of those LEDs must be controlled to reach the desired luminance. However, if adjacent areas of the image are different luminances, veiling glare places a limitation on the number of LEDs we can control. This interdependence implies that the range of luminances achievable for a feature depends on the size of the feature and the intensity of the surrounding area.

For the purposes of illustration, consider a uniform circular feature of one luminance on a uniform background of another luminance. This simplification ignores some of nuances of more complex patterns, but the underlying principles are the same. With a larger radius of the feature, there are more LEDs that have the same desired luminance. If the feature is brighter than the background, then there are more LEDs contributing to the intensity at each pixel in the feature. If the feature is darker than the background, then the brighter background LEDs are further away and contributing less light. If we are only considering the falloff of the PSF in Figure 4.2, this does not seem to be a significant issue. While the contribution from an LED decreases with larger distances, the number of LEDs at any given distance is proportional to that distance, and more contribute the further away you go. Figure 4.3 shows the minimum and maximum backlight

luminances achievable for features of different radii on a 500 cd/m^2 background.

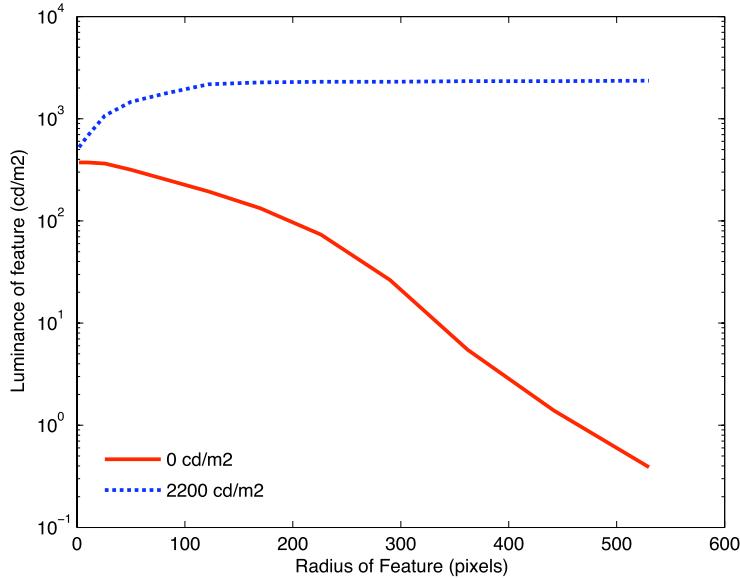


Figure 4.3: Spatial response of the HDR display. Note that while the bright feature increases rapidly towards the desired luminance, the dark feature decreases much more slowly. Because the tail of the diffuser PSF extends for a wide distance around the LED, a large area is required to achieve the desired result.

The compensation in the LCD panel image can alleviate this problem, but as discussed in Section 3.3 it does so at the cost of representing less high-frequency and color detail. At some point, clamping occurs and all detail is lost. For features sufficiently brighter or darker than their surrounding areas, the LCD is limited in its ability to alter the intensity of areas, because the LCD can only modulate the available light within a certain range. These limitations are most prominent for features near the peak intensity of the display, where the LCD value before compensation is already high and there is little room above for correction. Despite the measures in Section 3.4, it can be impossible to make a feature the correct intensity. In that case, the panel clamps to either 0 or 1 and results in an objectionably large area of no texture detail.

Effective Peak Luminance. This insight into the relationship between feature size and luminance begs the question of what is the value of I_{max} , effective peak luminance

of the display. The display has limitations in the contrast between features and their surrounding areas, while the image data clearly does not. For this reason, the effective peak intensity must be less than the luminance obtained from a uniform image formed by turning all LEDs to full. However, it is not prudent to optimize for the other extreme of a single pixel at peak intensity and everything else black. In this scenario, the HDR display is limited by the contrast of the LCD panel and offers no benefits over a conventional display.

As with many other aspects of the display, this value is strongly dependent on the images to be displayed and the requirements of the user. We must balance the distribution of pixel values in the image and the median contrast of boundaries with how accurate the user requires the final image to be. Test patterns often have features of peak luminance adjacent to images of zero luminance. Similarly, scientific data, such as medical images, can contain small features of high contrast, and require accurate representation since the image content is being used to make informed decisions.

HDR photographs and images of natural scenes have a comparatively wider distribution of these characteristics and it is less likely that the brightness of a small feature cannot be accurately represented. We have not experimented widely with scientific images, but have empirically determined a reasonable starting value for photographs. We compute the backlight that results from driving every LED at full intensity, and compute the average of all the pixel luminances to account for falloff towards the edges. We then choose I_{max} to be 75% of this average value.

Chapter 5

Evaluation

This chapter presents the results of our techniques and evaluates the quality of the images produced by the hardware compared with the desired image. We make use of Mantiuk et al's [45] HDR VDP to perform the comparison and show that, while the hardware limitations prevent reproducing the exact luminances of the original, a human observer cannot readily detect the majority of the differences. We first review several details important to interpreting the results, and then proceed to the discussion of results.

5.1 Preliminaries

Demonstration of Ocular Scatter. A fundamental claim of the hardware that the veiling glare resulting from ocular scattering is able to mask the inability of the LCD panel to completely compensate for the low-frequency of the backlight. Before discussing its effects on image perception, we demonstrate that this claim is true. It is not possible to demonstrate the claim without the utilization of an HDR display, but we can make it clear that the claim is feasible. Figure 5.1 contains a tone mapped HDR photograph of a red square being shown on the DR-37P. Around the square there is a large white bloom that is the result of the inability of the LCD panel to completely compensate for the low frequency of the backlight. The additional much smaller red bloom just adjacent to the square is the result of scattering in the optics of the camera. If the optics of the human eye were comparable to a modern SLR camera, then the claim would not hold, however as described in Section 2.1.1, this is not the case and the human eye introduces significantly more scattering than optical quality glass. The larger amount of scattering implies that the veiling glare would be larger and the red bloom should be much wider.

The assumption is true if the bloom appears completely red to an observer, and this is the case with the actual system.



Figure 5.1: Demonstration of the difference between veiling glare from optics and low frequency of backlight diffuser. Note: Original is color image.

HDR VDP Output. As mentioned above, we use the HDR VDP to compare the quality of the image produced to the original. The method takes both the original and displayed images on input, processes them for each stage of its pipeline, and produces a map of the probability of detecting differences at each pixel. The VDP predicts visibility of differences ie. if both the original and the modified are visible. In order to accurately interpret the results presented, it is necessary to be familiar with the output. Recall from Section 2.1.3 that after the veiling glare and luminance quantization steps the HDR VDP processes each with a series of filters sensitive to spatial frequency and orientation. The differences between respective filtered images are then passed to the contrast sensitivity and probability summation stages.

Because for a particular difference between the two images, the filtering operations are performed in the frequency domain, the location of a difference relative to the position of the associated feature is dependent on the spatial frequency of the difference. However, for computational efficiency, the VDP does not filter the images with respect to every spatial frequency or orientation and instead makes use of a smaller set of ori-

entations and frequency bands. This approximation results in banded areas of detection which, upon first inspection, appear unrelated to the feature. In reality these differences are much smoother, and if all bands and orientations were used, these features would be wider and more evenly defined. In this case, we do not interpret them as representative of the exact shape of the visible difference, but rather as an indication of the existence of a perceivable difference and its size and magnitude.

Figure 5.2 depicts the output of the HDR VDP applied to a sample pair of images. In the case of the HDR display, the pairs of red stripes on either side of each face of the square represent that the edge is not accurately reproduced in the displayed image. The red bars outside the square indicate that there is more backlight than can be compensated for while the red bars inside the squares indicate that the backlight is insufficient and that the LCD panel cannot let more light through. Similarly, the angled features inside the corners are representative of the fact that the backlight is lower frequency and cannot represent the sharp corners.

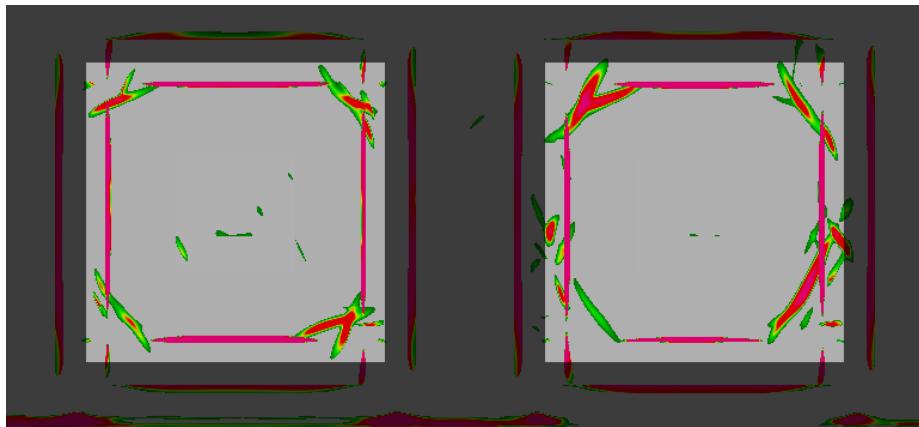


Figure 5.2: Example of HDR VDP output. Note: Original is color image.

These issues cause the probability map to be difficult to interpret directly. While it can be shown independently, the VDP results are considerably more readable if this map is displayed using the image as context, and is normally presented as an overlay of the original image. The overlay we use in Figure 5.2 and all other figures in this chapter displays all probabilities over 95% as solid red, probabilities between 75% and 95% as a

gradient from green to red, and does not display probabilities below 75%. The end result is an image containing probabilities of detection between 0% and 100%. Additionally, we list the percent of pixels in each of the two percentiles.

5.2 Algorithm Evaluation

In the evaluation of our methods, we compare the original image to a simulation of the luminances output by the display device. The measurements taken during the calibration process provide absolute luminance data, and we make use of it to accurately simulate the luminances produced by the display hardware.

For the discussion of our results, we compare the output of the HDR VDP for four images: two test patterns and two photographs. Each set is presented in the same way: the original image is on top, the displayed image is in the middle, and the VDP probability overlay is at the bottom. Since both the original and displayed images are HDR, they are first tone mapped to 8 bits using Reinhard et al's photographic tone mapping operator [61] for display. All images were processed using the same method and were produced using the software implementation of the algorithm running on a 3 GHz Xeon processor running Linux.

Test Pattern Figure 5.3 is a combination several different features. In the center, are vertical and horizontal frequency gratings of different spacings, while the horizontal white bars above and below are linear gradients. There are solid rectangles on the left, and the outlined boxes on the right can be used to check alignment of the display. The black level is set to $1\text{ cd}/\text{m}^2$ and the peak intensity is set to $2200\text{ cd}/\text{m}^2$. 1.42% of the pixels had more than a 75% probability of detection while 0.71% had more than a 95% probability.

This image is a very difficult image to reproduce correctly, especially the right side where there is no acceptable set of values. Given the size of the solid rectangles on the left and the relation of intensity to spatial frequency in Section 4.3, we expect the inside red bars which indicate that there is not enough light. On the right, there are several issues that stem from the fact that none of the outlined boxes are big enough

to get the required light. There is too little white area to become bright enough to have the veiling glare obscure the excess backlight in the surrounding dark areas. The bars outside the vertical box, and the splotches on the dark areas indicate that there is too much backlight. The larger splotches are the result of the backlight being too bright for a large area. They do not appear adjacent to the outline rectangles because there is veiling glare to obscure the differences in those areas. The diagonal hashing in the vertical box indicates an LED arrangement that does not optimally match the box shape, and the LED grid pattern is visible.

Frequency Ramp Figure 5.4 consists of alternating white and black boxes of various widths and heights, similar to some of the DCT basis functions used by JPEG images. Once again, the black level is set to 1 cd/m^2 and the peak intensity is set to 2200 cd/m^2 . 1.15% of the pixels had more than a 75% probability of detection while 0.79% had more than a 95% probability. Considering the edge contrasts and feature sizes, the algorithm performs well, but shows the common problem of failing to maintain peak intensity towards edges of features. The red bars inside the white rectangles indicate where the LCD panel switched to full white causing a perceivable discontinuity. The red bars in the corners of dark areas indicate excessive light being spilled from the two adjacent bright areas. The number of visible differences in the upper right is due to the relation between the feature shape and the LED grid. The packing of the LED grid is aligned horizontally, so while thin horizontal features can be accurately depicted, thin vertical features will cause a saw-tooth like vertical pattern that is detected. This orientation difference is why the error is detected in the upper right, but not in the lower left where the same features have been rotated 90 degrees.

Apartment Figure 5.5 is the first of the two photographs of real scenes and depicts an indoor scene. The values are roughly calibrated to absolute photometric units, and the minimum value is 0 cd/m^2 and the maximum value is 1620 cd/m^2 . 0.26% of the pixels had more than a 75% probability of detection while 0.16% had more than a 95% probability. Compared to the test patterns, it has noticeably less error. Most natural images do not contain such drastic contrast boundaries as the test patterns, and the

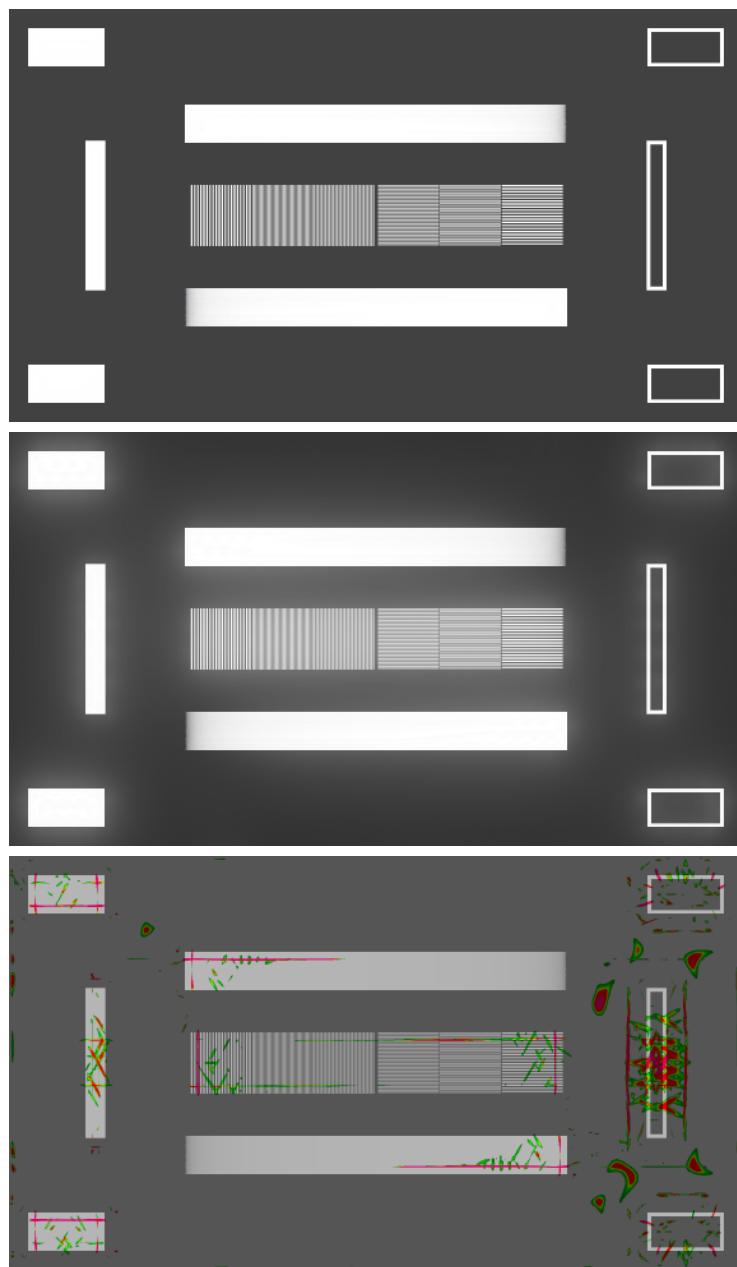


Figure 5.3: TestPattern.

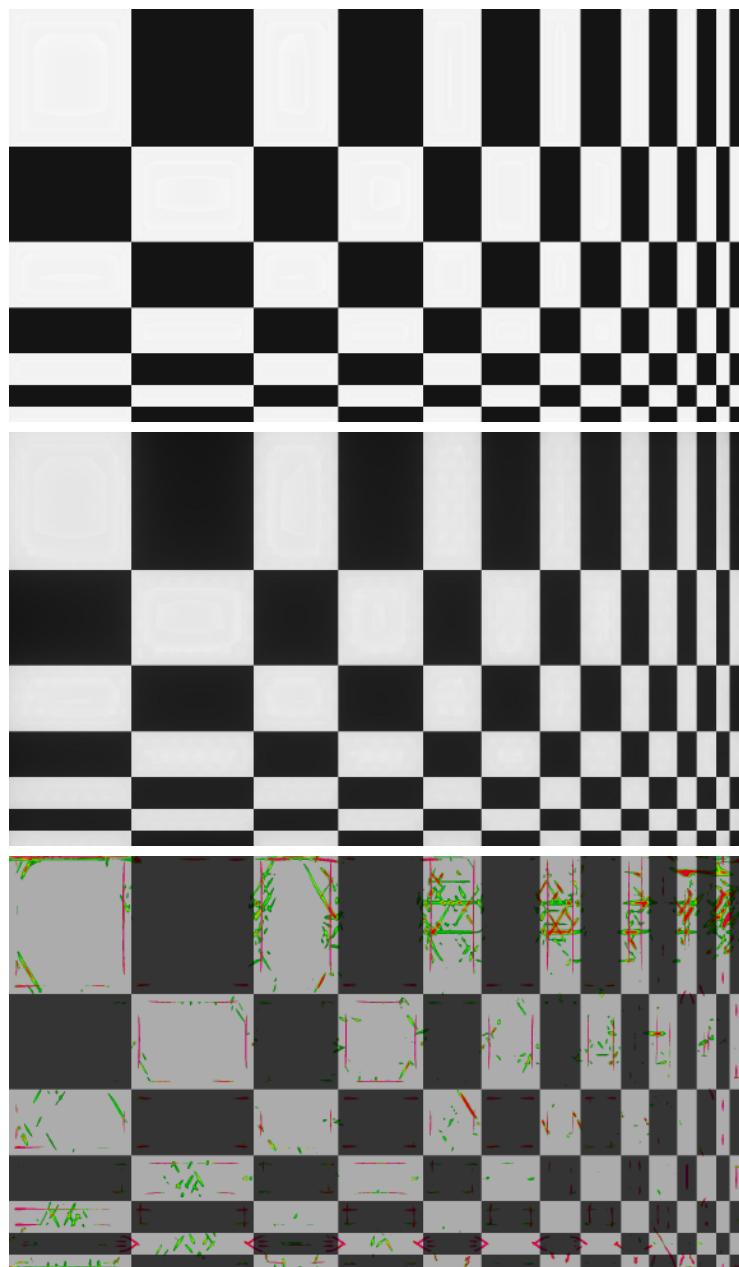


Figure 5.4: FrequencyRamp.

result is considerably fewer areas where the display is not able to accurately represent the image. Most of the error is in the small bright reflections on the balcony, or in the reflection of the lamp in the TV, which probably are not of much consequence.

Moraine Figure 5.6 is a sample of an outdoor scene. Again, the values are roughly calibrated to absolute photometric units. For this image, the minimum value is $0\text{ cd}/\text{m}^2$ and the maximum value is $2200\text{ cd}/\text{m}^2$. This image is an example of an image that is perfectly represented on the display with 0.0% of the pixels had more than a 75% probability of detection. All features are within the tolerance range for the size-to-luminance relationship described in Section 4.3, and no boundaries are so extreme that we cannot accurately reproduce luminance and detail on both sides. While we have shown that this is not the case for all images, this does validate that there is nothing intrinsic in the display hardware that prevents producing artifact-free images.

Distance-Dependent Sensitivity. An important property of the veiling glare is that it is not related to the size or shape of the object causing the glare, only to the luminance of that object. Walking away from, or towards, a bright object will not change the angle subtended by the veiling glare. The size of the glare *appears* to change because the angle subtended by the objects in view does change when moving closer or further away. Moving closer causes the size of the veiling glare to decrease relative to the objects in the scene, while moving further away causes the size of the veiling glare to increase relative to the objects in the scene.

This property has important implications for the design of the optical package of HDR displays. The distance between the viewer and the panel plays into the spacing of the LEDs and the diffuser used. Viewing a display designed for close viewing from much further away means that there are extraneous LEDs consuming power because, as dictated by the relation of the PSF of the diffuser and PSF of the eye, they are packed closer than necessary. Conversely, viewing a display designed for distant viewing from up close means that the LED PSFs will be too wide with respect to the PSF of the eye and the inaccuracies will be visible.

The design of the optical package of the DR-37P takes this into account and the



Figure 5.5: Apartment.

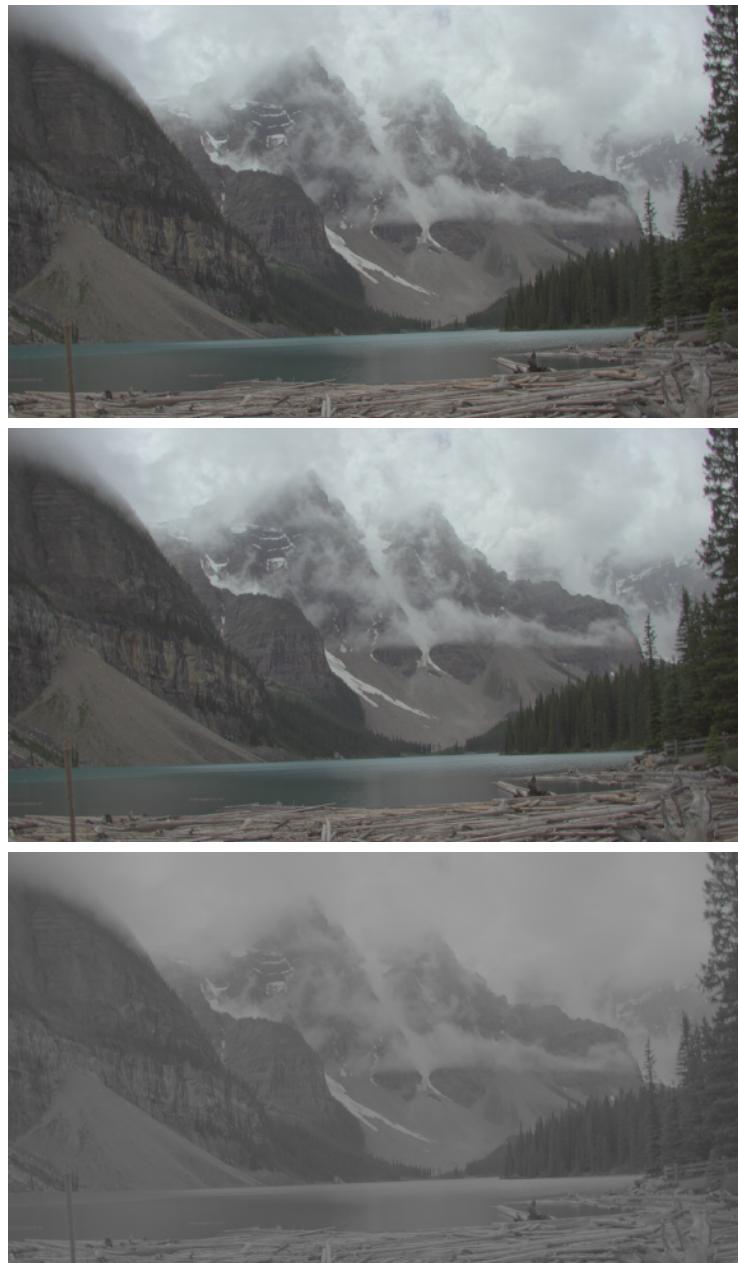


Figure 5.6: Moraine. (Image courtesy of Greg Ward.)

display has an optimal viewing distance of roughly $3.5m$, the distance between the viewer and the television in an average living room setup. The comparison of visible differences at two viewing distances shows the significance of this effect, and to further illustrate this, we run the VDP at two different assumed distances for the same pair of images. We compare the results at a distance of $3.5m$, the average viewing distance at home, to $1.2m$, which is about as close as you can approach and still view the entire display. We repeat this for both **Test Pattern** in the Figure 5.7 and the **Frequency Ramp** in Figure 5.8, and observe that there is significantly more error in the $1.2m$ distance since the veiling glare is unable to obscure as many of the inaccuracies. Table 5.1 lists the percent of pixels with probability of visible differences for both distances for all images.

| | | >75% | >95% |
|-----------------------|------|-------|-------|
| Test Pattern | Near | 4.48% | 3.15% |
| | Far | 1.42% | 0.71% |
| Frequency Ramp | Near | 4.07% | 3.55% |
| | Far | 1.15% | 0.79% |
| Apartment | Near | 0.29% | 0.19% |
| | Far | 0.26% | 0.16% |
| Moraine | Near | 0.00% | 0.00% |
| | Far | 0.00% | 0.00% |

Table 5.1: Table of percent of total pixels at or above a detection level

5.3 Discussion

Finally, it is worth noting that the assigned probabilities are based on our ability to detect differences in a direct comparison. Without the original image to compare against, the user must rely on other less accurate mechanisms of determining whether a feature indicates a difference. In simple cases, such as detecting a frequency grating or determining that the backlight is spilling across a boundary and is visible on a uniform patch of color, there is enough information that the probability of detection does not change significantly in the absence of the comparison image. In more complex cases, this is not the case. For many applications, the user will not be comparing the display to any

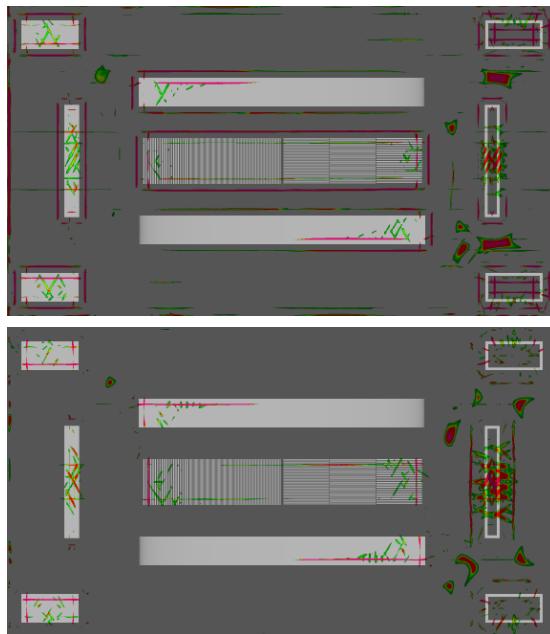


Figure 5.7: TestPattern distance comparison.

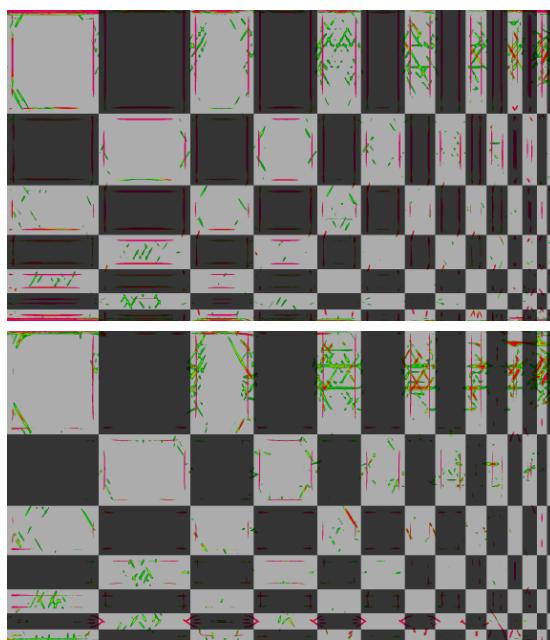


Figure 5.8: FrequencyRamp distance comparison.

ground truth and we can expect that detection probabilities will decrease in many areas. As long as the difference does not look out of place or explicitly wrong, the displayed image will appear as valid as the original. In many cases, even if the user can ascertain that there are differences between the actual and desired images, those differences do not necessarily affect the perceived image quality.

Chapter 6

Conclusions

This thesis addresses the accurate depiction of photometric images on dual-modulator HDR displays. The steady increase in HDR imaging research has created a strong desire to display the additional image data those techniques provide. Display hardware with the potential to fulfill these needs is now available, but due to material limitations the images produced are not pixel-perfect copies of the original image. A study of the psychophysics involved in human perception shows that while the vision system is incredibly powerful, observers have significant limitations in their ability to detect certain features in images.

Much as the motion depicted in video playback at 60 Hz is sufficiently indistinguishable from the original for human observers, the design of HDR display hardware incorporates assumptions about the observer. The hardware is based on the claim that the requirements for accurate depiction of local contrast are significantly less than stringent than is required for the depiction of global contrast, and as long as the hardware is capable of achieving those requirements, the images are indistinguishable. We have shown that it is possible, and that it can be done efficiently and accurately.

The remainder of this chapter is organized into a summary of our contributions, followed by a discussion of possible future directions of research, and some closing remarks.

6.1 Contributions

In Chapter 3, we presented the main contribution: the image processing algorithms for the display hardware. We have researched the properties of the display hardware and the areas of human psychophysics pertinent to that hardware, and incorporated those observations into the design of our image processing methods employed in producing

the input data required by the LCD/LED construction of the hardware device. Because the real-time display of input is a fundamental requirement of any display device, we have devised approximate solutions that operate within the enforced time constraints while still achieving high quality results.

In Chapter 4, we described the differences between our idealized model of the display device and the physical hardware. Compared to the errors introduced by the approximate solutions using an accurate hardware model, we observe that similar magnitude inaccuracies in the hardware models used, even when employed by a full solution, can result in disproportionately large errors in displayed images. The chapter covers the measures required to correctly produce images given the nuances of the physical device.

In Chapter 5, we evaluated our work using the same perceptual principles that inspired the hardware design and our approach. This evaluation shows that to a human observer the displayed images are nearly indistinguishable from the original, even though some of the steps we have performed are approximations and the display hardware cannot perfectly recreate most images.

Beyond its academic contributions, the work we presented also has significant commercial relevance. The research has been actively incorporated into the production pipeline at BrightSide Technologies, the company responsible for the development of the HDR display hardware. The processing algorithms have been in commercial use for 18 months, and the hardware controller inside the monitor runs the algorithm on a FPGA as implemented by The Moving Pixel Company and based on our research. Additionally all real-time image processing for the LED display has been performed using the contents of this thesis, and has been shown internationally at numerous conferences, conventions, and trade shows.

6.2 Future Work

Our work here has only addressed a small portion of the overall problem of accurate depiction on HDR displays, and there are numerous opportunities for extension and improvement. Some of the possible areas include the topics mentioned in the introduction including color, motion, and dependency on spatial frequency. Similar to how

we addressed luminance, we can incorporate the pertinent perceptual background into devising new methods.

As discussed in Chapter 3, we have not addressed the topics of remapping images with pixel values outside the displayable space of the monitor. Hence, there is the opportunity to improve tone mapping techniques from very high dynamic range images to HDR images that the monitor supports, and color space transformations given the extra considerations required over larger contrast ranges. These topics and others are all aimed at more accurate color appearance models, which are needed for the accurate display of images. Fundamentally, all the same constraints found with LDR display systems still apply to HDR displays, but have been loosened. Limits on peak intensity, feasible chromaticities, and other characteristics still exist. Research needs to be conducted in how well current practices work on the problems and how they should, or can, be improved.

Issues remain concerning the improvement of the hardware design and image processing algorithms, as well as addressing the challenges inherent to the combination of LCD and variable backlight. The color saturation of an LCD panel changes when it changes in intensity. Due to the achromatic light leaking through blacks, the darker the color the less saturated that color is. The color space collapses as luminances approach black, and the XZ components of the RGB channels desaturate and move towards the white point. In LDR display calibration, because the poor sensitivity of the HVS to saturation differences for lower luminances, this characteristic is approximated as a constant to be subtracted from all channels. This is not the case with a variable backlight setup of the HDR displays. Since the back illumination is not constant, it cannot be modeled as a scalar value and subtracted out. The challenge is to design a method to remap the gamut based on the backlight brightness and LCD driving level.

6.3 Closing Remarks

The crowning achievement of computer graphics and display science will be to produce images indistinguishable from reality. While there are many steps left on this path, we feel that HDR displays represent a significant advancement. However, more is required than the ability to display contrasts beyond those that are reproducible by existing

displays. Humans perceive the same contrast differently at different background luminances, and thus knowledge not only of the relative values of scene luminances, but also of absolute intensities, is required.

Traditional imaging pipelines have lacked some or all of the additional data necessary to accommodate this requirement. LDR pixel representations do not express the range of values needed to store scene luminances, and device-referred image formats do not include the necessary information to recover the initial values. Even HDR images, if uncalibrated, only contain information about how much of a quantity is contained at each pixel but fail to unambiguously specify what that quantity is. The repercussions of this absence can be observed everywhere from the nuisance of calibrating a home theater setup to the significant overhead costs associated with fields, such as film production and medical imaging where preserving this information is critical. As the stages of the imaging pipeline develop and means of including and processing the additional information to recover photometric data appear, we feel that there will be a movement towards making use of this information.

Approaching the image processing problem in terms of mapping photometric inputs to photometrically calibrated luminances has provided us with many insights into solving the problem we would not otherwise have gained. While not the only requirement in realizing a calibrated, photometric imaging pipeline, the ability to accurately display data is a key element. We feel that accurate depiction provides a piece of the foundation for advanced imaging pipelines based on scene-referred images, and another step towards the eventual realization of the ideal display.

Bibliography

- [1] A. Adams. *The Negative*. Bulfinch, 1995.
- [2] P. Alessi, M. Fairchild, K. Hashimoto, R. Hunt, et al. The CIE 1997 interim colour appearance model – simple version, CIECAM97s. Technical report, International Commission on Illumination, 1998.
- [3] J.V.C. Antwerp. Automatic brightness control apparatus. *United States Patent 4,514,727*, April 1985.
- [4] M. Ashikhmin. A tone mapping algorithm for high contrast images. In *Proceedings of Eurographics Workshop on Rendering 2002*, pages 145–156, 2002.
- [5] B. Barbier, S. Ediar, and J. Brun. System for the display of luminous data with improved readability. *United States Patent 5,057,744*, October 1991.
- [6] C. J. Bartelson and E. J. Breneman. Brightness perception in complex fields. *Journal of the Optical Society of America*, 57(7):953–957, March 1967.
- [7] P. Barten. Physical model for the contrast sensitivity of the human eye. In *Proceedings of the SPIE*, volume 1666, pages 57–72, 1992.
- [8] P. Barten. Spatio-temporal model for the contrast sensitivity of the human eye and its temporal aspects. In *Proceedings of the SPIE*, volume 1913-01, 1993.
- [9] P. Barten. *Contrast Sensitivity of the Human Eye and its Effects on Image Quality*. Proceedings of the SPIE, 1999.
- [10] P. Barten. Formula for the contrast sensitivity of the human eye. In *Image Quality and System Performance*. Edited by Miyake, Yoichi; Rasmussen, D. Rene. Proceed-

- ings of the SPIE, Volume 5294, pp. 231-238 (2003)., pages 231–238, December 2003.
- [11] BrightSide Technologies Inc. DR-37P HDR display, 2005.
<http://www.brightsidetech.com>.
- [12] 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 1993*, pages 245–253, 1993.
- [13] CIE. CIE 19.2.1, An analytic model for describing the influence of lighting parameters upon visual performance, Volume 1. technical foundations. Technical report, International Organisation for Standardization, 1981. Technical Committee 3.1.
- [14] CIE. CIE 15.2004, Colorimetry. Technical report, International Organisation for Standardization, 2004. Technical Committee 1.48.
- [15] S. Daly. The visible differences predictor: an algorithm for the assessment of image fidelity. *Digital images and human vision*, pages 179–206, 1993.
- [16] P. Debevec and J. Malik. Recovering high dynamic range radiance maps from photographs. In *SIGGRAPH 1997: Proceedings of the 24th annual conference on Computer graphics and interactive techniques*, pages 369–378, New York, NY, USA, 1997. ACM Press/Addison-Wesley Publishing Co.
- [17] R. Deeley, N. Drasdo, and W. Charman. A simple parametric model of the human ocular modulation transfer function. *Ophthalmology and Physiological Optics*, 11:91–93, 1991.
- [18] K. Devlin. A review of tone reproduction techniques. Technical Report CSTR-02-005, Department of Computer Science, University of Bristol, November 2002.
- [19] DICOM. *Digital Imaging and Communications in Medicine (DICOM)*, chapter Part 14: Grayscale Standard Display Function.
<http://www.medical.nema.org/dicom/2001.html>, 2001.

- [20] F. Drago, W. Martens, K. Myszkowski, and H. P. Seidel. Perceptual evaluation of tone mapping operators with regard to similarity and preference. Research Report MPI-I-2002-4-002, Max-Planck-Institut für Informatik, Stuhlsatzenhausweg 85, 66123 Saarbrücken, Germany, August 2002.
- [21] F. Durand and J. Dorsey. Fast bilateral filtering for the display of high-dynamic-range images. *ACM Transactions on Graphics*, 21(3):257–266, 2002.
- [22] M. Fairchild and G. Johnson. Meet iCAM: An image color appearance model. In *Proceedings of the IS & T/SID 10th Color Imaging Conference*, Scottsdale, 2002.
- [23] M. Fairchild and G. Johnson. The iCAM framework for image appearance, image differences, and image quality. *Journal of Electronic Imaging*, 2004.
- [24] M. Fairchild, G. Johnson, J. Kuang, and H. Yamaguchi. Image appearance modeling and high-dynamic-range image rendering. In *APGV 2004: Proceedings of the 1st Symposium on Applied perception in graphics and visualization*, pages 171–171, New York, NY, USA, 2004. ACM Press.
- [25] R. Fattal, D. Lischinski, and M. Werman. Gradient domain high dynamic range compression. *ACM Transactions on Graphics*, 21(3):249–256, 2002.
- [26] J. A. Ferwerda, S. N. Pattanaik, P. Shirley, and D. P. Greenberg. A model of visual adaptation for realistic image synthesis. In *SIGGRAPH 1996: Proceedings of the 23rd annual conference on Computer graphics and interactive techniques*, pages 249–258, New York, NY, USA, 1996. ACM Press.
- [27] A. Ghosh, M. Trentacoste, and W. Heidrich. Volume rendering for high dynamic range displays. In *Proceedings of Volume Graphics 2005*, pages 91–98, 2005.
- [28] A. Ghosh, M. Trentacoste, H. Seetzen, and W. Heidrich. Real illumination from virtual environments. In *Proceedings of Eurographics Symposium on Rendering 2005*, pages 243–252, 2005.
- [29] D. Hood and M. Finkelstein. *Handbook of Perception and Human Performance, Volume 1, Sensory Processes and Perception*, chapter Sensitivity to Light. Wiley, 1986.

- [30] IMS Chips. HDRC VGAx, 2004. <http://www.hdrc.com>.
- [31] ITU. ITU-R BT.709, basic parameter values for the HDTV standard for the studio and for international programme exchange. Standard Recommendation 709, International Telecommunication Union, Geneva, 1990.
- [32] G. M. Johnson. Cares and concerns of CIE TC8-08: spatial appearance modeling and HDR rendering. In *Image Quality and System Performance II*, volume 5668, pages 148–156. Proceedings of the SPIE, 2005.
- [33] F. Kainz, R. Bogart, and D. Hess. The OpenEXR image file format. In *SIGGRAPH Technical Sketches*. <http://openexr.com>, 2003.
- [34] S. Kang, M. Uyttendaele, S. Winder, and R. Szeliski. High dynamic range video. *ACM Transactions on Graphics (special issue SIGGRAPH 2003)*, 22(3):319–325, 2003.
- [35] G. Krawczyk, M. Goesele, and H. P. Seidel. Photometric calibration of high dynamic range cameras. Research Report MPI-I-2005-4-005, Max-Planck-Institut für Informatik, Stuhlsatzenhausweg 85, 66123 Saarbrücken, Germany, April 2005.
- [36] J. Kuang, H. Yamaguchi, G. Johnson, and M. Fairchild. Testing HDR image rendering algorithms. In *Color Imaging Conference*, pages 315–320, 2004.
- [37] G. Ward Larson. LogLuv encoding for full-gamut, high dynamic range images. *Journal of Graphics Tools*, 3(1):15–31, 1998.
- [38] S. Lau, W. Walters, and T.s Wan. LED array calibration for linearity and uniformity. Technical report, Structured Surface Physics Laboratory, University of British Columbia, 2005.
- [39] P. Ledda, A. Chalmers, and H. Seetzen. HDR displays: a validation against reality. In *IEEE International Conference on Systems, Man and Cybernetics 2004*, October 2004.

- [40] P. Ledda, A. Chalmers, T. Troscianko, and H. Seetzen. Evaluation of tone mapping operators using a high dynamic range display. *ACM Transactions on Graphics (special issue SIGGRAPH 2005)*, 24(3):640–648, 2005.
- [41] C. Lloyd and R. Beaton. Design of a spatio-chromatic human vision model for evaluation full-color display systems. In *Proceedings of the SPIE*, volume 1249, pages 23–37. SPIE, 1990.
- [42] J. Lubin. A visual discrimination model for imaging system design and evaluation. *Visual Models for Target Detection*, pages 245–283, 1995.
- [43] Lumetrix. Iqcam imaging photometer.
- [44] S. Mann and R. Picard. Being 'undigital' with digital cameras: Extending dynamic range by combining differently exposed pictures. Technical Report 323, M.I.T. Media Lab Perceptual Computing Section, 1994. Also appears, IS&T's 48th annual conference, Cambridge, MA, May 1995.
- [45] R. Mantiuk, S. Daly, K. Myszkowski, and H.P. Seidel. Predicting visible differences in high dynamic range images - model and its calibration. In Bernice E. Rogowitz, Thrasivoulos N. Pappas, and Scott J. Daly, editors, *Human Vision and Electronic Imaging X, IS&T/SPIE's 17th Annual Symposium on Electronic Imaging (2005)*, volume 5666, pages 204–214, 2005.
- [46] R. Mantiuk, G. Krawczyk, K. Myszkowski, and H.P. Seidel. Perception-motivated high dynamic range video encoding. *ACM Transactions on Graphics (special issue SIGGRAPH 2004)*, 23(3):733–741, 2004.
- [47] R. Mantiuk, K. Myszkowski, and H.P. Seidel. Visible difference predictor for high dynamic range images. In *Proceedings of IEEE International Conference on Systems, Man and Cybernetics*, pages 2763–2769, 2004.
- [48] D. Marimont and B. Wandell. Matching colour images: the effects of axial chromatic aberration. *Journal of the Optical Society of America*, 11(12):2113–3122, 1994.

- [49] N. Miller, P. Ngai, and D. Miller. The application of computer graphics in lighting design. *Journal of the Illuminating Engineering Society*, 14(1):6–26, 1984.
- [50] T. Mitsunaga and S. Nayar. Radiometric self calibration. In *Proceedings of IEEE CVPR*, pages 472–479, 1999.
- [51] P. Moon and D. Spencer. Visual data applied to lighting design. *Journal of the Optical Society of America*, 34(605), 1944.
- [52] P. Moon and D. Spencer. The visual effect of non-uniform surrounds. *Journal of the Optical Society of America*, 35(3):233–248, 1945.
- [53] N. Moroney, M. Fairchild, R. Hunt, C. Li, M. Ronnier Luo, and T. Newman. The CIECAM02 color appearance model. In *Color Imaging Conference*, pages 23–27, 2002.
- [54] E. Muka and G. Reiker. Reconsidering bit depth for radiological images – is eight enough? In *Proceedings of the SPIE*, volume 4686, pages 177–188, 2002.
- [55] F.L. Van Nes and M.A Bouman. Spatial modulation transfer in the human eye. *Journal of the Optical Society of America*, 57:401–406, 1967.
- [56] R. Normann, B. Baxter, H. Ravindra, and P. Anderton. Photoreceptor contributions to constast sensitivity: Applications in radiological diagnosis. In *IEEE Transactions on Systems, Man, and Cybernetics*, volume SMC-13, pages 946–953, 1983.
- [57] A. Oppenheim, R. Schafer, and T. Stockham. Nonlinear filtering of multiplied and convolved signals. *Proceedings of the IEEE*, 56(8):1264–1291, 1968.
- [58] S. H. Park and G. M. Montag. Evaluating tone mapping algorithms for rendering non-pictorial (scientific) high-dynamic-range images. *Submitted to ACM Transactions on Applied Perception*, 2005.
- [59] S. Pattanaik, J. Ferwerda, M. Fairchild, and D. Greenberg. A multiscale model of adaptation and spatial vision for realistic image display. In *Proceedings of ACM SIGGRAPH 1998*, pages 287–298, 1998.

- [60] C. Poynton. *Digital Video and HDTV: Algorithms and Interfaces*. Morgan Kaufmann, 2003.
- [61] E. Reinhard, M. Stark, P. Shirley, and J. Ferwerda. Photographic tone reproduction for digital images. *ACM Transactions on Graphics (special issue SIGGRAPH 2002)*, 21(3):267–276, 2002.
- [62] E. Reinhard, G. Ward, P. Debevec, and S. Pattanaik. *High Dynamic Range Imaging*. Morgan Kaufmann, 2005.
- [63] M. Robertson, S. Borman, and R. Stevenson. Dynamic range improvements through multiple exposures. In *Proceedings of International Conference on Image Processing (ICIP) 1999*, pages 159–163, 1999.
- [64] H. Seetzen, W. Heidrich, W. Stuerzlinger, G. Ward, L. Whitehead, M. Trentacoste, A. Ghosh, and A. Vorozcova. High dynamic range display systems. *ACM Transactions on Graphics (special issue SIGGRAPH 2004)*, 23(3):760–768, 2004.
- [65] H. Seetzen, L. Whitehead, and G. Ward. A high dynamic range display using low and high resolution modulators. In *Society for Information Display Internatiational Symposium Digest of Technical Papers*, pages 1450–1453, 2003.
- [66] SMPTE. SMPTE RP 71, Setting chromaticity and luminance of white for color television monitors using shadow mask picture tubes. Technical report, Society of Motion Picture and Television Engineers, 1977.
- [67] G. Spencer, P. Shirley, K. Zimmerman, and D. P. Greenberg. Physically-based glare effects for digital images. In *Proceedings of ACM SIGGRAPH 1995*, pages 325–334, 1995.
- [68] SpheronVR AG. SpheroCamHDR, 2004. <http://www.spheron.com/>.
- [69] J. Stevens and S. Stevens. Brightness function: Effects of adaptation. *Journal of the Optical Society of America*, 53(3), 1963.
- [70] M. Stokes, M. Anderson, and R. Motta. Standard default color space for the Internet. Technical report, World Wide Web Consortium, 1996.

- [71] Sunnybrook Technologies Inc. Zeetzen 5 HDR display, 2004.
<http://brightsidetech.com>.
- [72] Thompson Grass Valley. Viper FilmStream, 2004.
<http://www.thomsongrassvalley.com>.
- [73] J. Tumblin and H. Rushmeier. Tone reproduction for realistic images. *IEEE Computer Graphics and Applications*, 13(6):42–48, 1993.
- [74] V. Virsu, P. Lehtio, and J. Rovamo. Contrast sensitivity in normal and pathological vision. In L. Maffei, editor, *Doc. in Ophthalmology Proceeding Series*, volume 30, pages 363–272, 1981.
- [75] J. Vos. Disability glare - a state of the art report. *CIE Journal*, 3(2):39–53, 1984.
- [76] G. Ward. *Graphics Gems IV*, chapter A Contrast-Based Scale Factor for Luminance Display, pages 415–421. Academic Press, 1994.
- [77] G. Ward. A wide field, high dynamic range, stereographic viewer. In *In Proceedings of PICS 2002*, April 2002.
- [78] G. Ward and M. Simmons. Subband encoding of high dynamic range imagery. In *APGV 2004: Proceedings of the 1st Symposium on Applied perception in graphics and visualization*, pages 83–90, New York, NY, USA, 2004. ACM Press.
- [79] A. Watson. The cortex transform: Rapid computation of simulated neural images. *Computer Vision, Graphics and Image Processing*, 39:311–327, 1987.
- [80] A. Yoshida, V. Blanz, K. Myszkowski, and H.P. Seidel. Perceptual evaluation of tone mapping operators with real-world scenes. In B. Rogowitz, T. Pappas, and S. Daly, editors, *Human Vision and Electronic Imaging X, IS & T/SPIE's 17th Annual Symposium on Electronic Imaging (2005)*, volume 5666 of *SPIE Proceedings Series*, pages 192–203, San Jose, USA, January 2005. SPIE.