

HIGH DYNAMIC RANGE IMAGE RECONSTRUCTION



Synthesis Lectures on Computer Graphics and Animation

Editor

Brian A. Barsky, *University of California, Berkeley*

High Dynamic Range Imaging Reconstruction

Asla Sa, Paulo Carvalho, and Luiz Velho

2007

High Fidelity Haptic Rendering

Miguel A. Otaduy, Ming C. Lin

2006

A Blossoming Development of Splines

Stephen Mann

2006

Copyright © 2007 by Morgan & Claypool

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means—electronic, mechanical, photocopy, recording, or any other except for brief quotations in printed reviews, without the prior permission of the publisher.

High Dynamic Range Image Reconstruction

Asla M. Sá, Paulo Cezar Carvalho, Luiz Velho

www.morganclaypool.com

ISBN: 1598295624 paperback

ISBN: 9781598295627 paperback

ISBN: 1598295632 ebook

ISBN: 9781598295634 ebook

DOI: 10.2200/S00103ED1V01Y200711CGR003

A Publication in the Morgan & Claypool Publishers series

SYNTHESIS LECTURES ON COMPUTER GRAPHICS AND ANIMATION #3

Library of Congress Cataloging-in-Publication Data

Series ISSN: 1933-8996 print

Series ISSN: 1933-9003 electronic

First Edition

10 9 8 7 6 5 4 3 2 1

HIGH DYNAMIC RANGE IMAGE RECONSTRUCTION

Asla M. Sá, Paulo Cezar Carvalho, Luiz Velho

IMPA-Instituto de Matematica Pura e Aplicada, Brazil

SYNTHESIS LECTURES ON COMPUTER GRAPHICS AND ANIMATION #3



MORGAN & CLAYPOOL PUBLISHERS

ABSTRACT

High dynamic range imaging (HDRI) is an emerging field that has the potential to cause a great scientific and technological impact in the near future. Although new, this field is large and complex, with non-trivial relations to many different areas, such as image synthesis, computer vision, video and image processing, digital photography, special effects among others. For the above reasons, HDRI has been extensively researched over the past years and, consequently, the related scientific literature is vast. As an indication that the field is reaching maturity, tutorials and books on HDRI appeared. Moreover, this new resource has already reached interested practitioners in various application areas. In this book, we do not aim at covering the whole field of high dynamic range imaging and its applications, since it is a broad subject that is still evolving. Instead, our intent is to cover the basic principles behind HDRI and focus on one of the currently most important problems, both theoretically and practically. That is, the reconstruction of high dynamic range images from regular low dynamic range pictures.

KEYWORDS

HDRI, High Dynamic Range Reconstruction, Photometric Calibration

Contents

List of Figures	ix
1. Introduction	1
2. Digital Image	3
2.1 Digital Image	3
2.2 Modeling Light and Color	4
2.3 Measuring Light	5
3. Imaging Devices and Calibration	7
3.1 Digital Photography	7
3.2 Tonal Range and Tonal Resolution	8
3.3 Color Reproduction	12
3.4 Camera Characteristics	12
3.5 Camera Calibration	15
4. HDR Reconstruction	19
4.1 Problem Statement	19
4.2 Analysis and Classification	21
4.3 Main Methods	24
4.4 Experiments and Comparison	33
4.5 Improving Robertson's Method	35
5. HDRI Acquisition and Visualization	37
5.1 Data Acquisition	37
5.2 Representation and HDR Formats	38
5.3 Tone Mapping	40
6. Tone Enhancement	43
6.1 Partial Reconstruction: Relative Tones	43
6.2 Active Range-Enhancement	45
6.3 Real-Time Tone-Enhanced Video	46
References	49
Author Biography	53

List of Figures

- | | | |
|-----|--|----|
| 2.1 | The original spectral distribution of an emitted light is shown in (a), the spectral response function of the three filters used to sample the signal are shown in (b), the sampled values are in (c) and the reconstructed signal is shown in (d). | 4 |
| 3.1 | These images illustrate the resultant-acquired images when exposure time is changed. In (a) the sensor has been exposed longer than in (b). | 8 |
| 3.2 | FujiChrome Provia 400F Professional [RHP III] data sheet, from FujiFilm. | 9 |
| 3.3 | Film characteristic response curves compared. | 10 |
| 3.4 | (a) CCD sensor and (b) CMOS sensor, from [17]. | 10 |
| 3.5 | Bayer pattern (from http://www.dpreview.com/learn/ by Vincent Bockeart). | 12 |
| 3.6 | Images taken from http://www.normankoren.com/Tutorials/MTF.html illustrate the effects of MTF on the input target. The blue curve below the target is the film MTF, expressed in percentage; the red curve shows the density of the bar pattern. | 13 |
| 3.7 | (a) Camera spatial resolution power test target. (b) A detail of the same image. The camera used in this test was a Canon EOS 350D and the image is from the site http://www.dpreview.com . | 14 |
| 3.8 | Some cameras' noise compared. Indicated ISO sensitivity is on the horizontal axis of this graph, standard deviation of luminosity (normalized image) on the vertical axis. Image from the site http://www.dpreview.com . | 15 |
| 4.1 | Image acquisition pipeline. | 21 |
| 4.2 | Exposure-intensity graph from samples. | 22 |
| 4.3 | Plot of the reconstructed response function from samples at three pixel locations and five different exposures. | 25 |
| 4.4 | Input images to HDR reconstruction experiments. | 34 |
| 4.5 | Response curves reconstructed with the Debevec and Malik algorithm with different values of the smoothness parameter. | 34 |
| 4.6 | Response curves reconstructed with the Robertson, Borman, and Stevenson algorithm with different values of the weighting function. | 35 |
| 4.7 | Response curves reconstructed the Mitsunaga and Nayar algorithm with polynomial reconstruction of different degrees. | 35 |
| 4.8 | Output f 10-bits of pixel depth. | 36 |

x LIST OF FIGURES

- | | | |
|-----|---|----|
| 5.1 | In (a) differently exposed LDR pictures from the same scene are processed to compose a HDR Radiance Map of the input images. A TMO is then applied to visualize the HDR image; figure (b) is visualized using linear TMO and figure (c) using histogram adjustment TMO. | 38 |
| 5.2 | Error profile of log and floating point encodings. | 39 |
| 6.1 | Example of two different exposed images with correspondent histograms, and the summation image with the combined histogram. To be visualized, the image sum was linearly tone mapped. | 44 |
| 6.2 | Absolute versus relative tone values. | 44 |
| 6.3 | Two consecutive input fields result in one frame. | 46 |
| 6.4 | Images (a) and (b) are the video input fields, while in (c) is shown the tonal-enhanced foreground. | 46 |
| 6.5 | Images (a) and (b) are the video input fields, while in (c) is shown the tonal-enhanced foreground. | 47 |
| 6.6 | Images (a) and (b) are the video input fields, while in (c) is shown the tonal-enhanced foreground. | 47 |

CHAPTER 1

Introduction

High dynamic range imaging is the field that studies digital images whose associated color information has full precision.

To make the above informal definition a little more rigorous, let's look at an image as a mathematical object. In that sense, a color image can be described as a function $f : U \subset \mathbb{R}^2 \rightarrow \mathbb{R}^3$, which associates values $c \in \mathbb{R}^3$ from some color space¹ to pixel locations $(u, v) \in U$ in the image plane. This concept is expressed mathematically as $c = f(u, v)$. Here, the image plane is the *domain*, while the color space is the *range* of function f .

Because we are concerned with digital images, this abstract mathematical description needs a concrete representation that can be processed by a computer and imaging devices, such as a digital camera or a monitor. To this end, the nature of digital technologies requires a discrete (binary) representation. So, our continuous mathematical model of the image function has to be *discretized*. This is done through a discretization of both the domain and range of the function f .

One of the characteristics associated with a discrete representation is its *resolution*. In the case of digital images, since the domain and range are discrete, there is the notion of *spatial resolution* and *tonal resolution*. The spatial resolution is related to the number of pixel locations (u, v) in the image. The tonal resolution is related to the number of bits in the representation and the set of possible values for each color component (c_1, c_2, c_3) of c . For example, if an 8-bit number is used to represent a color component, then it can assume $2^8 = 256$ different values.

Until recently, a significant emphasis has been given to the spatial resolution of digital images. This trend can be witnessed in the evolution of digital cameras that are featuring sensors with higher and higher resolutions. Consequently, the quality of a camera is often measured in megapixels.

Nonetheless, now the focus is shifting from spatial to tonal resolution and this is the point where high dynamic range imaging enters into the scene. In fact, arguably it can be said that tonal resolution is more important than spatial resolution for image processing tasks.

¹For example, $c = (r, g, b)$ values from the CIE-RGB color space, where the coordinates represent the red, green, and blue color components.

2 HIGH DYNAMIC RANGE IMAGE RECONSTRUCTION

Historically, the research in HDRI began motivated by photorealistic image synthesis using physics-based illumination. In this type of simulation, the setup is expressed in physical quantities and the resulting rendered images represent radiance maps. At that stage, the main issue was how to display such HDR images on LDR display devices. Therefore, the problem of tone mapping was given a great deal of attention. Subsequently, the technological evolution of digital cameras motivated the desire for acquisition of HDR images from real scenes. Finally, these three research topics (i.e., physics-based illumination, tone mapping, and HDR acquisition) combined have given rise to various applications of HDRI, such as image-based lighting and others.

As seen above, a part of the research on *high dynamic range imaging* (HDRI) is devoted to overcome sensors' tonal range limitations. The goal is to achieve better representation and visualization of images, as well as to recover the scenes actual radiance values. We will refer to the scenes' radiance values as *absolute tones*, since they are often related to a physical real quantity. The usual images acquired by cameras with limited dynamic range are the *low dynamic range images* (LDRI).

An important observation when studying HDRI is that the input light $C(\lambda)$ is read in an interval around a reference value and the output C is discrete. What is being questioned is that C discretization should not be imposed by sensors nor displays limitations, but it should be adapted to the scene's luminance range. Instead of being driven by the device representation, the discretization should adapt to scenes' tonal information.

This book is structured as follows. In Chapter 2, we review the basic concepts related to the digital images. In Chapter 3, we discuss imaging devices and their characteristics. We also study the problem of camera calibration. In Chapter 4, we describe the main methods for reconstructing high dynamic range images. In Chapter 5, we give an overview of various issues related to acquisition, representation, and visualization of HDR images. In Chapter 6, we introduce some recent developments in the area.

CHAPTER 2

Digital Image

In this chapter, we review the basic concepts related to digital images. The fact that the basic information to be stored in a photographic digital image is light intensity motivates the discussion on light modeling and measuring.

2.1 DIGITAL IMAGE

A *color image* can be mathematically modeled as a function $f : U \subset R^2 \rightarrow C \subset R^3$. Where its *domain* U , the image plane, is usually defined by a rectangular subset of R^2 , while its *range* C is a subset in R^3 in which values are related to some color space.

The basic element of a color image is the color information at each pixel locations $(u, v) \in U$ in the image plane, that is, $c = f(u, v)$. If the instrument used to capture the image is a photographic camera, a photographic image is obtained, that is, a projection of a real scene that passed through a lens system to reach a photosensitive area in the image plane. In a photographic image, light intensity information c is stored at each measurement point $(u, v) \in U$. The measured intensity values depend upon physical properties of the scene being viewed and on the light source's distribution as well as on photosensitive sensor characteristics.

A *digital image* is the discrete representation of the two-dimensional continuous color image. Pixel coordinates are related to image spatial resolution while color resolution is determined by how color information is captured and stored.

In order to digitize the image signal that reaches the sensor, a *sampling* operation is carried on. *Sampling* is the task of converting the continuous incoming light into a discrete representation. Photographic digital sensors are physical implementations of signal discretization operators [14].

To form a digital image, the scene is sampled at a finite number of points. In practice, a photographic digital sensor consists of a set of sensor elements spatially distributed in the image plane that measures incoming light simultaneously. The *spatial resolution* is the number of points where the signal was sampled per unit area. Each sensor element is responsible to sample the incoming light flux and store it as digital value. The sensor conversion of incoming flux into a discrete finite subset of the color space C defines the color space discretization. The *color resolution* of an image is limited by the number of bits used to store color information.

4 HIGH DYNAMIC RANGE IMAGE RECONSTRUCTION

To visualize the image, a *reconstruction* operation recovers the original signal from samples. Ideally, the reconstruction operation should recover the original signal from the discretized information; however, the result of reconstruction frequently is only an approximation of the original signal. Imaging devices, such as digital projectors, monitors, and printers, reconstruct the discrete image to be viewed by the observer. It should be noted that the observer's visual system also plays a role in the reconstruction process.

2.2 MODELING LIGHT AND COLOR

The physics of light is usually described by two different models, *Quantum Optics* describes photons behavior, and *Wave Optics* models light as an electromagnetic wave [13]. The most relevant light characteristics that will be useful to us are well described by the electromagnetic model; *Geometric Optics* is an approximation of wave optics. We will adopt the electromagnetic approach and geometric optics when it is convenient.

Light sources radiate photons within a range of wavelengths. The energy of each photon is related to its wave frequency by the *Planck's* constant h , that is, $E = hf$. Once the frequency is determined, the associated wavelength λ is also known through the relation $c = \lambda f$. The emitted light can be characterized by its *spectral distribution* that associates to each wavelength a measurement of the associated radiant energy, as illustrated in Fig. 2.1(a). A source of radiation that emits photons all with the same wavelength is called *monochromatic* [14].

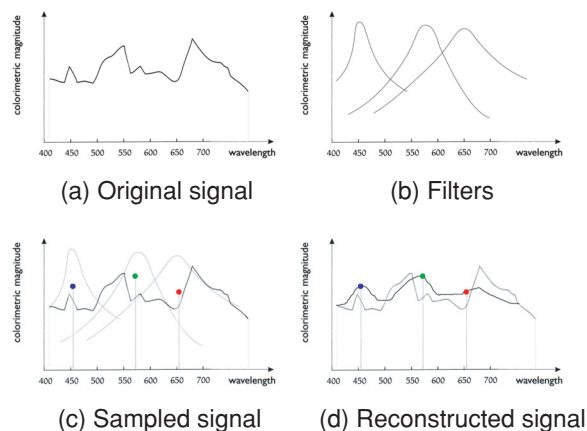


FIGURE 2.1: The original spectral distribution of an emitted light is shown in (a), the spectral response function of the three filters used to sample the signal are shown in (b), the sampled values are in (c) and the reconstructed signal is shown in (d).

A photosensitive sensor is characterized by its *spectral response function* $s(\lambda)$. If a sensor is exposed to light with spectral distribution $C(\lambda)$, the resulting measurable value is given by

$$w = \int_{\lambda} C(\lambda)s(\lambda)d\lambda.$$

The human eye has three types of photo sensors with specific spectral response curves. Light sensors and emitters try to mimic the original signal with respect to human perception. The standard solution adopted by industry is to use red, green, and blue filters as *primary colors* to sample and reconstruct the emitted light, illustrated in Fig. 2.1(b). The interval of wavelengths perceived by the human eye is between the range of 380 nm and 780 nm, known as the *visible spectrum*. The spectrum of emitted light together with sensor response curves defines the color perceived by an human observer or an imaging system.

2.3 MEASURING LIGHT

The intensity value registered by a sensor is a function of the incident energy reaching it. It corresponds to the integration of the electromagnetic energy flux both in time and in a region of space that depends upon the shape of the object of interest, the optics of the imaging device and the characteristics of the light sources.

TABLE 2.1: Radiometric Quantities and Their Photometric Counterparts [13]

QUANTITY DESCRIPTION	RADIOMETRIC QUANTITY [UNITY]	PHOTOMETRIC QUANTITY [UNITY]
Q : basic quantity of energy transported by the wave	Radiant Energy [Joule] $J = \frac{\text{kg m}^2}{\text{s}^2}$	Luminous Energy [talbot]
Energy per unit of time $\Phi = \frac{dQ}{dt}$	Radiant Flux [Watt] $W = \frac{J}{s}$	Luminous Flux [lumens] $\text{lm} = \frac{\text{talbot}}{s}$
Flux per solid angle (ω) $I := \frac{d\Phi}{d\omega}$	Radiant Intensity $\frac{W}{\text{sr}}$	Luminous Intensity [candelas] $\text{cd} = \frac{\text{lm}}{\text{sr}}$
Flux per area $u := \frac{d\Phi}{dA}$	Irradiance/Radiosity $\frac{W}{\text{m}^2}$	Illuminance/Luminosity [lux] $\text{lx} = \frac{\text{lm}}{\text{m}^2}$
Flux through a small area from a certain direction $L := \frac{d^2\Phi}{\cos\theta \cdot dA \cdot d\omega}$	Radiance $\frac{W}{\text{m}^2 \text{sr}}$	Luminance [nit] $\text{nit} = \frac{\text{cd}}{\text{m}^2}$

6 HIGH DYNAMIC RANGE IMAGE RECONSTRUCTION

Radiometry is the field that study electromagnetic energy flux measurements. The human visual system is only responsive to energy in a certain range of the electromagnetic spectrum, that is the *visible spectrum*. If the wavelength is in the visible spectrum, the radiometric quantities are also described in *photometric* terms. Photometric terms are simply radiometric terms weighted by the human visual system spectral response function [8].

In the literature, the description of the visible spectrum can come in radiometric quantities as well as in photometric quantities. This can be a source of confusion and a table relating both quantities is given in Table 2.1.

Most radiometric quantities can be measured in practice with lab instruments. Digital photographic cameras can also be used to measure light, but since they were projected to produce beautiful images and not to the measurement instruments, their nonlinear response to light intensity must be characterized. This is a fundamental step in HDRI recovery known as *camera characteristic function recovery*. In the next chapter, we will discuss digital cameras and calibration in more detail.

CHAPTER 3

Imaging Devices and Calibration

As we mentioned in the previous chapter, a digital camera is a device containing a sensor consisting of a grid of photosensitive cells that convert incident radiance into digital values. In this chapter, we study in detail the issues related to image devices, camera calibration, and the relationships with reconstruction of high dynamic range images.

3.1 DIGITAL PHOTOGRAPHY

A digital photograph is acquired by exposing the camera sensor to light during a certain period of time, called *exposure time*. During the exposure time, the sensor keeps collecting charge. At the end, the total electric charge collected is converted into digital brightness values. In Fig. 3.1, the effect of varying exposure time while keeping all other camera parameters fixed is illustrated.

The fundamental information stored in a digital image is the pixel *exposure*, that is, the integral of the incident radiance on the exposure time. A reasonable assumption is that incident radiance is constant during the exposure time, specially when small exposure times are used. Thus, exposure is given by the product of incident radiance by the total exposure time. The incident radiance value w_{ij} is a function of the scene's radiance, optical parameters of the system and the angle between the light ray and system's optical axis.

The most obvious way to control exposure is by varying exposure time, that is, by varying the total time that the sensor keeps collecting photons; but other camera parameters can also be controlled to alter exposure in different ways:

- controlling lens aperture;
- changing film/sensor sensitivity (ISO);
- using neutral density filters;
- modulating intensity of light source.

To vary time (as illustrated in Fig. 3.1) and lens aperture is easy and all professional and semi-professional cameras have these facilities. The disadvantages are related to limitations in applications since long exposures can produce motion blur while lens aperture affects the depth of focus, which can be a problem if the scene has many planes of interest. Film and

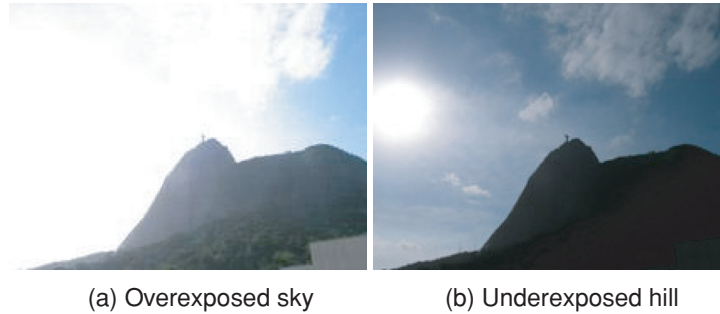


FIGURE 3.1: These images illustrate the resultant-acquired images when the exposure time is changed. In (a) the sensor has been exposed longer than in (b).

sensor sensitivity can be altered but the level of noise and graininess are also affected. Density filters are common photographic accessories but its usage depends on an implementation in hardware, or to be manually changed, which may not be practical. The use of controllable light source is tricky since intensity change depends on the distance of objects to the light source, that is, fails to change constantly on the scene and, in addition, produces shadows.

The photographic industry has been studying acquisition, storage and visualization of images for chemical emulsions since the beginning of the history of photography. Many concepts concerning image quality and accuracy were established since then [1, 2, 3]. Technical information about photographic material is available for reference in data sheets provided by manufacturers. The standard information provided is storage, processing, and reproduction information, as well as its technical curves, shown in Fig. 3.2.

Film technical curves guide the characterization of emulsions and are the technical base to chose an emulsion adequate for each scene and illumination situation. To characterize emulsion light intensity response, the useful curve is the *characteristic response curve*. *Spectral response curves* considers problems directly related to color reproduction. The *MTF curve* describes the spatial frequency resolution power of the sensible area.

The classical characterization of emulsions can also guide the study of digital sensors, although this information is usually not provided by digital sensors manufacturers. We now turn to the understanding of their role in digital image formation.

3.2 TONAL RANGE AND TONAL RESOLUTION

Considering light intensity, the behavior of an imaging sensor is described by its characteristic response function f . The distinct values registered by the sensor are the image *tones*.

In the classical photographic process, the film's photosensitive emulsion is exposed to light during the exposure time. The film is then processed to transform the emulsion's latent

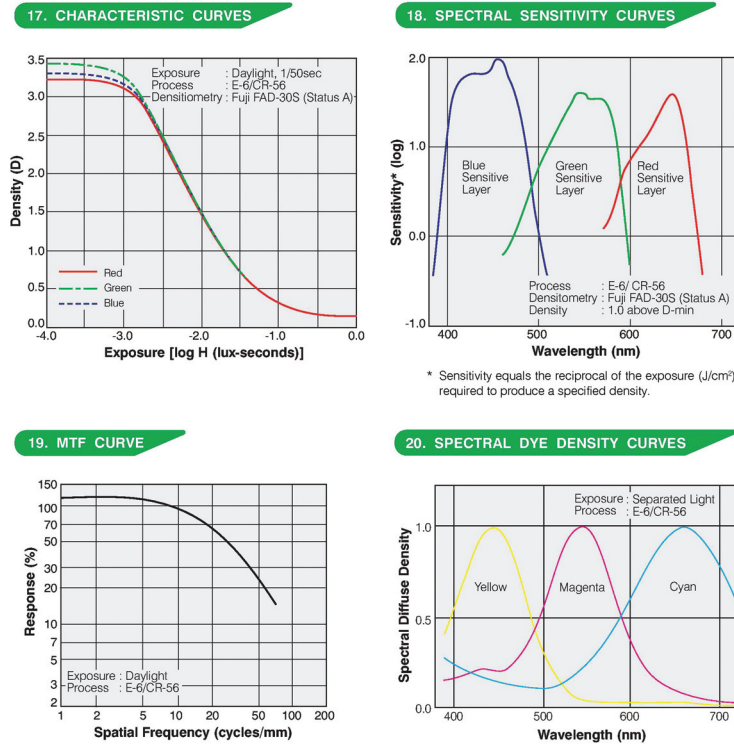


FIGURE 3.2: FujiChrome Provia 400F Professional [RHP III] data sheet, from FujiFilm.

image into *density* values. The concept of density is central in photography and relates the incoming and outgoing light; for films it is a transmission ratio $D_T = -\log_{10} T$ and for photo papers $D_R = \log_{10} 1/R$ is the reflection ratio, with both T and R in the interval $[0, 1]$ [2]. The characteristic curve of a film is the curve that relates the exposure and density. In Fig. 3.3, the characteristic curves of different film emulsions are compared. Observe that the film sensitivity to light (ISO) is different for each emulsion.

In digital photography, the film emulsions are replaced by photosensitive sensors. The sensor behavior depends on its technology. Usually, the stored electrical charge is highly linearly proportional to radiance values. In this case, if the sensor is capable to store a total number of d electrons, then the maximum number of distinct digital brightness values, that is, its *tonal resolution*, will potentially be equal to d . In practice, the digitization process influences on final image tonal resolution. Another important concept is that of *tonal range*, that is the difference between the maximum and the minimum exposure values registered by the sensor.

The leading sensors technologies are charge-coupled device (CCDs) and complementary metal oxide semiconductor (CMOS) sensors. The main difference between them is that in

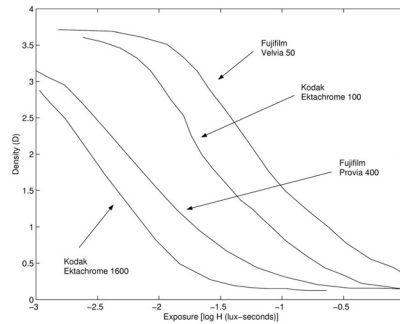


FIGURE 3.3: Film characteristic response curves compared.

CCDs every pixel charge is transferred through a very limited number of output nodes to be converted to voltage, as shown in Fig. 3.4(a). Differently, in CMOS sensors each pixel has its own charge-to-voltage conversion, shown in Fig. 3.4(b). This difference implies in several other differences ranging from noise level to manufacturing costs and sensor size [17].

CMOS sensors sensitivity to light is decreased in low light conditions because a part of each pixel photosensitive area is covered with circuitry that filters out noise and performs other functions. The percentage of a pixel devoted to collecting light is called the pixels *fill factor*. To compensate for lower fill-factors, micro-lenses can be added to each pixel to gather light from the insensitive portions of the pixel and focus it down to the photosensitive area.

Although the sensor’s natural behavior is linear, due to perceptive reasons the final brightness value stored in the image is nonlinearly related to radiance. This nonlinear behavior is characterized by the camera response curve.

The function $f : [E_{\min}, E_{\max}] \rightarrow [0, M]$ actually maps sensor exposure to brightness values, where E_{\min} and E_{\max} are, respectively, the minimum and the maximum exposure values

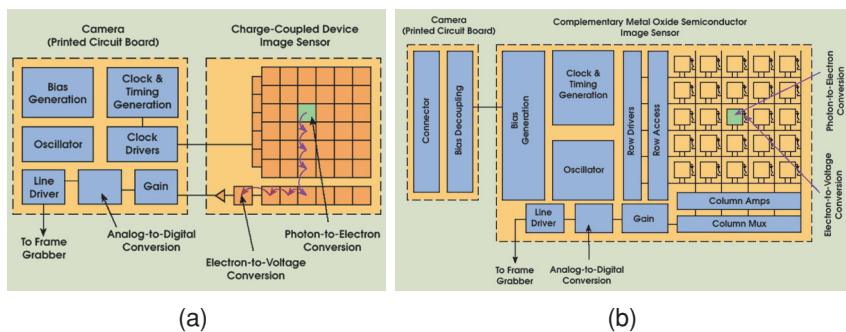


FIGURE 3.4: (a) CCD sensor and (b) CMOS sensor, from [17].

measurable by the sensor, and M is the maximum digitized value. The function f is in the core of the image formation process. In most cases, f is nonlinear and the application of f^{-1} is required to make meaningful comparisons between brightness values of differently exposed images.

Another important concept is that of *dynamic range*, it is the ratio of the highest to the lowest in a set of values; in the image context, these values are light intensity values. The fact that the range is dynamic is due to the possibility to control the exposure by varying camera parameters, thus changing the maximum and the minimum radiance values related to the same exposure range.

In photography, dynamic range—also referred to as film or photopaper *latitude*—is given in terms of stops, which is a \log_2 scale. Films produce a density range of about seven stops (that is, 128:1, or two orders of magnitude in base 10). Photographic paper has a much lower dynamic range, equivalent to four or five stops (approximately 20:1). Several techniques are adopted in the printing process to overcome this gap. The design of photographic materials has evolved to the goal of optimal response for human viewing under a variety of viewing conditions, and is well known that contrast plays a huge role in achieving good images.

Sensors convert an analog signal into a digital signal, so its characteristics define the signal discretization step. The dynamic range defines the range of tones that can be stored by the chosen media. Digital sensors and displays, independent of their accuracy, represent a discrete interval of the continuous infinite range of real luminances, so tonal resolution is influenced by the number of bits n used to describe it.

The difference between tone resolution and the tonal range of radiances spanned in an image is resumed as follows:

- *Tonal range* is the total size of the intensity interval that can be perceived by the sensor.
- *Tonal resolution* is the sample density, that is, the number of tones represented given a fixed interval.

Tonal range can be changed without altering the number of bits n used to store intensity information while changing n not necessarily changes the total range; both changes have an influence on the final resolution. Intuitively, the total range is the maximum contrast reproduced by the media, while the resolution influences on the tonal smoothness reproduction.

The distinction between tonal range and tonal resolution is subtle and important to the understanding of some simple operations that can be done to enhance images' tonal quality without using the powerful tools of HDRI concept.

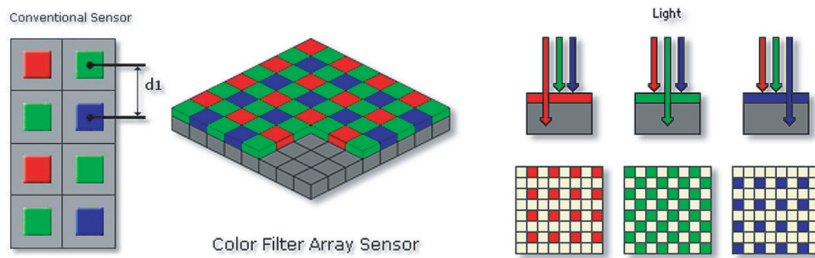


FIGURE 3.5: Bayer pattern (from <http://www.dpreview.com/learn/> by Vincent Bockeart).

3.3 COLOR REPRODUCTION

As mentioned before, light sensors and emitters try to mimic the scene's light signal concerning human perception; it is the human perception that is important concerning colors reproduction. Inspired on the trichromatic base of the human eye, the standard solution adopted by industry is to use red, green, and blue filters, referred as RGB base, to sample the input light signal and also to reproduce the signal using light-based image emitters. Printers work on a different principle, a discussion about them is out of the scope of this work.

Photographic color films usually have three layers of emulsion, each with a different spectral curve, sensitive to red, green, and blue light, respectively. The RGB spectral response of the film is characterized by spectral sensitivity and spectral dye density curves (see Fig. 3.2).

Electronic sensors are, by nature, sensitive to the entire visible spectrum and also to infrared wavelengths. In order to sample the input signal in RGB tristimulus base, colored filters are attached to the sensors. To each sensor pixel only one filter is attached, this implies the adoption of solutions like the usage of a Bayer pattern as shown in Fig. 3.5. In the Bayer pattern, the green channel is sampled twice more than the red and blue channels. This design choice is also based on the human perception.

Other solutions can also be adopted. The most common alternative is to use three sensors, one for each channel. However, since this solution is more expensive, most consumer cameras use Bayer patterns to sample light in the RGB base.

3.4 CAMERA CHARACTERISTICS

The characteristics of a camera need to be taken into account when reconstructing high dynamic range images. These characteristics are mainly related to sensor resolution and noise.

3.4.1 Spatial Resolution

The total size of the sensor, the size of an individual pixel and their spatial distribution determine the image spatial resolution and the resolution power of the sensor. The spatial resolution power

is related to the image sampling process. A tool of fundamental importance in the study of spatial resolution and accuracy issues is the sampling theorem [14].

[The Shannon–Whittaker Sampling Theorem]

Let g be a band-limited signal and Ω the smallest frequency such that $\text{sup } \hat{g} \subset [-\Omega, \Omega]$, where \hat{g} is the Fourier transform of g . Then g can be exactly recovered from the uniform sample sequence $\{g(m\Delta t) : m \in \mathbb{Z}\}$ if $\Delta t < 1/(2\Omega)$.

In other words, if the signal is bandlimited to a frequency band going from 0 to ω cycles per second, it is completely determined by samples taken at uniform intervals at most $1/(2\Omega)$ seconds apart. Thus, we must sample the signal at least two times every full cycle [14]. The sampling rate $1/(2\Omega)$ is known as the *Nyquist limit*. Any component of a sampled signal with a frequency above this limit is subject to *aliasing*, that is, a high frequency that will be sampled as a low frequency.

The number of sensor's pixels defines the image grid, that is, its spatial resolution in classical terms; but their physical size and spatial distribution also influences on the resolution power of the imaging device. Since the pixel spacing δ is uniform, the sensor Nyquist frequency is given by $1/(2\delta)$. Note that the adoption of Bayer pattern alters this δ -value altering sensor Nyquist frequency for each color channel.

Between light and sensor there is the camera lens system, which has its own resolution that influences on the final camera resolution. Lenses, including eye, are not perfect optical systems. As a result when light passes through it undergo a certain degree of degradation. The modulation transfer function (MTF) (see Figs. 3.6 and 3.2) shows how well a spatial frequency information is transferred from object to image. It is the Fourier transform of the

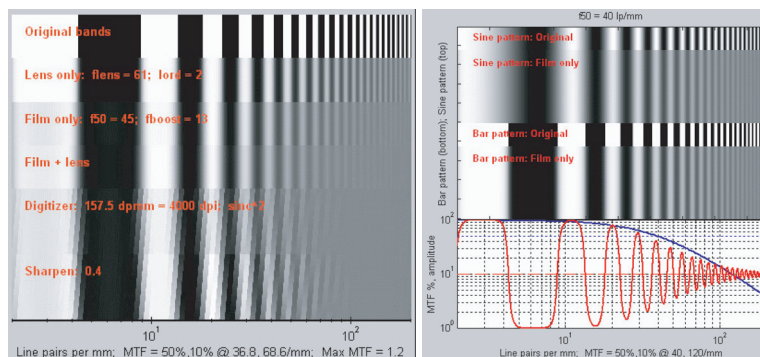


FIGURE 3.6: Images taken from <http://www.normankoren.com/Tutorials/MTF.html> illustrate the effects of MTF on the input target. The blue curve below the target is the film MTF, expressed in percentage; the red curve shows the density of the bar pattern.

14 HIGH DYNAMIC RANGE IMAGE RECONSTRUCTION

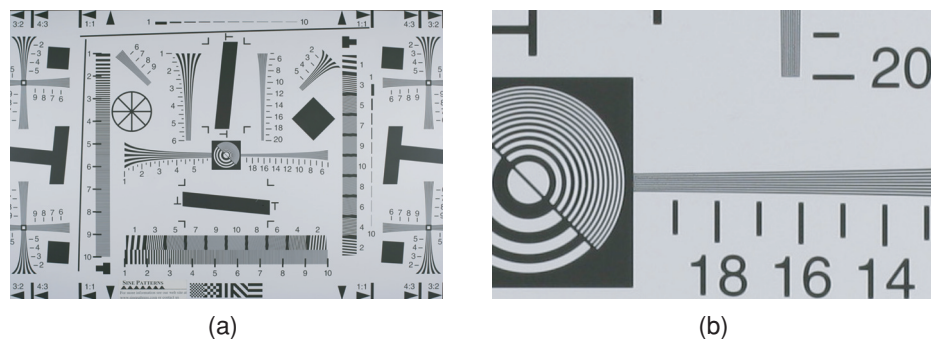


FIGURE 3.7: (a) Camera spatial resolution power test target. (b) A detail of the same image. The camera used in this test was a Canon EOS 350D and the image is from the site <http://www.dpreview.com>.

point spread function (PSF) that gives the scattering response to an infinitesimal line of light and is instrumental in determining the resolution power of a film emulsion or a lens system.

Lens and film manufacturers provide the MTF curves of their lenses and film emulsions.

In summary, image spatial resolution power is determined by imaging system lenses and sensor's pixels physical spacing. Usually, if the chosen lens is of high quality, pixel dimension and spacing is the critical information to be considered to evaluate spatial resolution power.

Spatial Resolution Power

To complete the characterization of a camera device, its spatial resolution power should be considered. This issue is related to the characterization of its MTF curve. In Fig. 3.7, the image of a test target used to analyze the camera resolution power is shown. It can be observed that the finest spatial frequencies are not well reproduced in the acquired image.

3.4.2 Noise, Aberrations, and Artifacts

Many distortions on measurement can be caused by electric phenomena such as dark current, thermal noise, charge overflowing to neighboring sensors, etc. Dark current means that a pixel may exhibit nonzero measurements even when there is no incoming photon. The longer the exposure time is, the more the dark current noise is accumulated. Cooling the sensor can be of great help since noise can double with every increase in temperature of about 6 K [13]. Bright spots can create large currents and the charge overflows to neighboring pixels leading to blooming artifacts.

Concerning pixel size, small pixels respond to fewer photons and can hold fewer electrons. Thus, although they allow for finer spacing, they suffer from increased noise, that is, poorer signal-to-noise ratio (SNR), reduced exposure range (fewer f-stops), and reduced sensitivity (lower ISO speed). Large pixels have good SNR, ISO speed and exposure range, but suffer from aliasing.

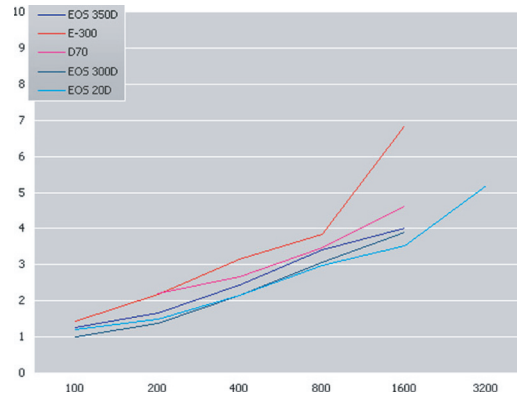


FIGURE 3.8: Some cameras’ noise compared. Indicated ISO sensitivity is on the horizontal axis of this graph, standard deviation of luminosity (normalized image) on the vertical axis. Image from the site <http://www.dpreview.com>.

Concerning sensors’ total size, a well-known issue for large format photography lovers, small sensors are affected by lens diffraction, which limits image resolution at small apertures—starting around $f/16$ for the 35 mm format. At large apertures— $f/4$ and above—resolution is limited by aberrations; these limits are proportional to the format size. Large sensors are costly. For reference, sensor diagonal measurements are 43.3 mm for full frame 35 mm film, up to 11 mm for compact digital cameras, and 22 mm and over for digital SLRs. Large format cameras are famous for their image resolution power.

Digital image post-processing can introduce artifacts. For instance, a side effect of Bayer pattern adoption is that the reconstruction of RGB values for each pixel uses information of neighboring pixels, the spatial measurement displacement can then introduce chromatic artifacts.

Noise

Sensor noise also influences on the image formation process. In this work, we referred to technical references of the adopted devices to choose parameters that minimize sensors noise. In Fig. 3.8, the behavior of noise with respect to the chosen ISO sensitivity for different camera models is illustrated.

3.5 CAMERA CALIBRATION

Camera calibration can be divided into three aspects: *geometric calibration*, *photometric calibration* and *spectral calibration*. For HDR reconstruction, the important aspects are photometric and spectral calibrations.

3.5.1 Photometric Calibration

Considering devices' photometric behavior, it has been already mentioned that, given a sensor's *spectral response function* $s(\lambda)$, if a sensor is exposed to light with spectral distribution $C(\lambda)$ the actual incoming value to be registered by the sensor is given by

$$w = \int_{\lambda} C(\lambda)s(\lambda)d\lambda.$$

It is also known that sensors pixels ij respond to exposure values

$$E_{ij} = w_{ij}\Delta t,$$

where Δt is the exposure time. Consequently, the actual digitized value d_{ij} is a function of the values w_{ij} . Thus, a full sensor photometric calibration should characterize the response function

$$d_{ij} = f(w_{ij}\Delta t)$$

as well as the RGB filters spectral functions $s(\lambda)$.

Note that the signal $C_{ij}(\lambda)$ cannot be recovered unless the calibration is done to each monochromatic wavelength λ and $s(\lambda)$ is known. In addition, it is possible that different input signals $C_{ij}(\lambda)$ at pixel ij produces equal responses w_{ij} , that is, the signals are sensor's *metameric* signals.

3.5.2 Spectral Calibration

The RGB values recorded for a color patch depends not only on light source spectral distribution and scenes reflective properties but also on the spectral response of the filters attached to camera sensors. To interpret meaningfully the RGB values, each spectral distribution should be characterized separately.

An absolute spectral response calibration is the complete characterization of RGB filter's, spectral behavior, that is, to recover $s(\lambda)$.

To the graphical arts and printing industry, color values have to be comparable in order to achieve consistent colors throughout the processing flow. What is done in practice is to adopt a color management system (CMS) to ensure that colors remain the same regardless of the device or medium used. The role of the CMS is to provide a profile for the specific device of interest that allows us to convert between its color space and standard color spaces. Usually, standard color charts are used to characterize the device color space, in this case, a photograph of the given chart is taken and based on the registered RGB values the device color space is inferred. The core information of a device profile is in most cases a large lookup table which

allows us to encode a wide range of transformations, usually nonlinear, between different color spaces [13].

Another issue related to color calibration is the compensation of light source spectral distribution. If scene illumination is different from white, what occurs with the most common illuminants like incandescent and fluorescent lighting, then the measured values will be biased by the illuminant spectral distribution.

The human eye has a chromatic adaptation mechanism that preserves approximately the colors of the scene despite the differences caused by illuminants. Digital imaging systems cannot account for these shifts in color balance, and the measured values should be transformed to compensate for illuminant chromatic distortions.

Many different algorithms can perform color balancing. A common approach usually referred as *white balance* is a normalize-to-white approach. There are several versions of white balance algorithm, but the basic concept is to set at white (W_R, W_G, W_B) a point or a region that should be white in the real scene. One version of the white balance algorithm sets values ($\max(R), \max(G), \max(B)$), the maximum values of the chosen white region, at a reference white (W_R, W_G, W_B). The colors in the image are then transformed using

$$(R', G', B') = \left(\frac{W_R}{\max(R)} R, \frac{W_G}{\max(G)} G, \frac{W_B}{\max(B)} B \right).$$

CHAPTER 4

HDR Reconstruction

In this chapter, we discuss the central problem of our study: *How to reconstruct a high dynamic range image using a digital camera?* That is, we would like to obtain the radiance map of a real 3D scene using image acquisition devices and considering the limitations of current technology.

Here we will focus on the conceptual understanding of the problem and its mathematical formulation. This approach has various practical benefits: on one hand, such a formulation is the key to derive effective computational procedures for generating HDR images with available equipment; on the other hand, such knowledge of basic principles is important for the development of next-generation solutions.

In order to acquire the radiance map, we will use a digital camera to measure radiance information of a scene. There are two main problems which we have to address: the first problem is that current cameras are designed to be only *sensing devices* and not *measuring devices*. We can say that the basic difference between them is that a measuring device is a properly calibrated sensing device. So, we need to calibrate! The second problem is that camera sensors have limited operating range and cannot capture the full range of radiances in a typical 3D scene. So, we need multiple exposures!

Therefore, the solution of the problem must involve photometric calibration to estimate the sensor's spectral response function, as well as, the use of multiple images differently exposed such that the range of scene radiances is completely covered. Note that, once the camera is calibrated, it is immediate to compute radiance values from the response function, given the measured intensity and the corresponding exposure.

4.1 PROBLEM STATEMENT

Photometric camera calibration is responsible for the characterization of the intensity response function f . As the image values d_{ij} are nonlinearly related to scene radiance values w_{ij} , it is mandatory to recover the characteristic sensor response function f in order to linearize data and perform meaningful comparisons between differently exposed d_{ij} values. Since f is reasonably assumed to be monotonically increasing, thus its inverse f^{-1} is well defined, making possible to compute w_{ij} from d_{ij} .

The recovery of f from observed data has been extensively studied in recent years. As stated previously, most methods are based on the usage of a collection of differently exposed images of a scene as input [4, 7, 10, 12].

A collection of K differently exposed pictures of a scene acquired with known variable exposure times Δt_k gives a set of $d_{ij,k}$ values for each pixel ij , where k is the index on exposure times. Although f is modeled as a continuous function, what can be observed are its discrete values registered by the sensor. The discrete response function \hat{f} associated to f includes in its modeling important sensors characteristics such as noise.

Considering sensor's noise η_{ij} , the actual value to be digitized is given by

$$z_{ij,k} = E_{ij,k} + \eta_{ij} = w_{ij}\Delta t_k + \eta_{ij}.$$

As the digitization function is discrete, if $z_{ij,k} \in [I_{m-1}, I_m)$, where $[I_{m-1}, I_m)$ is an irradiance interval, then $d_{ij,k} = \hat{f}(z_{ij,k}) = m$. The discrete response function \hat{f} is then

$$\hat{f}(z) = \begin{cases} 0 & \text{if } z \in [0, I_0), \\ m & \text{if } z \in [I_{m-1}, I_m), \\ 2^n & \text{if } z \in [I_{2^n-1}, \infty), \end{cases}$$

where $m = 0, \dots, 2^n$, with n the number of bits used to store the information (in practice, the maximum is not required to be equal to 2^n , but here we will consider this for notation simplicity).

The monotonically increasing hypothesis imposes that $0 < I_0 < \dots < I_m < \dots < I_{2^n-1} < \infty$. Thus, an inverse mapping can be defined by $\hat{f}^{-1}(m) = I_m$. The z values mapped to 0 are underexposed, while the z values mapped to 2^n are overexposed.

If $\hat{f}(z_{ij,k}) = m$ then $\zeta_{ij} = I_m - z_{ij,k}$ is the quantization error at pixel ij , thus

$$\begin{aligned} \hat{f}^{-1}(m) &= z_{ij,k} + \zeta_{ij} \\ &= w_{ij}\Delta t_k + \zeta_{ij}, \\ \hat{f}^{-1}(m) - \zeta_{ij} &= w_{ij}\Delta t_k \\ w_{ij} &= \frac{\hat{f}^{-1}(m) - \zeta_{ij}}{\Delta t_k}. \end{aligned}$$

To summarize, the scene radiance values are captured by a digital camera to produce discrete intensity values, d_{ij} , for each image pixel (ij). The corresponding *scene radiances*, w_{ij} , go through a three-stage pipeline inside the camera. First, the image is exposed by the camera lens and shutter to produce the *exposed sensor irradiances*, $E_{ij} = w_{ij}\Delta t$, where Δt is the exposure factor. Next, exposed irradiances are acquired by the camera sensor to give continuous pixel values, $z_{ij} = f(E_{ij} + \eta_{ij}) = f(w_{ij}\Delta t + \eta_{ij})$, possibly corrupted by some noise η_{ij} . Finally, the sensor's output values are digitized by the camera's quantization circuitry to generate the discrete pixel values $d_{ij} = \hat{f}(z_{ij})$. The entire pipeline is illustrated in Fig. 4.1.

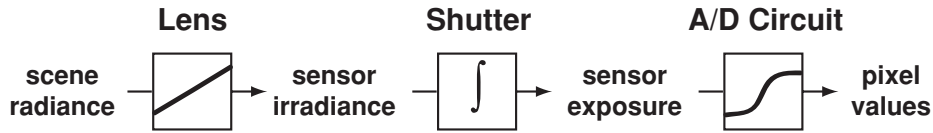


FIGURE 4.1: Image acquisition pipeline.

Here, the function f is the continuous response function of the camera, while the function \hat{f} is the discrete response function. In the following analysis, we may be interested either in computing f from z values, or of \hat{f} from d values, depending on whether we are concerned with the quantization step or not. For notation convenience, in each of these options, the function that is not under consideration will be assumed as the identity. We also may or may not consider the sensor noise model.

Thus, the problem can be stated as: “given a set of K images with exposure Δt_k , select a sufficient number of pixel values $d_{ij,k}$ such that the camera response function \hat{f} can be estimated.”¹

Then, compute radiance values the inverse of \hat{f} , as $w_{ij} = \frac{\hat{f}^{-1}(d_{ij,k})}{\Delta t_k}$.

4.2 ANALYSIS AND CLASSIFICATION

At this point, a question can be posed: *What is the essential information necessary and sufficient to obtain the camera characteristic response function from images?*

The answer to this question will make precise a part of the problem statement that says: “select a sufficient number of pixel values”. This is important because it defines the conditions used on the data to solve the problem. Different ways to collect and process the data will lead to the different methods for photometric calibration and HDR reconstruction. Each of these methods has its advantages and drawbacks. In what follows, we will investigate this issue and formulate a classification of the HDR reconstruction methods.

Intuitively, the photometric calibration amounts to creating a function that relates scene radiances with measured values. Note that this function must be invertible and the effect of exposure has to be factored out.

A conceptually simple way to generate such a function would be to capture a set of known radiance values and record the measured intensities. From these pairs (w, d) of data, it would be trivial to build a table that represents the function f . Figure 4.2 shows an example of sample values collected for calibration and plotted as an exposure–intensity graph.

¹Or equivalently, the same for f and $z_{ij,k}$.

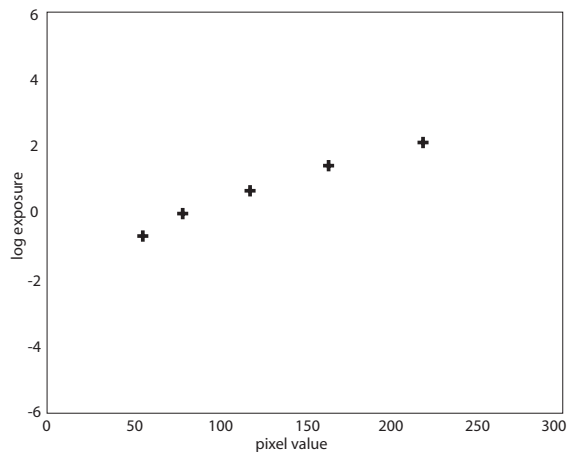


FIGURE 4.2: Exposure–intensity graph from samples.

If enough different irradiance values are measured—that is, at least one meaningful digital value is available for each mapped irradiance interval—then \hat{f}^{-1} mapping can be recovered for the discrete values.

This is, in fact, the basis of methods that are called *chart-based*, because they usually employ a calibration chart, such as the Macbeth chart [19], which include patches of known relative reflectances. The calibration procedure requires to capture a uniformly lit image of the chart in a controlled environment. The chief disadvantage of such methods is that they depend too much on the conditions of the setup. Thus, they are mostly suited to camera manufacturers.

Methods that do not use calibration charts are called *chartless*. Those are the techniques in which we will be interested here because they are more flexible and can integrate calibration with HDR reconstruction.

Chartless methods resort to a collection of differently exposed images captured without a priori restrictions on the subject. Note that, in this setting we do not have knowledge of the radiances in the scene.² We only know the exposures of each image and their pixel values.

Now, it becomes apparent that the exposure variation is the information which we should exploit in a chartless method. The basic idea is that if we vary the exposure Δt_k and record the pixel values $d_{ij,k}$ of a patch of unknown constant radiance w_{ij} , we effectively get a set of pairs $(E_k, d_{ij,k})$, where $E_k = w_{ij} \Delta t_k$ and $d_{ij,k} = f(E_k)$, which are samples of the camera response function—pretty much in the same way we had in a chart-based method. Since the unknown radiance is constant, it only affects the outcome by a multiplicative factor and can be disregarded.

²Actually, the goal of HDR reconstruction is to compute the radiances.

Once again, if we collect enough pairs of values, we are able to build a table that represents the response function.

There are four relevant issues that need to be analyzed in this basic approach.

First, we are assuming a constant radiance w_{ij} at a pixel (ij) for all the exposures Δt_k , with $k = 1, \dots, K$. This implies that both the camera and the scene must remain static while images are captured. This assumption can be alleviated by establishing a correspondence between pixel values across images with different exposures.

Second, not all recorded values $d_{ij,k}$ will be meaningful. Since the sensor has a limited operating range, values below the lower limit will be underexposed and mapped to 0, while values above the maximum limit will be overexposed and mapped to the highest value. In this respect, a weighting function could be used to give more importance to well exposed values.

Third, we are not explicitly taking into account sensor noise, which affects the result in various ways. For example, there is significantly more noise for low exposure values.

Fourth, the basic idea of this approach considers the values of just one pixel (ij) . This would require a large number of images to perform the calibration, and moreover, it would be wasteful since an image has many pixels exposed to different radiances. In principle, using information from all the pixels could reduce the number of images required for calibration and also make the estimation more robust. However, this is not a trivial task. In fact, the main difference between the various chartless methods lies in how they handle this issue. Generally, the solution will involve some kind of optimization.

In order to further discriminate the methods for HDR reconstruction, we can classify them according to three criteria: (1) type of function for camera response, (2) structure of the algorithm, and (3) strategy for selecting pixel values.

Regarding the type of the camera response function, the methods can use a *non-parametric* or a *parametric* representation. Non-parametric representations are usually discrete functions based on tabulated samples, as we mentioned above. These samples can be interpolated to obtain an approximation of the function in all its continuous domain. However, some assumptions must be imposed on the function such as continuity or smoothness restrictions. Parametric representations are often given by a polynomial function of high order. In some cases, parameterized models can be too restrictive and some real-world curves may not match the model. Grossberg and Nayar [12] study the space of camera response curves. It is observed that, although the space $W_{RF} := \{f \mid f(0) = 0, f(1) = 1 \text{ and } f \text{ is monotonically increasing}\}$ of normalized response functions is of infinite dimension, only a reduced subset of them arise in practice. Based on this observation it is created a low-parameter empirical model to the curves, derived from a database of real-world camera and film response curves.

Regarding the algorithmic structure, the methods can divide the computations in *two stages* or in *one stage*. Two-stage methods first estimate the response curve and subsequently

apply the curve to recover the radiance map. One-stage methods compute the response curve and the radiance map in the same step. Note that the pixel values determine both the calibration and HDR reconstruction. This interdependency can be exploited in different ways depending on the structure of the algorithm.

Regarding the pixel-value selection, the methods can be classified into *pixel-based* and *histogram-based*. Pixel-based methods select a subset of pixels in the image, assuming that there is a correspondence between scene points for each pixel location in the sequence of exposed images. Histogram-based methods work directly with values from accumulated histograms of the images, assuming that there is a correspondence between the M brightest pixels in each image of the sequence.

4.3 MAIN METHODS

In the rest of this chapter, we will study in detail the main methods for reconstruction of high dynamic range images. We will organize them according to the classification given in the previous chapter. An overview of these methods can be seen in the following table:

- Two-stage
 - Non-parametric
 - * Pixel-based [5]
 - Parametric
 - * Pixel-based [21, 23]
 - * Histogram-based [10–12]
- One-stage
 - Non-parametric
 - * Pixel-based [27].

4.3.1 Two-Stage, Non-Parametric, Pixel-Based Methods

Arguably, the most popular technique for HDR reconstruction is the method of Debevec and Malik. Their SIGGRAPH 1997 paper [4] was very influential in the graphics research community and gave the initial push to make HDR applications viable.

This technique solves the HDR reconstruction problem using a two-stage algorithm that estimates a non-parametric response function from image pixels and then recovers the radiance map. The algorithm gives a simple and robust solution to combine the pixel information from different exposures and image locations using a linear optimization.

The intuition behind the algorithm can be understood as follows: consider the sequence of values $d_{ij,k}$ from a pixel at image location (ij) with exposures Δt_k with $k = 1, \dots, K$. The

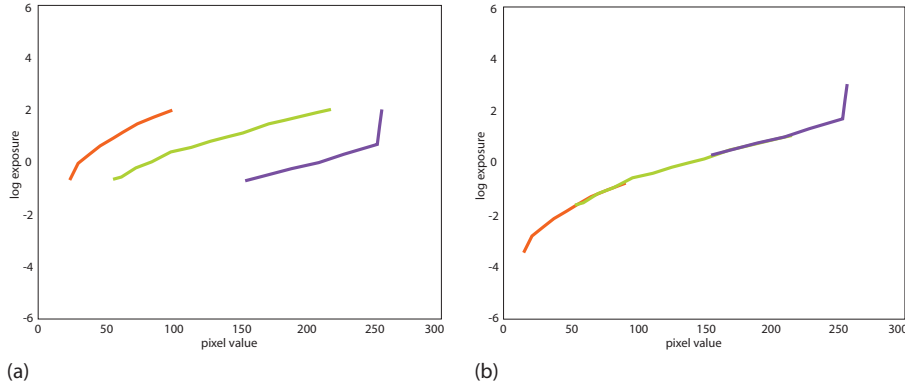


FIGURE 4.3: Plot of the reconstructed response function from samples at three pixel locations and five different exposures.

pairs $(d_{ij,k}, w_{ij}\Delta t_k)$ form a piece of the response curve indetermined up to a multiplicative constant given by the radiance w_{ij} . The same can be stated for each image location. So, all the pixel locations in the images provide us with many pieces of the curve. The problem is that, because the constants w_{ij} are different and unknown, it is not clear how these fragments fit together. Therefore, the problem is to find a curve that joins these pieces smoothly into a unique intensity response function. Figure 4.3 illustrates the problem, Fig. 4.3(a) shows the curve fragments corresponding to three different pixel locations with five different exposures, and Fig. 4.3(b) shows the response function that best fits the fragments.

Debevec and Malik solved this problem by relating the exposure values with the inverse of f . Note that the sensor irradiances $E_{ij} = w_{ij}\Delta t$ are given by $E_{ij} = f^{-1}(z_{ij,k})$. Since both right-hand sides of these equations are equal, we obtain the following expression:

$$w_{ij}\Delta t = f^{-1}(z_{ij,k}), \quad (4.1)$$

where the indices (ij) run over pixel locations and k runs over the set of differently exposed images.

In order to simplify further manipulation of the above expression, we can define $g = \ln f^{-1}$. Then, we have

$$g(z_{ij,k}) = \ln f^{-1}(z_{ij,k}) = \ln w_{ij} + \ln \Delta t_k. \quad (4.2)$$

In order to reconstruct the function g and the irradiances E_{ij} , we will need to minimize the function:

$$\mathcal{O} = \sum_{(ij)k} [g(z_{ij,k}) - \ln w_{ij} - \ln \Delta t_k]^2. \quad (4.3)$$

As we have a large number of equations and \mathcal{O} is quadratic in the E_{ij} and $g(z)$, minimizing it is a linear least-squares problem. This type of minimization can be successfully solved using singular value decomposition (SVD) or QR factorization. In the paper, they solved it using SVD with very good results.

However, in order to obtain more realistic response functions the above objective function needs to be modified in the following ways.

First, we have to add a constraint to establish a scale factor since the solution of the minimization problem is unique up to a single scale factor. So, we can make

$$g\left(\frac{z_{\min} + z_{\max}}{2}\right) = 0, \quad (4.4)$$

where $z_{\min} = 0$ and $z_{\max} = 255$. This constraint is enforced by adding equation (4.4) to the linear system.

Next, we introduce a weighting function in order to give more emphasis to well-exposed pixels. This will force the solution to have a better fit by anticipating the basic shape of the response function. Since $g(z)$ will typically have a steep slope near z_{\min} and z_{\max} , we expect that $g(z)$ will be less smooth and will fit the data more poorly near the extremes. The weighting function $h(z)$ is a “hat” function and is defined as

$$h(z) = \begin{cases} z - z_{\min} & \text{for } z < 1/2(z_{\min} + z_{\max}), \\ z_{\max} - z & \text{for } z > 1/2(z_{\min} + z_{\max}). \end{cases}$$

Finally, to control the smoothness of the response function, a new term is added to the objective function:

$$\lambda \sum_{z=1}^{254} [h(z)g''(z)]^2 \quad (4.5)$$

The parameter λ controls the importance we give to smoothness in our model. As our model is discrete, the second derivatives of g are approximated using finite differences $g''(z) = g(z-1) - 2g(z) + g(z+1)$.

With these changes, the final objective function becomes

$$\tilde{\mathcal{O}} = \sum_{(ij)k} \{h(z_{ij,k}) [g(z_{ij,k}) - \ln w_{ij} - \ln \Delta t_k]\}^2 + \lambda \sum_{z=1}^{254} [h(z)g''(z)]^2. \quad (4.6)$$

The formulation of the optimization problem suggests the use of every pixel in the image to compose the linear system. However, this will produce a large number of equations, even if the image size is small. Then, it is natural to ask ourselves how many images and pixels per image are needed to reliably recover the response curve. Clearly, we will need at least two

differently exposed images, but it will be better to use more. As far as the number of pixels, for a range of 256 values, it will be sufficient to use 50 pixels to obtain an overdetermined system of equations. The pixel selection is a delicate matter. In general, the selected pixels should satisfy the following properties:

1. have a good spatial distribution;
2. cover, as well as possible, the entire irradiance interval in the image; and
3. be selected from zones with small irradiance variance.

The solution of the minimization problem gives an optimal value for $g(z)$ in the range of $z \in [0, 255]$ for an 8-bit image. Then, to reconstruct the radiance map g is applied to recover the actual scene irradiance. This can be done by choosing an arbitrary image k of the sequence and use the expression

$$\ln w_{ij} = g(z_{ij,k}) - \ln \Delta t. \quad (4.7)$$

However, a more robust way is to use all the available information by means of the following expression:

$$\ln w_{ij} = \frac{\sum_k h(z_{ij,k})(g(z_{ij,k}) - \ln \Delta t)}{\sum_k h(z_{ij,k})}. \quad (4.8)$$

This algorithm have been implemented in the software HDRShop [15].

4.3.2 Two-Stage, Parametric, Pixel-Based Methods

In this chapter, we present two techniques for HDR reconstruction using parametric representations of the camera response function.

Mann and Picard Algorithm

The method developed by Mann and Picard was a seminal work in the field of HDR reconstruction. The technique proposed in their 1995 paper [23] introduced the main ideas of photometric self-calibration and generation of high dynamic range images by combining multiple images from a digital camera.

The algorithm models the response curve by a parametric function based on density curve for photographic films. It is given by the function

$$f(E) = \alpha + \beta E^\gamma, \quad (4.9)$$

where (α, β, γ) are the curve parameters. In photography applications, this function is interpreted as a “density versus exposure” curve, in which α indicates the minimum density, β is a scaling factor, and γ is the film contrast parameter.

In order to estimate the parameter α , they suggest taking an image with the lens covered. The value obtained will be subtracted from the other images.

To compute the parameter γ , the algorithm exploits the exposure variation between images and assumes a pixelwise correspondence of scene points.

Consider two images S_1 and S_2 taken respectively with different exposures e_1 and e_2 . The exposure ratio $R = e_1/e_2$ relates measurements in the two images as follows: $f(E) = f(RE)$, where E is some constant unknown irradiance. Therefore, if we start with a pixel p_0 in image S_1 with value $z_{0,1} = f(E_0)$, we know that its corresponding value in image S_2 will be $z_{0,2} = f(RE_0)$. On the other hand, by reciprocity, a different pixel p_1 in image S_1 with the same value $z_{0,2}$ will have a value $z_{1,2} = f(R^2 E)$ in image I^2 .

The algorithm uses this idea to build the sequence of value pairs

$$(z_0, f(E)), (z_1, f(RE)), (z_2, f(R^2 E)), \dots, (z_n, f(R^n E)) \quad (4.10)$$

by alternating the search for values in one image that map to corresponding values in the other.

Once the sequence (4.10) is generated, the parameter γ of the model is estimated by regression.

As in other methods, the response curve is determined up to a constant factor (the parameter β), that can be arbitrarily set to 1.

The main drawback of this method is that the parametric model of the camera response function, although inspired in the photographic film, is highly restrictive. Therefore, in the case of a general imaging system it will give only a qualitative calibration result.

Nonetheless, Mann and Picard introduced in their work many important concepts, such as the relation of quantization with noise models, and the use of histograms, that laid the ground for subsequent research.

Mitsunaga and Nayar Algorithm

The method of Mitsunaga and Nayar improves on previous work by proposing a more realistic parametric model of the camera response function.

In their algorithm [21], they approximate the response function by a higher order a polynomial g of degree P :

$$g(z) = \sum_{p=0}^P c_p z^p, \quad (4.11)$$

where the coefficients $\{c_p\}_{p=0,\dots,P}$ are the parameters of the model.

As in the Mann and Picard method, they exploit the exposure ratio between two subsequent images, S_k and S_{k+1} . One of the advantages of their technique is that the formulation of

the algorithm makes possible to re-estimate the exposure ratios as part of the numerical solution. This feature can be relevant in practical cases where the image exposures are not known exactly, such as when working with low-cost cameras.

The exposure ratio can be defined as

$$R_{k,k+1} = \frac{w \Delta t_k}{w \Delta t_{k+1}}. \quad (4.12)$$

However, by noting that $w_{ij} \Delta t_k = f^{-1}(z_{ij,k})$, we can write the exposure ratio also as

$$\frac{f^{-1}(z_{ij,k})}{f^{-1}(z_{ij,k+1})} = R_{k,k+1}. \quad (4.13)$$

Hence, we get the following equation:

$$f^{-1}(z_{ij,k}) - R_{k,k+1} f^{-1}(z_{ij,k+1}) = 0. \quad (4.14)$$

Then, by substituting $f^{-1}(z) = g(z) = \sum_p c_p z^p$ in equation (4.14), we arrive at an error function \mathcal{E} , in terms of the polynomial model, such that the minimizer of \mathcal{E} yields the best fitting parametric reconstruction curve.

The objective function of this optimization problem is

$$\mathcal{E} = \sum_{(ij)k} \left[\sum_p c_p z_{ij,k}^p - R_{k,k+1} \sum_p c_p z_{ij,k+1}^p \right]^2, \quad (4.15)$$

where the indices (ij) run over the image pixels, the index k enumerates the multiple exposures, and p indexes the coefficients of the polynomial approximation of the response function.

The above minimization is solved by setting to zero the partial derivatives of \mathcal{E} with respect to the polynomial coefficients.

$$\frac{\partial \mathcal{E}}{\partial c_p} = 0. \quad (4.16)$$

This gives us a system of $P + 1$ linear equations to be solved. However, this system is underdetermined, due to the fact that the response function can only be computed up to some arbitrary scaling factor. Therefore, in order to reduce the dimensionality of the linear system, we impose that $f(1) = 1$ and get the additional constraint

$$c_p = 1 - \sum_p c_p. \quad (4.17)$$

Intuitively, this restriction normalizes all measurements to the range $0 \leq z \leq 1$ and scales the maximum irradiance value to 1.

The final $P \times P$ linear system can be written as

$$\begin{aligned} & \begin{pmatrix} \sum_q d_{q,0}(d_{q,0} - d_{q,P}) & \cdots & \sum_q d_{q,0}(d_{q,0} - d_{q,P}) \\ \cdots & \cdots & \cdots \\ \sum_q d_{q,P-1}(d_{q,0} - d_{q,P}) & \cdots & \sum_q d_{q,P-1}(d_{q,0} - d_{q,P}) \end{pmatrix} \times \begin{pmatrix} c_0 \\ \cdots \\ c_{P-1} \end{pmatrix} \\ & = \begin{pmatrix} -\sum_q d_{q,0} - d_{q,P} \\ \cdots \\ -\sum_q d_{q,0} - d_{q,P} \end{pmatrix}, \end{aligned} \quad (4.18)$$

where, for notation simplicity, we use the multi-index $q = ij, k$ and define

$$d_{q,p} = z_{ij,k,p} - R_{k,k+1} z_{ij,k+1,p}. \quad (4.19)$$

The above formulation of the problem assumes that the exposure ratios $R_{k,k+1}$ are known precisely. However, if this is not the case, as mentioned previously, the method makes possible to improve the estimation of exposure ratios by recomputing them iteratively.

The iterative scheme is as follows: at each step (n) the previous exposure ratio estimates $R_{k,k+1}^{(n-1)}$ are used to compute next set of coefficients $c_p^{(n)}$. Then, these coefficients are used to update the estimates

$$R_{k,k+1}^{(n)} = \sum_{(ij)} \frac{\sum_p c_p^{(n)} z_{ij,k}^p}{\sum_p c_p^{(n)} z_{ij,k+1}^p}, \quad (4.20)$$

where the initial exposure ratios $R_{k,k+1}^{(0)}$ are part of the input to the method.

The algorithm have as stopping condition

$$|f^{(n)}(z) - f^{(n-1)}(z)| < \epsilon, \forall z$$

that indicates convergence of the method.

As in the other two-stage methods, once the camera response function is computed, the radiance map can be reconstructed from the exposure images and f^{-1} by applying the basic equation for one image S_K :

$$w_{ij} = \frac{f^{-1}(d_{ij,k})}{\Delta t_k}, \quad (4.21)$$

or using a more robust scheme that uses a combination of the information in all images, as mentioned in chapter 4.3.1.

This algorithm is implemented in the MacOS program Photosphere [25] by Greg Ward and in the RASCAL software [26] developed by the authors of the method.

4.3.3 One-Stage, Non-Parametric, Pixel-Based Methods

Robertson, Borman, and Stevenson developed a technique which estimates the camera response curve and reconstructs the radiance map in a single integrated stage. Their algorithm [27] employs a non-parametric representation of the response function and solves a discrete version of the problem by an iterative optimization process. The algorithm iteratively improves an estimate of the response function \hat{f} based on pixel values $d_{ij,k}$ and use this estimation to compute a better approximation of the irradiance values w_{ij} , until convergence.

This method has several advantages: first, since the representation is non-parametric, \hat{f} do not need to be assumed to have a shape described by some previously defined class of continuous functions; second, the formulation takes explicitly into account the sensor noise model; and third, the structure of the algorithm naturally uses the all information available from image pixels for the estimation of \hat{f} and reconstruction of $\{w_{ij}\}$.

The discrete version of the problem is modeled as follows: the measured value $z_{ij,k}$ is corrupted by additive sensor noise $\eta_{ij,c}$, so $z_{ij,k} = w_{ij}\Delta t_k + \eta_{ij,c}$. The discrete response function \hat{f} maps real irradiance values $w_{ij}\Delta t_k$ to integers m . Therefore, the reconstruction of radiance values defined by the inverse mapping $\hat{f}^{-1}(m) = I_m$ has a quantization error $\zeta_{ij} = I_{ij} - z_{ij,k} = I_{ij} - w_{ij}\Delta t_k$. Here, $\zeta_{ij} = \eta_{ij,c} + \eta_{ij,d}$ is a Gaussian random variable that accounts for the accumulated effects of sensor noise $\eta_{ij,c}$ as well as the dequantization uncertainty $\eta_{ij,d}$.

In this setting, the inverse response function is defined by the values I_m that are independent random variables and can be estimated using a maximum-likelihood (ML) approach. However, since these values depend on the unknown radiances w_{ij} they have to be estimated simultaneously.

The objective function to be minimized is

$$O(I, w) = \sum_{(i,j),k} \sigma(m)(I_m - w_{ij}\Delta t_k)^2, \quad (4.22)$$

where the function $\sigma(m)$ is a weighting function chosen based on the confidence on the observed data. In the original paper, $\sigma(m) = \exp(-4\frac{(m-2^{n-1})^2}{(2^{n-1})^2})$.

Gauss–Seidel relaxation can be used to minimize this equation. First, (4.22) will be minimized with respect to w and subsequently with respect to I .

The minimization with respect to the unknown w is done by setting the gradient $\nabla O(w)$ to zero, the optimum w_{ij}^* at pixel ij is given by

$$w_{ij}^* = \frac{\sum_k \sigma(m)\Delta t_k I_m}{\sum_k \sigma(m)\Delta t_k^2}. \quad (4.23)$$

32 HIGH DYNAMIC RANGE IMAGE RECONSTRUCTION

In the initial step f is supposed to be linear, and the I_m values are calculated using f . The second step iterate f given the w_{ij} . Again the objective function 4.22 is minimized, now with respect to the unknown I . The solution is given by

$$I_m^* = \frac{1}{\#(\Omega_m)} \sum_{((i,j),k) \in \Omega_m} w_{ij} \Delta t_k, \quad (4.24)$$

where

$$\Omega_m = \{((i, j), k) : d_{ij,k} = m\} \quad (4.25)$$

is the index set and $\#(\Omega_m)$ is its cardinality.

A complete iteration of the method is given by calculating (4.23) and (4.24), then scaling of the result. The process is repeated until some convergence criterion is reached. In the paper, the authors suggest using the rate of decrease in the objective function.

This technique is implemented as part of the PFStools software suite [24].

4.3.4 Two-Stage, Parametric, Histogram-Based Methods

Grossberg and Nayar developed a technique that, instead of using pixel values directly from the images, use the image histograms to estimate the camera response function. The main advantage of this method is that the image histograms summarizes the pixel-value information from the images and, at the same time, make the estimation process more robust to camera motion and movements in the scene.

The central result in this method is a function that relates pixel values in two images with different exposures, very much in the spirit of the Mann and Picard's method.

In their paper [10], the authors define the intensity mapping function $\tau : [0, 1] \rightarrow [0, 1]$ as the function that correlates the measured brightness values of two differently exposed images. This function is defined at several discrete points by the accumulated histograms H of the images, given by $\tau(d) = H_2^{-1}(H_1(d))$ ³, and expresses the concept that the m brighter pixels in the first image will be the m brighter pixels in the second image for all m .⁴ Then, the following theorem is derived:

[Intensity Mapping Theorem [10]]

The histogram H_1 of one image, the histogram H_2 of a second image (of the same scene) is necessary and sufficient to determine the intensity mapping function τ .

³Supposing that all possible tones are represented in the input image, the respective accumulated histogram H is monotonically increasing, thus the H^{-1} inverse mapping is well defined.

⁴ H and τ are considered as continuous functions although in practice they are observed at discrete points and their extension to continuous functions deserves some discussion.

The referred function τ is given by the relation between two corresponding tones in a pair of images.

Let

$$\begin{aligned} d_{ij,1} &= f(w_{ij}; \Delta t_1) \\ d_{ij,2} &= f(w_{ij}; \Delta t_2), \end{aligned} \quad (4.26)$$

then

$$\begin{aligned} d_{ij,1} &= f\left(\frac{f^{-1}(d_{ij,2})}{\Delta t_2} \Delta t_1\right) \\ &= f(\gamma f^{-1}(d_{ij,2})), \end{aligned} \quad (4.27)$$

where $\gamma = \frac{\Delta t_1}{\Delta t_2}$,
that is

$$\begin{aligned} d_{ij,1} &= f(\gamma f^{-1}(d_{ij,2})) \\ &= \tau(d_{ij,2}). \end{aligned} \quad (4.28)$$

Based on the above theoretical result they develop an algorithm which exploits the fact that τ , together with the exposure times ratio are necessary and sufficient to recover f . Here the conclusions were derived from an ideal camera sensor without noise. But note that τ is obtained observing the accumulated histograms, that are less sensitive to noise than the nominal values.

In [10], the images accumulated histograms are used to obtain the intensity mapping function τ . Then, a continuous f^{-1} is obtained assuming that it is a sixth-order polynomial, and solving the system given by equation $f^{-1}(\tau(d)) = \gamma f^{-1}(d)$ on the coefficients of the polynomial. Two additional restrictions are imposed: that no response is observed if there is no light, that is, $f^{-1}(0) = 0$, assuming also that $f : [0, 1] \rightarrow [0, 1]$, $f^{-1}(1) = 1$ is fixed, which means that the maximum light intensity leads to the maximum response. Note that there is no guarantee that the obtained f is monotonically increasing.

The usage of accumulated histograms has some advantages: all the information present in the image is used instead of a subset of pixels, the images need not be perfectly registered since spatial information is not present on histograms, and they are less sensitive to noise. Note that, if an histogram approach is used, the spatial registration between image pixels is not necessary to find f , but pixel correspondences cannot be neglected when reconstructing the radiance map.

4.4 EXPERIMENTS AND COMPARISON

In this chapter, we describe some experiments made with the main methods for photometric camera calibration and HDR reconstruction.

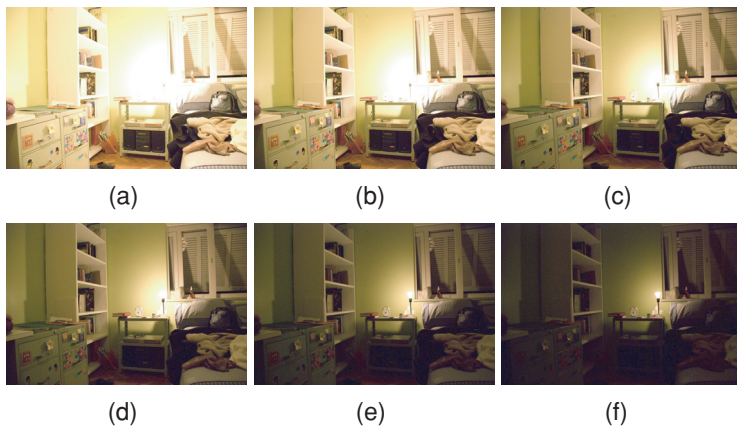


FIGURE 4.4: Input images to HDR reconstruction experiments.

For these experiments, we use as input the set of images shown in Fig. 4.4.

For each of the experiments, we reconstruct the camera response curve using one of the methods described in this chapter using different parameters of the method.

The first experiment is with the Debevec and Malik algorithm, where we varied the smoothness parameter λ . The result can be seen in Fig. 4.5. In Fig. 4.5(a) we set the parameter $\lambda = 0.1$, and in Fig. 4.5(b) we set the parameter $\lambda = 1.0$. Note that the response curve became smoother.

The second experiment is with the Robertson, Borman, and Stevenson algorithm, where we changed the variance of the weighing function. The result can be seen in Fig. 4.6. In Fig. 4.6(a) we set the parameter $\sigma = 16$, in Fig. 4.6(b) we set the parameter $\sigma = 2$, and in Fig. 4.6(c) we set the parameter $\sigma = 40$. Note that the response curve achieves the best regularity with $\sigma = 16$.

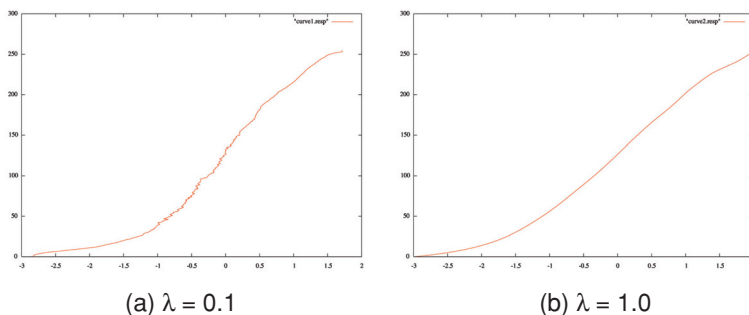


FIGURE 4.5: Response curves reconstructed with the Debevec and Malik algorithm with different values of the smoothness parameter.

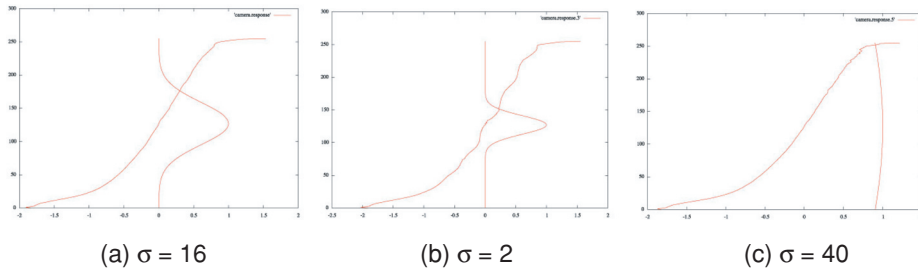


FIGURE 4.6: Response curves reconstructed with the Robertson, Borman, and Stevenson algorithm with different values of the weighting function.

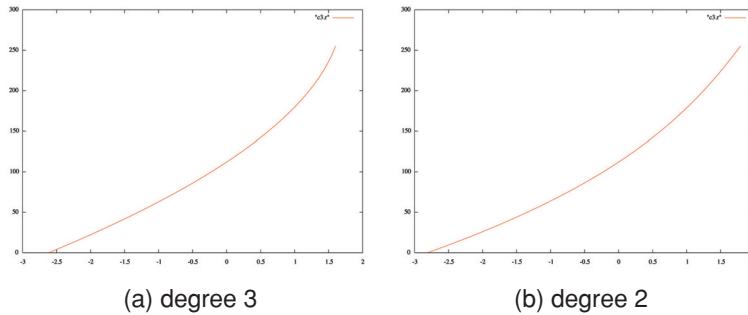


FIGURE 4.7: Response curves reconstructed the Mitsunaga and Nayar algorithm with polynomial reconstruction of different degrees.

The third experiment is with the Mitsunaga and Nayar algorithm. Were we changed the degree of the polynomial function. The result can be seen in Fig. 4.7. The reconstruction in Fig. 4.7(a) was with a degree 3 polynomial with coefficients $\{c_0 = -0.0144178, c_1 = 0.447053, c_2 = -1.0337, c_3 = 1.60107\}$. In Fig. 4.7(b), the reconstruction was with a degree 2 polynomial obtained by making $c_3 = 0$.

4.5 IMPROVING ROBERTSON'S METHOD

The HDR reconstruction algorithm by Robertson et al. described in chapter 4.3.3 is very appealing for practical usage. However, it can be improved with minor changes.

We observe that, as originally formulated, there is no guarantee that the values I_m obtained in (4.24) are monotonically increasing. Especially in the presence of noise this assumption can be violated. If I_m^* are not increasing, then the new w_{ij}^* can be corrupted, and the method does not converge to the desired radiance map. The correct formulation of the objective

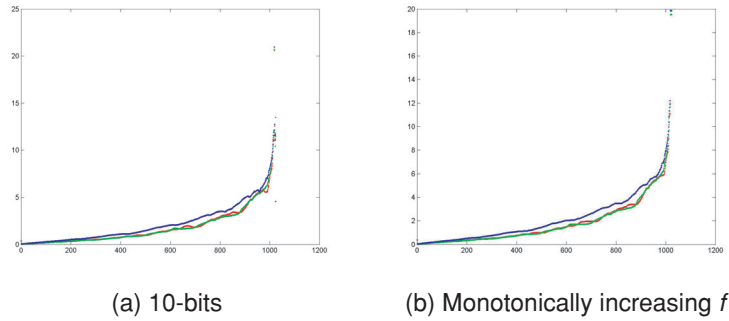


FIGURE 4.8: Output f 10-bits of pixel depth.

function should include the increasing restrictions:

$$O(I, w) = \sum_{(i,j),k} \sigma(m)(I_m - w_{ij} \Delta t_k)^2 \quad (4.29)$$

$$s.a. \quad 0 < I_0 < \dots < I_m < \dots < I_{2^n-1} < \infty.$$

This new objective function is not easily solved to the unknown I as the original one. In the next chapter, the iterative optimization method is applied to recover the f response function of the cameras used in the proposed experiments. The approaches used to deal with the increasing restrictions will then be discussed.

Another observation is that although the I_m values were modeled as the extreme of radiance intervals, the calculated I_m^* are an average of their correspondent radiance values.

As expected, using the original formulation of the algorithm the obtained function is not monotonically increasing. Especially where the input data is poor, the obtained f function is likely to be non-monotonically increasing. The heuristic adopted to guarantee that the function is monotonically increasing is very simple, it is based on linear interpolation, we simply ignore the values where some descent is observed and recalculate the values by linear interpolation considering the first non-descent occurrence. Figure 4.8(b) shows the final monotonically increasing f obtained from Fig. 4.8(a). To apply the linear interpolation, we work on $\log 2$ of the data that is more reasonably assumed to be well interpolated by linear parts.

CHAPTER 5

HDRI Acquisition and Visualization

In this chapter, we discuss practical issues related to acquisition and visualization of high dynamic range images.

5.1 DATA ACQUISITION

Algorithms that recover HDRI by software usually combine collections of LDRI. As discussed in chapter 4, the sensors' characteristic response curve is recovered from the input images and then applied to obtain radiance values, as illustrated in Fig. 5.1.

To generate the collection of differently exposed images, several exposure camera parameters can be controlled. In the reconstruction phase, the knowledge on which parameter has been changed should be considered since they affect image formation in different ways, e.g. introducing motion blur, defocus, etc.

To correctly reconstruct the radiance values, pixels correspondence between images is crucial. The correspondence depends on scenes features and temporal behavior as well as on which camera parameter is being changed to vary exposure. The pixel correspondence problem is usually stated as an optimization problem. It can be badly defined and hard to solve. A fast, robust, and completely automatic method for translational alignment of hand-held photographs is presented in [35].

The f recovery was discussed in chapter 4.3.4. With f in hand, the actual scene radiance values are obtained applying its inverse f^{-1} to the set of correspondent brightness values d_{ij}^k observed in the differently exposed images, where k is an index on the differently exposed images and ij are pixel coordinates. Different weights can be given according to the confidence on d_{ij}^k . If the pixel is almost over- or underexposure, a lower weight is given to it, augmenting the influence of the middle of the f curve, where sensors (and films) are well behaved. It is required that at least one meaningful digital value is available for each pixel, that is, at least one pixel value in a set of correspondent pixel has a good confidence on the measured data.

In case one has an HDR sensor, the knowledge of its characteristic function is necessary to recover the w values (we recall that w are the radiance values), but the correlation step is often unnecessary since only one exposure is enough to register the whole range of intensities

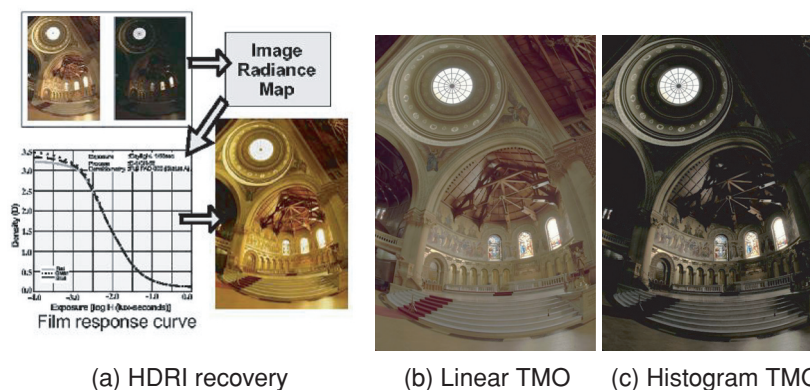


FIGURE 5.1: In (a) differently exposed LDR pictures from the same scene are processed to compose a HDR Radiance Map of the input images. A TMO is then applied to visualize the HDR image; figure (b) is visualized using linear TMO and figure (c) using histogram adjustment TMO.

present in the scene (at least what is expected from such devices). Scenes of high latitude range can be also synthetically generated by physically-based renderers like RADIANCE [34].

5.2 REPRESENTATION AND HDR FORMATS

The HDR Radiance Map representation is similar to the usual image representation. What changes is the accuracy on tone acquisition. The simplest way to enhance tone resolution is to augment the number of bits used to describe it, but there is a subtle difference between augmenting the tone resolution and enhancing the range of radiances spanned in an image. We can enhance resolution without a range enhancement and can enhance the range maintaining the original low resolution.

The main issue then is how to represent color information accurately and efficiently. In this context, the image representation is a separate problem. Nonetheless, some color spaces, such as Yuv, are better suited for image compression. In this chapter, we will discuss mainly how to represent color information at image pixel level and will not go into the discussion of image format standards.

A very comprehensive analysis of image encodings for HDR Radiance Maps has been written by Larson [16]. The following discussion is largely based on this work.

5.2.1 Color Encoding for HDR Maps

As pointed out above, HDR data requires accurate color representation. This could be done by just using more bits to encode color coefficient values. However, such option could be highly inefficient. For example, to represent 30 orders of magnitude would be necessary for

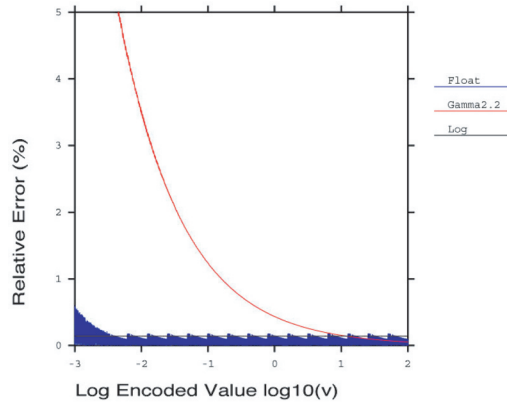


FIGURE 5.2: Error profile of log and floating point encodings.

an integer representation with 100 bits. So, the solution for an efficient representation is to use a representation suited for real numbers, such as floating point.

Since the digital format can only encode discrete quantities, quantization is inevitable. Thus, the key point is to control the error of a high range of values.

There are two main encodings for HDR values: the floating point encoding and the log encoding. The main difference between them is how the error is distributed over the range of values.

In the floating point encoding, a value v is represented using exponential notation as $x \cdot 2^y$, where x is the mantissa and y is the exponent.

In the log encoding, a value v is represented as $a \cdot (\frac{b}{a})^v$, where a and b are the minimum and maximum values of the output range.

Note that these two encodings can be tuned for a particular HDR application. For example, it may not be necessary to represent negative numbers for HDR values.

The main difference between the floating point and log encodings is how the error is distributed over the range of values.

It is easy to see that adjacent values in the log encoding differ by $(\frac{b}{a})^{\frac{1}{N}}$, where N is the number of quantization steps of the discretization. Therefore, the quantization error is constant throughout the whole range. This is in contrast with the floating point encoding, which does not have perfectly equal step sizes, but follow a sawtooth pattern in its error envelope. Figure 5.2 shows a graph comparing the error profile of these two encodings.

5.2.2 HDR Formats

Here we follow the convention that an HDR format is not tuned for a particular output device (i.e., the so-called *output referred standard*), but it is designed to represent the range of captures

scene values (i.e., the scene referred standard). Thus, we are interested in pixel encodings that extend over at least four orders of magnitude and can encompass the visible color gamut. They also should have a luminance step size below one resolution. An encoding format that meets these requirements will be able to represent an image with fidelity compatible with the human perception.

The main HDR formats are:

- Radiance RGBE,
- SGI LogLuv,
- ILM's OpenEXR,
- TIFF HDR encoding.

5.3 TONE MAPPING

Usually, HDR data are to be visualized on low dynamic range displays. The reduction of the range of image radiances to display brightness in a visual meaningful way is known as the tone mapping problem. Tone mapping operators (TMO) have been firstly proposed to solve the problem of visualization of synthetic images generated by physically based renderers before HDR from photographs became popular [18].

In [5], the tone mapping research is reviewed and TMOs are classified in three main classes: *spatially uniform time-independent*, *spatially uniform time-dependent*, and *spatially varying time-independent* operators. Another survey in TMO research is [22]. Recently, specific TMOs for HDR video visualization has been proposed. Below, some simple TMOs are described as well as the main reasoning that guides their intuition.

- *Linear mapping*: the simplest way to reduce the range of an HDRI is by using linear tone mapping, obtained as follows:

$$d = \frac{(w - w_{\min})}{R_w} R_d + d_{\min}, \quad (5.1)$$

where $R_w = (w_{\max} - w_{\min})$ and $R_d = (d_{\max} - d_{\min})$.

- *Histogram adjustment*: in [18] is proposed the histogram adjustment heuristic, it is inspired on the fact that in typical scenes luminance levels occur in clusters rather than being uniformly distributed throughout the dynamic range. The algorithm proposed is

$$d = P(w)R_d + d_{\min}, \quad (5.2)$$

where

$$\begin{aligned}
 P(B_w) &= \frac{[\sum_{b_i < w} b(b_i)]}{[\sum_{b_i < w_{\max}} b(b_i)]} \\
 &= \frac{H(w)}{H(w_{\max})}.
 \end{aligned}
 \tag{5.3}$$

- *Imaging system simulation*: techniques adopted in photographic printing process are a source of inspiration to create TMOs. A possible approach is the application of a specific film characteristic curve intending to simulate its look. More complex systems can also be simulated considering film development and enlargement techniques used in photographic laboratory and they are a source of inspiration to create TMOs, for instance, see [9, 28, 33].
- *Human eye simulation*: the behavior of human vision is not the same for all illumination conditions, specially in very dim scenes or when there are abrupt illumination changes in time-varying scenes. The reality of visualized images is achieved only if human vision behavior is modeled. In [18], the histogram adjustment algorithm is extended in many ways to model human vision. Also in [32] this approach is explored.

In Fig. 5.1, the difference of image visualization depending on the chosen TMO is illustrated; after reconstructing the church HDR image, linear TMO was applied to visualize (b) and histogram adjustment was used to visualize (c). Comparison between TMOs is mainly perceptual, see [6]. In [20], perception-based image quality metrics are presented.

The tone mapping problem can be thought in image color domain as the problem of adjusting the range of image luminances to display luminances maintaining color chrominances. Thus, it is a quantization problem in radiance domain and TMOs available in the literature can be interpreted from this point of view. The uniform quantization algorithm is comparable to linear TMO, the populosity quantization is similar to histogram adjustment. Reasoning like that, a family of new TMOs derived from quantization algorithms can be proposed. The main difference between color quantization and tone mapping is that human eye is more sensitive to spatial changes in luminances than in color. This explains the focus of several TMOs in working on spatial domain, such as in human eye perception simulation and in dodging-and-burning [28].

Another branch of research in HDRI visualization is on HDR displays that are able to represent all tones encoded in an HDR file format [30].

CHAPTER 6

Tone Enhancement

In this chapter, we discuss the high dynamic range for tone enhancement and some applications to video processing.

The main question answered in this chapter is: *What can be done in terms of tonal enhancement of an image without knowledge about the camera characteristic f function and image radiance values?*

We start with a brief review on the recent HDR images research. Then the concept of relative tones is introduced and applied to obtain real-time tone enhanced video sequences.

6.1 PARTIAL RECONSTRUCTION: RELATIVE TONES

As we have seen, the camera characteristic f function is the tool that correlates radiance values to camera brightness values, and is necessary to recover absolute radiance values from captured images. Here we introduce the conceptual difference between the absolute and relative tone values.

We remind that image histograms are necessary and sufficient to recover the intensity mapping function τ , as shown in chapter 4.3.4, useful to reconstruct the radiance map. The image histogram comparison expresses the concept that the m brighter pixels in the first image will be the m brighter pixels in the second image for all m discrete values assumed by the image.

We observe that a simple summation of two images preserves the information present on the histogram of the original images. The sum operation potentially doubles the number of distinct tones in the resulting image, consequently it requires one bit more to be stored. In Fig. 6.1, we show an example of LDRI histograms, and the combined information present in the image sum.

We then pose the following question: *What can be recovered from the sum of the images, if one does not know neither the exposures nor the response curve?* Surprisingly, the answer is: many things!

Of particular interest to our discussion is the following theorem [11]:

[Simple Summation Theorem]

The sum of a set of images of a scene taken at different exposures includes all the information in the individual exposures.

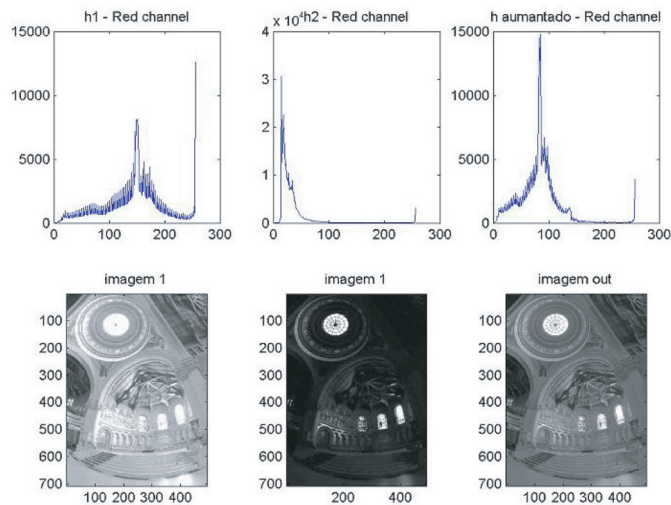


FIGURE 6.1: Example of two different exposed images with correspondent histograms, and the summation image with the combined histogram. To be visualized, the image sum was linearly tone mapped.

This means that, if one knows the exposure times used to obtain the images and the response curve of the camera, the radiance values and the individual images can be recovered from the image sum. In [11], the authors use this result to optimize the camera acquisition parameters.

We then define the *relative tones* m as the values present in the summation image, while *absolute tones* w are the real correspondent radiances. The relative m values are unique indices to real radiance values. Thus, with the response function f and the exposure camera parameters in hand a lookup table can be generated mapping q to w values, i.e., $F_{f,\Delta_r} : [0, 2] \rightarrow [E_{\min}, E_{\max}]$. In Fig. 6.2, we illustrate the relation between the quantization levels m and the

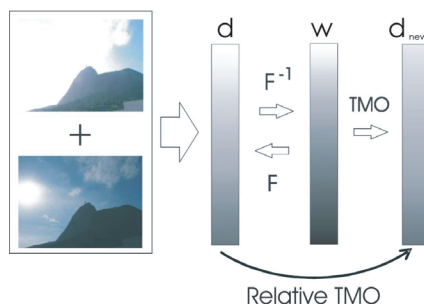


FIGURE 6.2: Absolute versus relative tone values.

absolute tone values w . We observe that, assuming that f is monotonically increasing, this mapping F is 1-1.

Absolute tones are directly related to physical quantities while relative tones preserve absolute order but do not preserve magnitude.

Usually, TMOs are described in terms of absolute tones. However, since there is a 1-1 mapping between the relative and absolute tone values, we conclude that TMOs to be applied directly to the relative tones can be proposed.

6.2 ACTIVE RANGE-ENHANCEMENT

When generating differently exposed images, instead of varying the exposure time, alternative ways to vary pixel exposure can be explored. By controlling light intensity, while keeping all other camera parameters fixed, the pixel irradiance w_{ij} is altered, and hence its correspondent exposure value. This is not usually done, once the illumination intensity varies irregularly throughout the scene. However, the concept of relative tone values introduced here makes possible the use of active illumination in recovering relative tones to enhance tonal resolution on the actively illuminated area.

Changing light intensity the tonal partial ordering on actively illuminated areas is preserved once reflectance is proportional to incident radiance. Thus, relative tones are naturally recovered by simple summation in those regions.

The fact that illumination mostly affects the foreground pixels can be explored to perform foreground–background segmentation.

The algorithm discussed in [31] can be used to produce a segmentation mask M , where 1 is assigned to the foreground pixels and 0 to the background, defining the actively illuminated regions.

Using active illumination, the foreground objects not only can be extracted but also can be tone-enhanced using relative tones concept. Note that the background cannot be tone-enhanced since there is no exposure variation in these regions. The summation image is given by $S(d_{ij}) = L(d_{ij}) + L(d_{ij})$, with $S(d_{ij}) \in [0, 2]$. As exposure values are different in two subsequent images, tonal resolution is enhanced by simple summation.

To visualize the foreground tones Larson's histogram adjustment is applied, $TMO_{s,v} : [0, 2] \rightarrow [0, 1]$, to the actively illuminated region. Chromaticities of both images are linearly combined to attenuate problems in over- or underexposed pixels.

The final image is obtained by a simple image composition:

$$C_{ij} = M_{ij}F_{ij} + (1 - M_{ij})B_{ij},$$

where F_{ij} are the tone-enhanced foreground pixels and B_{ij} are the background pixels that

remain unchanged. A low-pass filter is applied to image M to avoid the perception of discontinuities in transition of segmented regions.

All discussion above assumes that one knows the pixel correspondence in differently exposed images. This is trivial for static scenes captured by cameras mounted on a tripod. For moving scenes, pixel correspondence has to be solved before applying the method outlined above.

6.3 REAL-TIME TONE-ENHANCED VIDEO

For video implementation, a setup composed of a video camera synchronized with a projector is used. A video signal, where each field has constant gray color values ρ_1 and ρ_2 with $\rho_1 \neq \rho_2$ is projected onto the scene. This signal is connected to the camera *gen-lock* pin, which guarantees projection/capture synchronization. Each light exposure lasts for $1/59.54$ s using the NTSC standard. The fields I^1 and I^2 of each captured frame represent the same object illuminated differently. This is sufficient for actively segment and tone-enhance the image pair.

Input video images are in the Yuv color space. Thus, the processing is performed using the luminance defined as $L(p_{ij}) = Y(d_{ij})$, where Y is the video luminance channel. In this scheme, the tone-enhancement can be applied to any two consecutive fields. This produces an output video stream with the same input frame rate, as shown in Fig. 6.3.

We assume that the frame rate is high compared to the object's movement, thus, the effects of moving objects are small between a pair of fields. In the segmentation step, graph-cuts

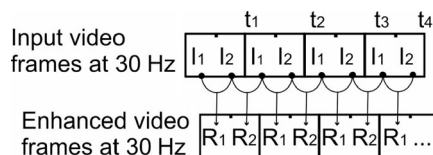


FIGURE 6.3: Two consecutive input fields result in one frame.



FIGURE 6.4: Images (a) and (b) are the video input fields, while in (c) is shown the tonal-enhanced foreground.



FIGURE 6.5: Images (a) and (b) are the video input fields, while in (c) is shown the tonal-enhanced foreground.

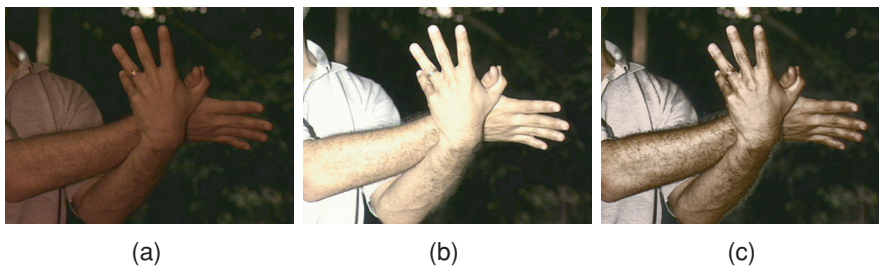


FIGURE 6.6: Images (a) and (b) are the video input fields, while in (c) is shown the tonal-enhanced foreground.

optimization is not applied, only the initial seed is used. To minimize undesirable effects of shadow regions, the projector is positioned very close to the camera. This implies that background should be far enough to not be affected by active illumination. The working volume can be determined by analyzing the projector intensity decay with distance (see chapter 4.3.4).

Figures 6.4, 6.5 and 6.6 show two consecutive video fields with different illumination, and their respective resulting tone-enhanced foreground. The projected gray values used were $\rho_1 = 0.35$ and $\rho_2 = 1$. The luminance threshold L_{\min} used is 0.08. One can note that our method has some trouble segmenting low-reflectance objects, such as hair. However, the resulting tone-enhanced images are still quite satisfactory.

A home-made version of the system was also implemented, it is composed of a web-cam synchronized with a CRT monitor playing the role of active illuminant [29]. Some results are available at <http://www.impa.br/~asla/ahdr>.

References

- [1] A. Adams, *The Camera, The Ansel Adams Photography Series*. Little, Brown and Company, Bulfinch Press UK—London, UK, 1980.
- [2] ——— *The Negative, The Ansel Adams Photography Series*. Little, Brown and Company, Bulfinch Press UK—London, UK, 1981.
- [3] ——— *The Print, The Ansel Adams Photography Series*. Little, Brown and Company, Bulfinch Press UK—London, UK, 1983.
- [4] P. Debevec and J. Malik, “Recovering High Dynamic Range Radiance Maps From Photographs,” *Proc. ACM SIGGRAPH ’97* (1997), pp. 369–378.
- [5] K. Devlin, A. Chalmers, A. Wilkie and W. Purgathofer, “Star: Tone Reproduction and Physically Based Spectral Rendering,” *State of the Art Reports, Eurographics 2002* (September 2002), pp. 101–123.
- [6] F. Drago, W. Martens, K. Myszkowski and H. Seidel, “Perceptual Evaluation of Tone Mapping Operators With Regard to Similarity and Preference,” *Tech. Report MPI-I-2202-4-002* (2002).
- [7] M. Goesele, W. Heidrich and H. Seidel, “Color Calibrated High Dynamic Range Imaging With ICC Profiles,” *Proc. 9th Color Imaging Conference Color Science and Engineering: Systems, Technologies, Applications, Scottsdale* (November 2001), pp. 286–290.
- [8] A. Glassner, *Principles of Digital Image Synthesis*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1994.
- [9] J. Geigel and F. Musgrave, “A Model for Simulating the Photographic Development Process on Digital Images,” *Comp. Graph. Proc.* (2003), pp. 135–142.
- [10] M. Grossberg and S. Nayar, “Determining the Camera Response From Images: What is Knowable?” *IEEE Trans. PAMI* 25, 11 (November 2003), 1455–1467.
- [11] A. O. Akyüz and E. Reinhard, “Noise Reduction in High Dynamic Range Imaging,” Academic Press—Orlando, FL, USA, 18 (May 2007), pp. 366–376.
- [12] S. K. Nayar, “Modeling the Space of Camera Response Functions,” *IEEE Trans. PAMI* 26, 10 (October 2004), 1272–1282.
- [13] M. Goesele, *New Acquisition Techniques for Real Objects and Light Sources in Computer Graphics*. Verlag, Saarbrücken, Germany, 2004.

- [14] J. Gomes and L. Velho, *Image Processing for Computer Graphics*. Springer-Verlag, New York—Secaucus, NJ, USA, 1997.
- [15] HDRSHOP, <http://gl.ict.usc.edu/HDRShop/>.
- [16] G. W. Larson, “High Dynamic Range Image Encodings,” http://www.anywhere.com/gward/hdrenc/hdr_encodings.html.
- [17] D. Litwiller, “CCD vs. CMOS: Facts and Fiction,” <http://www.dalsa.com>.
- [18] G. Larson, H. Rushmeier and C. Piatko, “A Visibility Matching Tone Reproduction Operator for High Dynamic Range Scenes,” *IEEE Trans. on Vis. and Comp. Graph.* 3, 4 (1997), 291–306. [10.1109/2945.646233](https://doi.org/10.1109/2945.646233)
- [19] Macbeth, “Gretag Macbeth,” <http://www.gretagmacbeth.com>.
- [20] A. MacNamara, “Visual Perception in Realistic Image Synthesis,” *Comput. Graph. Forum* 20, 4 (2001), 221–224.
- [21] T. Mitsunaga and S. Nayar, “Radiometric Self Calibration,” *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)* (June 1999), Vol. 1, pp. 374–380.
- [22] K. Matkovic, L. Neumann and W. Purgathofer, *A Survey of Tone Mapping Techniques*. Tech. Rep. TR-186-2-97-12, Institute of Computer Graphics and Algorithms, Vienna University of Technology, Favoritenstrasse 9-11/186, A-1040 Vienna, Austria, April 1997. human contact: technical-report@cg.tuwien.ac.at.
- [23] S. Mann and R. W. Picard, “Being Undigital With Digital Cameras: Extending Dynamic Range by Combining Differently Exposed Pictures,” in *Proceedings of the 48th IS&T’s Annual Conference*, (May 1995), pp. 442\226448, Washington, DC, USA.
- [24] PFStools. <http://www.mpi-inf.mpg.de/resources/pfstools/>.
- [25] Photosphere. <http://www.anywhere.com/>.
- [26] RASCAL. <http://www1.cs.columbia.edu/CAVE/software/rascal/rrhome.php>.
- [27] M. Robertson, S. Borman and R. Stevenson, “Dynamic Range Improvement Through Multiple Exposures,” *Proc. IEEE Int. Conf. on Image Processing* (Kobe, Japan, October 1999), Vol. 3, IEEE, pp. 159–163.
- [28] E. Reinhard, M. Stark, P. Shirley and J. Ferwerda, “Photographic Tone Reproduction for Digital Images,” *Proc. ACM SIGGRAPH ’02* (2002).
- [29] A. Sa, M. Bernardes, A. Montenegro, P. C. Carvalho and L. Velho, “Actively Illuminated Objects Using Graph-cuts,” *Proc. of SIBGRAPI 2006—XIX Brazilian Symposium on Computer Graphics and Image Processing* (Manaus, October 2006), SBC—Sociedade Brasileira de Computacao, IEEE Press.
- [30] H. Seetzen, W. Heidrich, W. Stuerzlinger, G. Ward, L. Whitehead, M. Trentacoste, A. Ghosh and A. Vorozcovs, “High Dynamic Range Display Systems,” *Proc. ACM SIGGRAPH ’04* (2004).

- [31] A. Sá, M. Vieira, P. Carvalho and L. Velho, “Range-Enhanced Active Foreground Extraction,” *Proc. ICIP* (2005).
- [32] J. Tumblin, J. Hodgins and B. Guenter, “Two Methods for Display of High Contrast Images,” *ACM Trans. Graph.* 18, 1 (1999), 56–94. [10.1145/300776.300783](https://doi.org/10.1145/300776.300783)
- [33] J. Tumblin and H. Rushmeier, “Tone Reproduction for Realistic Images,” *IEEE Comput. Graph. Appl.* 13, 6 (1993), 42–48. [10.1109/38.252554](https://doi.org/10.1109/38.252554)
- [34] G. Ward, “The radiance lighting simulation and rendering system,” *Proc. ACM SIGGRAPH '04* (1994), pp. 459–472.
- [35] ——— “Fast, robust image registration for compositing high dynamic range photographs from hand-held exposures,” *J. Graph. Tools* 8, 2 (2003), 17–30.

Author Biography

Luiz Velho is a Full Researcher/Professor at IMPA—Instituto de Matematica Pura e Aplicada of CNPq , and the leading scientist of VISGRAF Laboratory. He received a BE in Industrial Design from ESDI/UERJ in 1979, a MS in Computer Graphics from the MIT/Media Lab in 1985, and a Ph.D. in Computer Science in 1994 from the University of Toronto under the Graphics and Vision groups.

His experience in computer graphics spans the fields of modeling, rendering, imaging and animation. During 1982 he was a visiting researcher at the National Film Board of Canada. From 1985 to 1987 he was a Systems Engineer at the Fantastic Animation Machine in New York, where he developed the company's 3D visualization system. From 1987 to 1991 he was a Principal Engineer at Globo TV Network in Brazil. In 1994 he was a visiting professor at the Courant Institute of Mathematical Sciences of New York University. He also was a visiting scientist at the HP Laboratories in 1995 and at Microsoft Research China in 2002.

