Development of High Speed High Dynamic Range Videography

By: David John Griffiths

Dissertation submitted to the faculty of the Virginia Polytechnic Institute and State University in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
In
Mechanical Engineering

Alfred L. Wicks, Chair
Lynn A. Abbott
Michael J. Roan
Brian Vick
Jianhua Xuan

December 8, 2016
Blacksburg, VA

Keywords: high dynamic range video, high dynamic range image, high speed video, camera error analysis, synthetic images, stochastic image combination

Development of High Speed High Dynamic Range Videography

David J. Griffiths

ABSTRACT

High speed video has been a significant tool for unraveling the quantitative and qualitative assessment of phenomena that is too fast to readily observe. It was first used in 1852 by William Henry Fox Talbot to settle a dispute with reference to the synchronous position of a horse's hooves while galloping. Since that time private industry, government, and enthusiasts have been measuring dynamic scenarios with high speed video. One challenge that faces the high speed video community is the dynamic range of the sensors. The dynamic range of the sensor is constrained to the bit depth of the analog to digital converter, the deep well capacity of the sensor site, and baseline noise. A typical high speed camera can span a 60 dB dynamic range, 1000:1, natively. More recently the dynamic range has been extended to about 80 dB utilizing different pixel acquisition methods.

In this dissertation a method to extend the dynamic range will be presented and demonstrated to extend the dynamic range of a high speed camera system to over 170 dB, about 31,000,000:1. The proposed formation methodology is adaptable to any camera combination, and almost any needed dynamic range. The dramatic increase in the dynamic range is made possible through an adaptation of the current high dynamic range image formation methodologies. Due to the high cost of a high speed camera, a minimum number of cameras are desired to form a high dynamic range high speed video system. With a reduced number of cameras spanning a significant range, the errors on the formation process compound significantly relative to a normal high dynamic range image. The increase in uncertainty is created from the lack of relevant correlated information for final image formation, necessitating the development of a new formation methodology.

In the proceeding text the problem statement and background information will be reviewed in depth. The development of a new weighting function, stochastic image formation process, tone map methodology, and optimized multi camera design will be presented. The proposed methodologies' effectiveness will be compared to current methods throughout the text and a final demonstration will be presented. Observations will be discussed indicating some of the strengths and weaknesses of the proposed solution.

Development of High Speed High Dynamic Range Videography

David J. Griffiths

GENERAL ABSTRACT

High speed video is a tool that has been developed to capture events that occur faster than a human can observe. The use and prevalence of high speed video is rapidly expanding as cost drops and ease of use increases. It is currently used in private and government industries for quality control, manufacturing, test evaluation, and the entertainment industry in movie making and sporting events.

Due to the specific hardware requirements when capturing high speed video, the dynamic range, the ratio of the brightest measurement to the darkest measurement the camera can acquire, is limited. The dynamic range limitation can be seen in a video as either a white or black region with no discernible detail when there should be. This is referred to as regions of over saturation or under saturation.

Presented in this document is a new method to capture high speed video utilizing multiple commercially available high speed cameras. An optimized camera layout is presented and a mathematical algorithm is developed for the formation of a video that will never be over or under saturated using a minimum number of cameras. This was done to reduce the overall cost and complexity of the setup while retaining an accurate image. The concept is demonstrated with several examples of both controlled tests and explosive tests filmed up to 3,300 times faster than a standard video, with a dynamic range spanning over 310,000 times the capabilities of a standard high speed camera.

The technology developed in this document can be used in the previously mentioned industries whenever the content being filmed over saturates the imager. It has been developed so it can be scalable in order to capture extremely large dynamic range scenes, cost efficient to broaden applicability, and accurate to allow for a fragment free final image.

Acknowledgments

The author would like to recognize some of the many individuals that helped throughout the creation of this document. Without their help, much of what will be described would not be possible.

First the author would like to thank the personnel at the Naval Surface Warfare Center Dahlgren Division for their support in the author's pursuit of inventing the high speed high dynamic range video technology. The enabler of this work was a grant provided by the Naval Surface Warfare Center Dahlgren Division's In house Laboratory Independent Research Program, which the author is extremely grateful for. The author would also like to thank the personnel at the Explosive Ordinance Area for their cooperation and professionalism in the collection of sample data. The author would also like to thank the personnel from the Potomac River Test Range for going well above and beyond the normal expectations to aid in setup, testing, and procurement of equipment to create and demonstrate the prototype system. The author would like to thank Mr. Blaise Corbette for his continual alacrity when presented with new ideas and Mr. Jon Strzelec for his aid and fervent support of the prototype system.

The author would like to acknowledge the continual support of his family, his sister Laura, parents, John and Pat, and his ever encouraging fiancée, Allison. With their never ending ardor, seemingly arduous tasks became manageable.

Finally the author would like to thank his academic advisor, Dr. Al Wicks, for taking a chance on a distance student, working full time while pursuing an advanced degree. Without that opportunity and the continual advisement, the work presented herein would not have transpired. Our conversations, and your advisement have been sincerely appreciated.

Table of Contents

For proper viewing of all figures in this document please use a monitor or printer that is calibrated to a gamma of 2.2, D65 white point.

List of Figures

List of Tables

List of Abbreviations

HDR	High Dynamic Range
EMM	Error Minimization Method
SNR	Signal to Noise Ratio
fps	Frames per second
pps	Pictures per second
CDS	Correlated double sampler
dB	Decibels
CMOS	Complementary metal oxide sensor
ADU	Analog Digital Unit
A/D	Analog to digital converter
HSL	Hue, Saturation, Lightness
CCD	Charged coupled device

Document Nomenclature

DR	Dynamic Range
max[·]	Maximum of argument
min[·]	Minimum of argument
E_λ	Energy from electromagnetic wave
ϕ_e	Radiant flux
h	Plank's constant
c	Speed of light
λ	Wavelength
Δt	Exposure time
ω	Solid angle
A	Pixel Area
L_e	Radaince
$L(\lambda)$	Spectral radiance
$V(\lambda)$	Spectral response curve
Y	Luminance
Z	Digital value
i	Pixel number index
j	Camera or image index
A_{phy}	Physical attenuation
ϕ	Radiant Exitance
ϕ_{max}	Maximum radiant exitance
ϕ_{HDR}	HDR exitance estimate
N_{well}	Deep well capacity
q_s	Quantization steps in A/D
q	Electrons per Coulomb
dc	Dark current
I_d	Current at 300K
E_g	Bandgap energy
T	Sensor temeprature
γ	Mean number of photons
e^-_{tot}	Charge accumulation at sensor site
$s(\psi_s)$	Charge accumulation from shot noise
st	stops
N_D	Neutral density filter value
R,T_r	Beam splitter reflectance or transmittance
σ_{ADU}	Measured noise
σ_{sw}	World shot noise
p	Points in exitance range
ζ	Base dynamic range
k	Iteration number
σ_{ss}	Shot noise as sensor site
n	Number of samples
ξ	Number of standard deviations
P(·)	Probability of an event
$\bar{\phi}$	Actual mean radiant exitance
c_i	Normal standard units
ϵ	Offset from mean
σ_F	Error in formation
γ_{HDR}	Estimated luminance
γ_i	Reference luminance
k	Stefan Boltzman constant
E	Irradaince
M	Luminous exitance
$f^{-1}(Z)$	Camera response function
$g(Z)$	Log_e of the camera response function
Log	Log base 10
Ln	Log base e
$w(Z)$	Weighting function of digital value Z
α	Smoothing term
J	Total number of cameras or images
σ_T	Total error
σ_s	Shot noise error
σ_{rst}	Reset noise error
σ_{dc}	Dark current noise error
σ_{rd}	Read noise error
σ_q	Quantization noise error
φ	Photon flux
f/#	f-number of the lens
η	Quantum efficiency
S	Charge accumulation
s_1	Distance to object
s_2	Distance to imager
τ_{optics}	Optical transmission
h_{sys}	Point spread function of system
L_{scene}	Scene spectral radiance
Δf	Bandwidth
R	Resistance
$\overline{\mu_{dc}}$	Mean dark current offset
μ_{DC}	Dark current noise (ADU)
μ_{rd}	Read noise (ADU)
δ	Percent of ADU noise
randn[·]	Normaly distributed random value
ME	Mean error
D	Pupil diameter
D_{max}	Maximum pupil diamter
Y_b	Background luminance
a	Area field of view
e	Number of eyes
y	Simulated age
y_0	Reference age, 28.58
E_r	Retinal irradiance
E_p	Pupil irradiance
R/R_{max}	Normilized cellular response
σ_L	Light adapted pupil response
σ_D	Half saturation point
σ_x	Cellular adaptation
σ_β	Inverse pigment bleaching state
o_v	Vertical offset
o_h	Horizontal offset
α	Sigmoidal amplitude
C	Photo dioded capacitance
g	Sensor gain

Chapter 1. Introduction

Every sensor that is designed to transform an external stimulus into a digitally representative value is constrained to an operational range. The total operational range is limited by either nonlinear sensor response, electrical saturation, thermal noise, or a physical damage threshold. The operational range is one of the primary tradeoffs considered when recording test events. Comparable sensors with differing dynamic ranges will typically have a sensitivity and error proportional to the resolvable range. A tradeoff must be made to either acquire high fidelity data at the risk of exceeding the operational range of the instrument, or collecting low fidelity data and ensure the capture of the information.

Photographers and videographers have been considering this tradeoff since the first camera was created. The film sensitivity, exposure time, and aperture must all be defined to allow the film to absorb enough energy to produce an adequate image of the subject matter being viewed. There are instances where the idealized subject matter acquisition comes at the expense of other portions of the image being improperly exposed. Should the scene be temporally varying, as in the case of video, the camera must adapt to the scene to acquire the most information possible, or the highest fidelity image of a particular subject. The continuous adaptation frequently leads to the loss of information on the extremes of the camera's resolvable range. This tradeoff can be seen every day in broadcast TV, standard point and shoot cameras, cell phone cameras and similar acquisition devices where parts of the video stream are either over or under saturated.

During the mid-1990's, the concept of combining the sensor output from multiple digital cameras into a single image was introduced to overcome the dynamic range limitation of a single camera sensor. Since that time, more software developers have been supporting this form of image creation with programs such as Adobe Photoshop. As of 2016, virtually all smart phones, tablets, and cameras have some level of high dynamic range image capture capability. In 2014, the Apple iPhone 6 was the first smart phone to introduce native HDR video capture, demonstrating the continuing demand for this kind of technology. Amazon.com Inc., Vudu.com, and Netflix Inc. are some of the first content providers of HDR video for consumer access. In 2016 the first series of high dynamic range blu-ray movies and televisions have also entered the consumer market. Two competing encoding technologies are currently in existence, Dolby Vision and HDR10. Dolby Vision can cover the greatest dynamic range, a minimum output of 0.0001 nits and a maximum of 10,000 nits, lending to a 160 dB dynamic range [1]. A camera

would have to have a 27 bit analog to digital converter with a deep well capacity of at least one million electrons to be able to acquire the full range available on a single sensor for this encoding method to be fully utilized. Such a camera does not currently exist, which is the motivation for a multi imager HDR video system.

Objective

The objective of this effort is to create a methodology that fuses the input of several spatially aligned optical sensors with varying amounts of attenuation into a finalized image that can be further modified for accurate display, where no portion of the image is either over or under exposed. This effort will focus on the video capture of extremely large dynamic range scenes, in excess of 100 dB, with as few optical sensors as possible. This becomes a challenging task as the current state of the art for high dynamic range image formation depends on a significant number of samples, typically greater than 5, to statistically combine and correct the images. The techniques presented herein are formulated to overcome the dependency on many images. The objective is to form a robust methodology to reduce the number of required images needed for reconstruction without image degradation. With a reduction in the number of required sensors while maintaining fidelity, the extension of the high dynamic imagery to high speed high dynamic range videography is a natural and plausible merger. The unique contributions and underlying hypotheses introduced in this work are:

1) Development of a simplified weighting function that allows for the rapid creation of high dynamic range video frames while minimizing noise.
   Hypothesis: A weighting function can be created based solely on knowledge of camera parameters to minimize the error on the high dynamic range image formation.
2) The creation of a video frame formation methodology, named Error Minimization Method, which can be used to improve the final frame formation even in the presence of image misalignment and flare degradation.
   Hypothesis: It is possible to reduce video frame formation uncertainty by selecting the least degraded images only for final formation with limited data samples and a prior knowledge of the imaging chain.
3) The creation of a tone mapping method to scale the high dynamic range information into a perceptually correct image for display.
   Hypothesis: The human vision system can be modeled to create a tone mapping method that makes a video of high dynamic range video frames appear perceptually correct.
4) The development of a noise optimized high speed high dynamic range video system utilizing the methods from contributions 1-3 with as few sensors as needed.
   Hypothesis: A reduced number of cameras, three, can be used to span greater than a 160 dB range.

Fundamental concepts review

There is a significant amount of photometric and radiometric terminology used throughout this document, therefore a quick summary of the terminology as it is used in this document shall be given along with diagrammatic representation for better understanding. Also there are significant references to the internal workings and vocabulary associated with the use of a camera, therefore a brief review of digital cameras will also be given.

Digital camera fundamentals

The objective of a camera is to focus light from a scene on a planer surface in an undistorted way. Polychromatic light enters a typically glass lens, which has the objective of focusing light on the sensor. The polychromatic light will converge differently due to the non-homogeneous permittivity of glass elements and air gaps. This leads to chromic aberrations seen in images as rainbow like halos around high frequency content. Another possible image degradation is vignetting, which causes a dimming of the image around the corners and is commonly observed when a large aperture with a small focal length is used. Most modern lenses are designed to reduce these artifacts, but they can still be observed under the right circumstances.

Light is further attenuated by the aperture. Light will pass through the aperture and it will act as a filter for light that is not collimated. The f number is defined as the ratio of the focal length and the diameter of the aperture opening. A small f number corresponds to a large opening relative to the focal length of the lens. This has the effect of altering the depth of field.

The depth of field is the range of distances from the optical axis that can be in focus simultaneously. The background and foreground blurring effect is typically referred to as the bokeh of the lens and can be modified by the geometry of the aperture. Reducing the diameter of the aperture, decreases the amount of uncollimated light allowed to pass and provides for a greater depth of field. This concept is what allows pinhole cameras to work. As the effective aperture gets smaller, a lens is not needed as the depth of field approaches infinity. However, the amount of light that is allowed to pass is significantly reduced, therefore very long exposures are required in order to form a sharp low noise image. Once the light passes through the lens the light gets further filtered, except in 3 CCD or Foveon X3 sensors, by a Bayer filter.

A Bayer filter, originally developed by Bryce Bayer and patented in 1976 [2], is a red, green, and blue optical filter overlaid on a charge coupled device (CCD) or an active pixel sensor

charged metal oxide sensor (APS CMOS). This filter consists of 50% green, 25% red, and 25% blue pixels per unit area. The filter was designed with a green passband emphasis to mimic the human vision system's luminance perception. The Bayer filter has the effect of attenuating the light further since only 33% of the information from the scene is available to each pixel site. The full RGB output is estimated using a demosaic algorithm, interpolating the remaining 66% based on the neighboring pixels, which leads to color artifacts, spatial blurring, and noise particularly in regions of high spatial frequencies.

Photons are converted into electrical charges that accumulate in charge storage wells. The quantum efficiency relates the photon flux to the electron flux in the well at a given wavelength. The electron accumulation is linearly related to the quanta of light incident on each pixel. Each well has a specific capacity and when nearing or at saturation, can bleed into the adjacent wells causing a blooming effect. Most modern cameras are designed to have a potential drain to bleed excess charge to and reduce this artifact (see Appendix) The larger the drain the smaller the fill factor of the pixel, and reducing the cameras overall sensitivity. After the photosensitive capacitors have been exposed for the desired time, the charged coupled device (CCD), shifts the accumulated charges into the adjacent cells until they reach the last cell in the array.

There are several forms of hardware error that arise in a CCD or CMOS type sensors [3] [4]. For the following discussion, a CCD sensor will be assumed since they are prevalent in machine vision applications. A similar discussion can be performed for a CMOS based active pixel sensor, but for a general overview there are many similarities. In general, a CCD based sensor is constructed of a silicon dioxide layer on a silicon wafer with a gate structure built over the oxide layer. When a photon impacts the silicone layer, an electron hole pair is generated and the freed electrons will be collected within one of the gate sites. This is done by applying a positive electrical potential to the gate.

At each pixel site the accumulated charge per integration time is shift registered to the adjacent site by alternating the potential of each gate within a column of the CCD. The columns of charge are moved in parallel to a serial register in series. The charge is amplified and digitized through an analog to digital converter. Throughout this process there are several sources of error that can arise, most of which are shared with CMOS type sensors.

The first source of noise comes from dark current noise. Dark current noise is independent of the incident light and dependent on the temperature of the sensor. This source of noise is commonly seen in low light scenes and can account for several analog to digital (A/D) counts.

Another source of noise, called shot noise, stems from the quantum nature of light [5]. Light emission from an object is not a continuous stream. On a small enough time scale the random emission of energy packets can be observed and manifests itself as another source of noise. Shot noise is dependent on the mean incident energy level. The next form of noise is read noise. Read noise stems from thermal noise and outside interference and is the noise associated with the sensors amplifying system. Fixed pattern noise manifests itself by spatial sensitivity variations from the manufacturing tolerance of the pixels. Each well has a different capacity and a different sensitivity. Fixed pattern noise can be seen on long exposures; an example can be observed in Figure 1-5 in the HDR reconstruction image as a small spike offset from the main illumination point.

One final form of noise from the camera system is introduced from the analog to digital converter. The information that is being measured must be truncated and can be treated as a uniformly distributed random variable. Once the voltages have be converted to digital information the cameras apply several geometric and colorimetric transforms before the data gets to the finalized form such as a lens correction, dynamic range compression, color gamut matching, denoising, white balance, image compression, and more depending on the camera's manufacturers settings.

Radiometric and photometric review

First the concept of dynamic range shall be introduced. The dynamic range, DR, of a given measurement device is defined by

$$DR = 20Log_{10}\left(\frac{\max[\cdot]}{\min[\cdot]}\right) \qquad 1\text{-}1$$

where max[·] is the maximum resolvable quantity and min[·] indicates the minimum resolvable quantity. The dynamic range is an indication of the ratio of the largest resolvable quantity to the lowest resolvable quantity in a particular system. An easily recognizable example of dynamic range includes sound pressure levels, such as the ratio of the average sound pressure level of a rock concert [6], 20 Pa, to minimum threshold of human hearing, about 20 μPa, which is a dynamic range of 120 dB. In visual displays the dynamic range can be defined by the ratio of the maximum output brightness to the minimum output brightness and can be expressed in units of stops. Stops are used in the photography profession as it is indicative of f stop numbers a given scene can range. The number of stops a particular scene can span is defined by

$$DR = Log_2\left(\frac{\max[\cdot]}{\min[\cdot]}\right) \quad 1\text{-}2$$

Therefore the dynamic range of the rock concert, should the sound pressure level be compared in stops of dynamic range, would encompass about a 20 stop range.

There are several radiometric terms that are used throughout this text such as radiance, radiant exitance, spectral radiance, irradiance, and spectral irradiance. The terms that have the prefix *irr* indicate light, thought of as a ray, arriving to a surface. A term without the prefix, such as radiant exitance, indicate that light is emanating or reflecting from the surface. When the word spectral is included in the term, it is an indicator of per unit wavelength. Without the word spectral, the total power of all wavelengths are integrated. All terms are diagrammatically shown in Table 1-1 with their associated units and document nomenclature.

The radiant flux is the total amount of energy in a given time [7]. The energy being referred to stems from the energy associated with the electromagnetic wave and can be described by

$$E_\lambda = \frac{hc}{\lambda} \quad 1\text{-}3$$

where h is Plank's constant, c is the speed of light, and λ is the wavelength of light. Equation 1-3 is expressed in units of Joules and is representative of the energy associated with a given wavelength of light. The radiant flux in units of Watts can be expressed as

$$\phi_e = \frac{hc}{\lambda \Delta t} \quad 1\text{-}4$$

where Δt is the time. Therefore, the radiance, L_e, can be defined as

$$L_e = \frac{d\phi_e}{d\omega dA_\perp} \quad 1\text{-}5$$

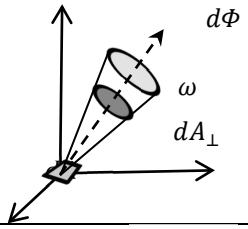
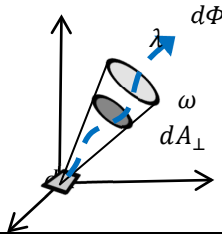
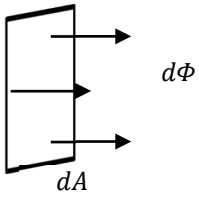
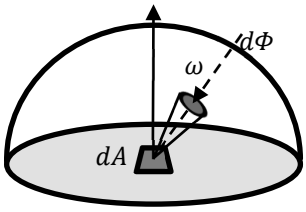
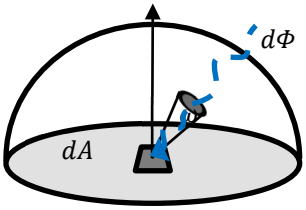
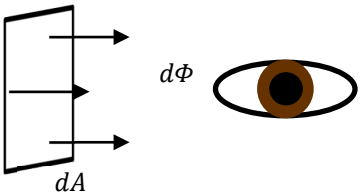
from a given projected differential area, $dA_\perp$, with a given solid angle, dω. Radiance can be described as the flux density per unit area per steradian. The spectral radiance is then the radiant flux density per unit area, per unit steradian, per unit wavelength. The radiant exitance is defined as the area density of radiant flux leaving a unit area. Irradiance is the total energy from all angles incident on a unit area. Spectral irradiance is then the energy per unit wavelength incident on a unit area.

The perception of the intensity of light is considered the photometric response [8]. It can be thought of as the radiance modified by the human vision system's ability to perceive the different wavelengths. The luminance, Y, from a scene can be defined as

$$Y = \int_{\lambda} L(\lambda)V(\lambda)\,d\lambda \qquad 1\text{-}6$$

where V(λ) is the spectral response curve and L(λ)is the spectral radiance. The units associated with the measurement of luminance is either presented in candelas per meter squared, a nit, or lumens per meter squared per steradian.

Table 1-1. Pictorial descriptions of selected radiometric and photometric terms used in this document.

Term	Diagram	Units
Radiance		$L_e = \frac{W}{m^2 sr}$
Spectral Radiance		$L(\lambda) = \frac{W}{m^2 sr\ nm}$
Radiant Exitance		$\phi = \frac{W}{m^2}$
Irradiance		$E = \frac{W}{m^2}$
Spectral Irradiance		$E(\lambda) = \frac{W}{m^2 nm}$
Luminous Exitance		$M = \frac{cd}{m^2}$

Genesis of high speed photography and videography

It is important to understand the origin of photography and videography to understand where high speed high dynamic range video falls into place. The American Society of Motion Picture and Television Engineers [9], founded in 1916, has suggested that any photograph of one millisecond or less is considered a high speed photography, and any video captured at a rate at or greater than 250 frames per second is considered high speed video. Typical video is recorded at an integer multiple of either 24 or 30 frames per second (fps). These baseline frame rates stem from the human persistence of vision, the discovery of the critical flicker frequency, and the addition of sound to video film.

One of the first scientific studies was performed by a French scientist, Chevalier d'Arcy [10], in 1765. His experiment consisted of whirling a hot ember in a circular motion at increasing rates until the circle looked to be continuous. At this point it was concluded the persistence of vision was about 0.13 seconds. Later in 1827 a Belgian physicist, Joseph Antoine Ferdinand Plateau, began to experiment further with the persistence of vision, published in his dissertation [11] in 1829. By 1832 Joseph Plateau created the first stroboscope, a device he called a phenakistoscope, which allowed a user to look through a window at a rotating disc with varying images. If the disk moved fast enough, it appeared to have a smooth animation. A projection of light through this device would later lead to modern cinema. By 1834, a British scientist, William Henry Fox Talbot became interested in the critical flicker rate and published what is now known as the Talbot-Plateau Law [12]. The law states that if there is a periodic stimulation on the retina at a high enough rate the event will appear as though it was a continuous event. The rate is dependent on the level of illumination and for average lighting scenarios is about 16Hz, the rate at which the first movies were replayed.

The concept of the critical flicker factor was being developed despite there not being any medium in which to really take advantage of this new found phenomena. One of the first patents that resembles a modern camera system was filed by Louis DuCos in 1864 [13], however a working prototype was never fully realized due to the lack of an adequate film. The first working film camera was realized in 1867 by Humbert de Morlad using a stationary lens.

Most original photographic captures originated from exposing silver salts to lights. The genesis of capturing a light began with four primary processes; daguerreotype, calotype, the wet collodion plate and twenty years later, the dry plate method. The daguerreotype exposure was introduced in 1839, invented by Louis Jaques Mande Daguerre. The process involved the

polishing of a silver plated copper substrate then sensitizing it by exposure to iodine and bromine fumes. The plate was exposed to the scene and developed using mercury fumes. The image was then fixed by washing the plate in a mixture of sodium thiosulphate. The fixed image was gilded with a heated solution of gold chloride and sodium thiosulphate.

The calotype was first introduced by William Henry Fox Talbot in 1841. The calotype film was a paper that was treated in silver halide, and when exposed to light would get darker. The original process took a very long time to process. Talbot switched to using silver iodide. The paper was exposed to light and blotted with a solution consisting of silver nitrate, acetic acid and gallic acid while warming the paper. Once the desired image was obtained the image was stabilized using a solution of potassium bromide and fixed in a solution of sodium thiosulphate. The main advantage of the calotype over the daguerreotype was the transparent nature which allowed for copies of the original.

The wet plate collodion process was introduced in 1851 attributed to an Englishman inventor, Frederick Scott Archer, published in March edition of *The Chemist,* and never patented. The process involved the addition of a soluble iodide to a cellulouse nitrate and coating a plate with the mixture. The plate was then dipped in a silver nitrate solution. The plate was placed into the camera, exposed and then rinsed with a solution of pyrogallic acid. The image was fixed with a solution of sodium thiosulfate.

The dry collodion process replaced the wet collodion process in 1871, by Richard Leach Maddox. With the dry collodion process a gelatin replaced the aqueous solution that the plate was dipped in. This allowed for the preparation of the film before it was needed and the storage of the film after it was exposed. George Eastman at the time a bank clerk, became interested in photography and the film making process. In 1879 Eastman moved to England and was awarded his first patent. Subsequently, he began the Eastman Dry Plate Company in 1881 to make dry plates commercial. In 1888 the Kodak company was founded and the first camera developed was the No 1. Kodak. It used a system that had the dry plates loaded in the camera when it was purchased. The paper backing on the film needed to be peeled, exposed then sent back for developing, procedurally similar in developing to the disposable cameras of the 1990's era. By 1902 Kodak had developed a nitrocellulose film and captured 90% of the world's use of film. This new form of film allowed for a new kind of camera to be developed.

As the film and technology advanced, the demand for photographs increased. The use of photography for qualitative or quantitative assessment is first attributed to William Henry Fox

Talbot in 1852 [14]. In this study Talbot, knowing films needed long exposure durations, made the leap to illuminate the subject with a high intensity light burst. The high light output was produced by a spark created with a Leydan jar. To demonstrate this, Talbot imaged a page of *The Times* newspaper while it was rotating at a high speed. When the film was developed the text could be read. The next reported used of a high speed photography was in 1858 by a photographer named Thomas Skaife. Skaife took photographs of cannon balls in mid-flight using wet collodion plates and the newly patented Pistolgraph camera was created. The Pistolgraph had a shutter mechanism that would snap shut once exposed. In 1860 [15] Sir John Herschel first wrote about the term "snap shot" as being associated with a photograph, and the term snapshot when referencing a picture is believed to stem from this.

One of the first ballistic uses of high speed photography was performed by Ernst Mach in the 1887 where he captured images of the shock wave produced by bullets [16]. This was accomplished by placing a candle just below the flight path of the bullet. As the bullet passed by the candle, the ionized gases got pushed into a switch gap and allowed the current to flow and trigger the camera. In 1892 Charles Boys of England used spark photography to study ballistics as well. He dropped the use of schlieren photography in favor of shadowgraphs for his investigations. In 1908 Arthur Mason Worthington [17] also utilized the spark illumination method to study the phenomena of fluid splashes; an example photograph from the text can be seen in Figure 1-1.

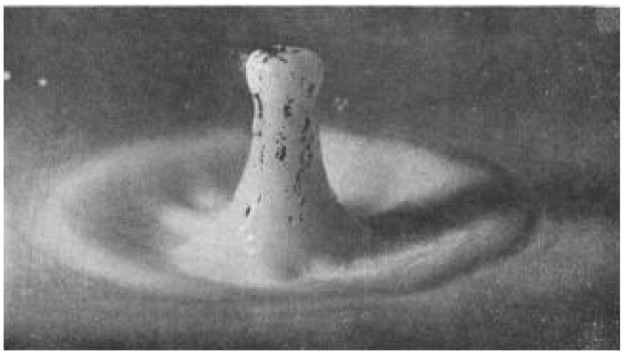


Figure 1-1. Splash picture produced from a spark gap illumination. Source. A Study of Splashes [17]

In 1912 Joseph Cranz [18], an Austrian, also used spark photography to measure the mode shapes of a gun barrel and rifle recoil. The method of spark illumination for instantaneous analysis of physical phenomena had come to maturity, being used more frequently in scientific analysis. The development of high speed video did not begin until 1872.

In 1872 an aspiring photographer born in Thames, England immigrated to the United States, originally named Edward Muggeridge, changed to Eadweard Muybridge on his arrival to America. Muybridge in 1872, was asked by the Governor of California to help resolve a dispute on the position of a horse's hooves while in a gallop. The first attempts used too slow of a wet collodion process and the images were not satisfactory. In 1873 [19] sharper views were achieved and the bet settled, as all four hooves could be seen in midair synchronously. More importantly a system was designed to take a series of images in rapid succession that could later be replayed as an animation, hence the first high speed video was captured.

By 1877 Muybridge had developed a system that had an exposure time of one millisecond. In 1881 a Frenchman, Étienne-Jules Marey, made a variation of the photographic gun. This was a device that looked like a rifle, similar to the Pistolgraph, where the barrel contained the optics, and instead of ammunition, the chamber held a rotating disc. This was a portable filming system that had an equivalent shutter speed of 1/720 s and could film at rates up to 12 pictures per second (pps). Armed with the flexible film now developed by Kodak, Marey in 1890 could run his camera up to 100 pps and the video camera was born.

From the early 1900's through the 1930's the development of movie industry began. The standards for recording and playback rate was between 16-18 fps. When sound was introduced and embedded on the film, the frame rate needed to be increased to 24 fps, so the film would be moving fast enough to capture and replay the sounds properly.

The development of the modern day equivalent of high speed videography didn't happen until the 1930's with the introduction of F.J Tuttle's rotating prism. Kodak utilized Tuttle's invention and produced the Eastman 1, which could run up to 250 pictures per second (pps), later replaced by the Eastman 2, which ran up to 2500 pps, and two years later the Eastman 3, which could run at 3000 pps. World War II spurred the development of high speed cameras further and faster than previously with cameras such as the Westrex Fastax being capable of a recording rate up to 8000 pps using the same rotating prism principle. The first major improvement to the rotating prism design, created by Red Lake Camera Design out of California, developed the Hycam, which improved the sharpness of the acquired images to a level consistent with a handheld camera. In the 1970's one major revision was made to include the Hycam's optical arrangement with the rotating prism design in the Photec camera, produced by Visual Instrumentation. This configuration was in use up to the 1990's where the rise of the digital high speed camera soon overtook the traditional film high speed camera.

The beginning of electric imaging took place in the 1950's through the 1960's [19]. The first cameras were composed of a three layer system. The first layer was a photocathode, the second an electronic shutter and the third was a fluorescing screen. The fluorescing screen could be imaged by exposing film to the screen to store the data. In 1970 a subsequently abandoned patent was filed by Willard Boyle and George E. Smith titled '*Electronics Design, New Surface-Charge Transistor has High Data Storage Potential*'. In this disclosure was the first introduction of the charged coupled device, which would soon revolutionize digital videography and photography. Both authors would be awarded the Nobel Prize in physics in 2009 for their invention.

A new patent [20] was filed in 1973 describing the CCD in full detail which was the enabling technology for digital imaging. The first patent to take advantage of the new CCD invention originating in 1972 by Tompsett [21]. Utilizing a CCD, in 1980 Eastman Kodak introduced the SP2000. The SP2000 was one of the first imaging systems with an electronic tape based recording that could capture images up to 2000 pps full frame or 12,000 pps split frame with a resolution of 240x192 and delivered instant playback. The instant playback and simple setup were considered great advantages over the film based systems which required tremendous power supplies and thousands of feet of film, and was one of the driving mechanism to the rise digital high speed videography.

By 1989 the Ektapro 1000 EM was introduced which was the first camera to use solid state memory instead of a magnetic tape reel. In 1991, Photron released at the time the fastest high speed digital camera, a Model 4540 which could acquire data at a rate of 40,500 pps. In 1992 Photron [22] released the FASTCAM Super10k which had a native resolution of 512x480 pixels with an 8 bit A/D. In 1995 Vision research was granted a patent [23] that covers the geometry of the modern day high speed camera. It consists of single body system where the data can be recorded and stored all in one unit. Prior to this patent the camera body was separated from the recording unit. In 2000 Photron made a new move to depart from using the then commonly used charged coupled device, to the complementary metal oxide sensor for high speed cameras, which are commonly used as the writing of this document.

As of 2016 the Photron FASTCAM SA5 [24], has a deep well capacity of $43000e^-$ and noise floor of $38e^-$ yielding a dynamic range of 61 dB digitized over a 72 dB range. The Vision Research v711 [25] has a deep well capacity of $23200e^-$ and a noise floor of $29e^-$ yielding a dynamic range of 57.7 dB digitized over a 72dB range. In either case these cameras utilize a CMOS sensor that has about a 1000:1 dynamic range. As discussed in this section, digital

camera have exceled when comparing ease of use to film cameras, however they still cannot compete with the dynamic range of film, about 80-90 dB depending on the film used. To overcome the dynamic range limitation on digital imagery, some manufacturers have used per pixel exposure time bracketing to increase the dynamic range such as Vision Research [26]. With extreme dynamic range mode turned on an increase to about 80 dB is achievable, however is still insufficient for the recording of highly energetic events, common with government applications. This work fills in that capability gap.

Introduction to high dynamic range imagery

The human eye adapts to a scene's range of light through photo bleaching and pupil dilation [27] to overcome the static dynamic range limitation. This adaptation mechanism extends the human vision system's functional range to about 12 orders of magnitude and allows for a resolvable 10,000:1 dynamic range for a given adaptation level [28]. High Dynamic Range (HDR) imaging was developed in response to the frequent breach of the upper limit of a single camera's acquisition range.

A solution to the dynamic range limitation was first presented in 1962 when Wyckoff [29] proposed the use of a composite film, composed of layers of varying speed photosensitive material with similar spectral sensitivity to capture a high dynamic range scene. Each layer of film would be printed to a different color and superimposed to form a pseudo color image that would represent how bright the various regions were.

In the 1980's, the invention of the charged coupled device created a wave of interest in digital imaging. However, digital imaging at the time had a significantly reduced dynamic range when compared to traditional film since the noise floor of the sensors were high relative to the deep well capacity. In 1995, Mann and Picard [30] introduced the concept of combining digital images of different exposures. They called each exposure level a Wyckoff layer where each layer is analogous to an image captured from exposure bracketing. Mann and Picard began their image formation by generating a camera sensitivity curve by comparing the value of differently exposed spatially aligned pixel values that were scaled by an unknown value. The amalgamation of the unknown values scaling parameters created a lookup table and is called the camera response function. The camera response function can be used for the linearization of the acquired data for further processing. Mann and Picard introduced the concept of fitting the data to a power curve via parametric regression to complete the lookup table for all potential digital

values. Mann and Picard also introduced the concept of using a weighted average as a method to reform a high dynamic range image from the series of images into an HDR image.

A high dynamic range image is defined as an image with a greater dynamic range then current media allows for representation. The current standard for display media is 8 bit output per color channel, or 256 shades of intensity. Typical monitor displays use a gamma mapping to represent the information in the form of light on a monitor or printed picture on paper. Gamma mapping scales the linearly recorded light intensity by a power function, typically either 1.8 or 2.22. If the acquired data spans a range greater than can be natively displayed it is considered high dynamic range. For this reason most cameras with 10 bit A/D or higher acquisition capabilities are considered high dynamic range capture. An example of a high dynamic range capture can be seen in Figure 1-2. In this figure a halogen light source was placed behind a color chart to produce an extremely large dynamic range scene. The dynamic range of this image is around 120 dB which far exceeds any current display's capability, making it impossible to natively display this information without some form of remapping of the acquired information to the display space.

Figure 1-2. Pseudo color representation of a very high dynamic range image.

The remapping of the data into the display space is called tone mapping. There are many different ways to tone map the data into some form of display space which will be reviewed in Chapter 4. The general idea is try to recreate an image in a way the preserves the pertinent information in the acquired scene and reproduce it either in a perceptually realistic method or in a way that looks artistic. The intensity graph presented in Figure 1-2 was scaled by converting the high dynamic range image into a decibel scale and applying a color map to each range. This

image represents one form of tone mapping, however does not represent the actual look of the scene.

How to form an HDR image

There have been several methods introduced in the past two decades to form a high dynamic range image. Akyuz [31] outlines the most common methods, stemming from the work of Debevec and Malik [32], Mitsunaga and Nayar [33], Robertson *et al* [34], Grossber and Nayar [35], and Dedrick *et al* [36]. Akyuz *et al* found that Grossberg and Nayar's method tended to recreate the most accurate images. Grossberg and Nayar introduced the EMoR (empirical model of response) algorithm to find the camera response function. Grossburg and Nayar created a database of over 200 different camera response models from different cameras and films. They reduced the parameter space that the camera response function must be within by stating that all camera response models must lie within the range of all of these samples. A principal component analysis is performed using 170 sets as training vectors and 26 as test vectors to find the eigenvectors from the principal component analysis that describe the camera response function in an energy hierarchal form. They found a three parameter estimate was adequate to accurately recreate the camera response function.

Debevec and Malik's and Mitsunaga and Nayar's algorithm were the most robust to noise within the calibration images. Mitsunaga and Nayar's method of camera response recovery tended to show the least amount of variance between the results from different scenes. Their method was to fit the camera response function to a polynomial of varying degrees. Each polynomial's error would be checked and an ideal polynomial order found. They found with a polynomial order of three there was little more to be gained for most instances.

One trait that all of these methods hold in common is the method in which the calibration is done. A single camera is used with sequential shots taken with varying exposure times. Debevec and Malik originally provided the frame work on how to do this by finding the camera response function from either a linear or nonlinear camera and forcing the camera response function to fit a logarithmic curve. Debevec and Malik start their derivation by stating the sensor output, Z_{ij}, is proportional to the irradiance and exposure time

$$Z_{ij} = E_i \Delta t_j \qquad 1\text{-}7$$

where E_i is the sensor irradiance on pixel i and Δt_j is the exposure time for some exposure, or physical attenuation as will be discussed later, j. This equation is simplified stating the stored pixel value, Z_{ij}, must be a function of the sensor irradiance and the exposure time yielding

$$Z_{ij} = f(E_i \Delta t_j) \tag{1-8}$$

for the *i* th pixel in the *j* th image. Assuming the process is monotonic and reversible

$$Ln\left[f^{-1}\left(Z_{ij}\right)\right] = Ln(E_i) + Ln(\Delta t_j) \tag{1-9}$$

simplifying the equation

$$g\left(Z_{ij}\right) = Ln(E_i) + Ln(\Delta t_j) \tag{1-10}$$

where $g(Z_{ij})$ is the image response function of the stored pixels. A series of points can be assembled from different images of the same scene taken at different exposures. A function can be fit using a linear optimization to find a smooth curve that minimizes the mean square error of the derived response function. The objective function for this linear optimization has been defined as

$$\vartheta = \sum_{i=1}^{N}\sum_{j=1}^{P}\{w\left(Z_{ij}\right)\left[g\left(Z_{ij}\right) - lnE_i - ln\Delta t_j\right]\}^2 + \alpha \sum_{z=Z_{min}+1}^{Z_{max}-1}[w(z)g"(z)]^2 \tag{1-11}$$

where w(Zij) is a weighting function, defined as a triangular function by Debevec *et al* to give priority to the pixels that are not either over or under exposed, and α is a scaling term used to smooth out the derived response function. The first term in this equation is the fitting term and the second term is a smoothing term that forces the data to fit more of a logarithmic shape. An example camera response function from a Sony RX100 and Prosilica 1900 C for each color channel derived using this method can be seen in Figure 1-3.

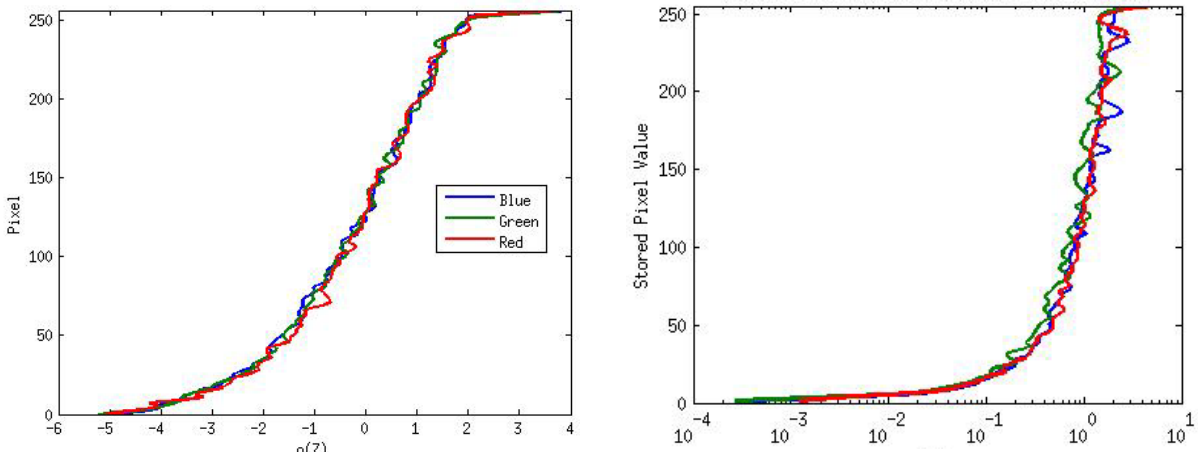


Figure 1-3. Left. Log_e of the camera response function for a Sony RX100 camera. Right. Camera response function for a linear Prosilica 1900 C camera.

To form the final high dynamic range image from the acquired exposures a weighted average should be used with the found camera response function.

Figure 1-4. Left. Six images taken at different exposure times. Right. The tone mapped reconstruction from the six images. The sky in the image is visible as well as the shadow detail in the chair.

The high dynamic range image, ϕ_{hdr}, is defined as

$$\phi_{hdr,i} = \sum_{j=1}^{J} w_{ij} \frac{f^{-1}(Z_{ij})}{\Delta t_j} \left(\sum_{j=1}^{J} w_{ij} \right)^{-1} \quad 1\text{-}12$$

where J stands for the number of images or cameras. An example of a tone mapped high dynamic range image can be seen in Figure 1-4. In this figure six images were acquired with the Sony RX100 camera at varying exposure times. Using the camera response function shown in Figure 1-3 and (1-12) with a triangle hat weighting function.

This method works well provided the scene is invariant to change between exposures, the sample population is sufficient to resolve the expected value, and all errors associated with the image formation are symmetrically distributed. Unfortunately a simple demonstration can show how each one of the mentioned conditions do not hold true, seen in Figure 1-5.

To demonstrate how this method fails, a scene utilizing 120 logarithmically spaced exposures was formed. If the premise is correct that (1-12) tends to the actual value given enough samples then a properly exposed image and an HDR image of the same scene should tend to the same value.

A point source of light was created with a halogen bulb source projected through a 32 blade aperture set to its smallest opening, about 2mm. The camera, a Prosilica 1900C was coupled with a fixed focal length 35 mm f-1.4 lens set to an f-16 to prevent out of plane focal artifacts and spatial resolution limitations due to diffraction from the lens aperture. The camera was placed sufficiently far away from the light source so the light source only encompassed a small number of pixels, about 9. The light intensity was adjusted so the full range of the scene could be acquired using a single camera.

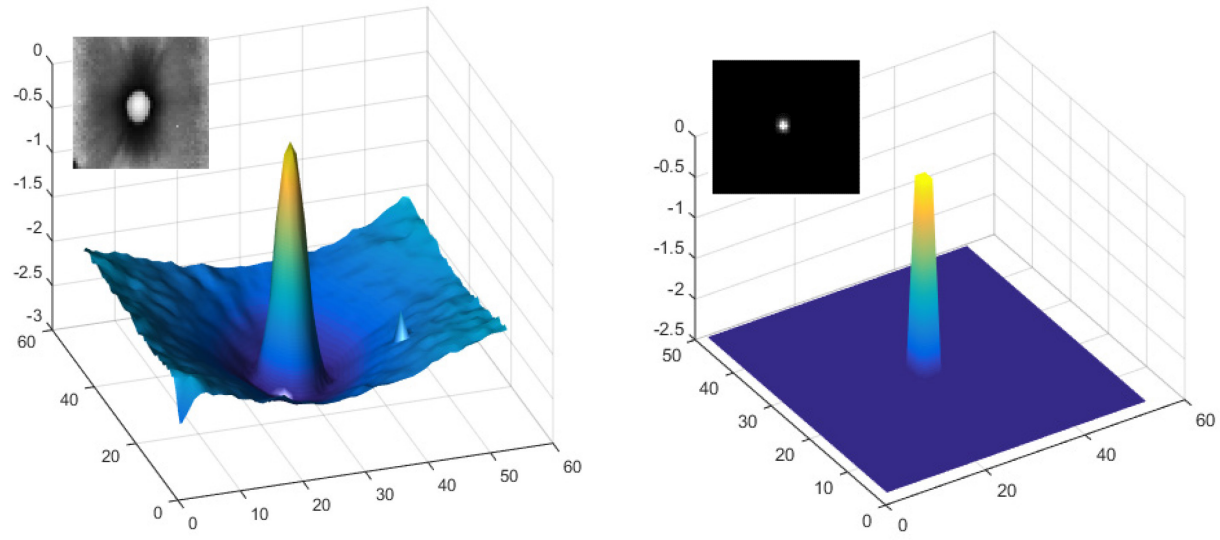


Figure 1-5. Left. HDR recreation of point source. Right. Single properly exposed image.

The left hand side of Figure 1-5 is the high dynamic range image recreated utilizing Debevec's method and is represented as a surface plot where the height of the plot is the log10 of the normalized magnitude. The right hand side image is an image from a single exposure that can capture the full range of the scene. Several errors are causing the significant degradation the HDR recreated image. First there is significant noise in the darker areas of the image. This is caused by the numerical amplification of the noise, indicating a problem with the weighting function. Second, the point source looks larger in diameter than it really is and does not have a sharp roll off when compared to the standard image. This is caused by veiling glare degradation and is not a symmetric error, therefore causes a bias in the final image formation. Another example of these problems in the presence of limited samples can be seen in Figure 1-6. The flare degradation manifests itself as color abnormality around the outside of the bulb. Third, it is impractical from the standpoint of forming a video to acquire 120 different images, so it must be determined how many samples will be needed at a minimum to span a given range and how to setup the camera system. The proceeding chapters will discuss each one of these concerns at length and propose new methods to avoid them when forming a high speed high dynamic range video.

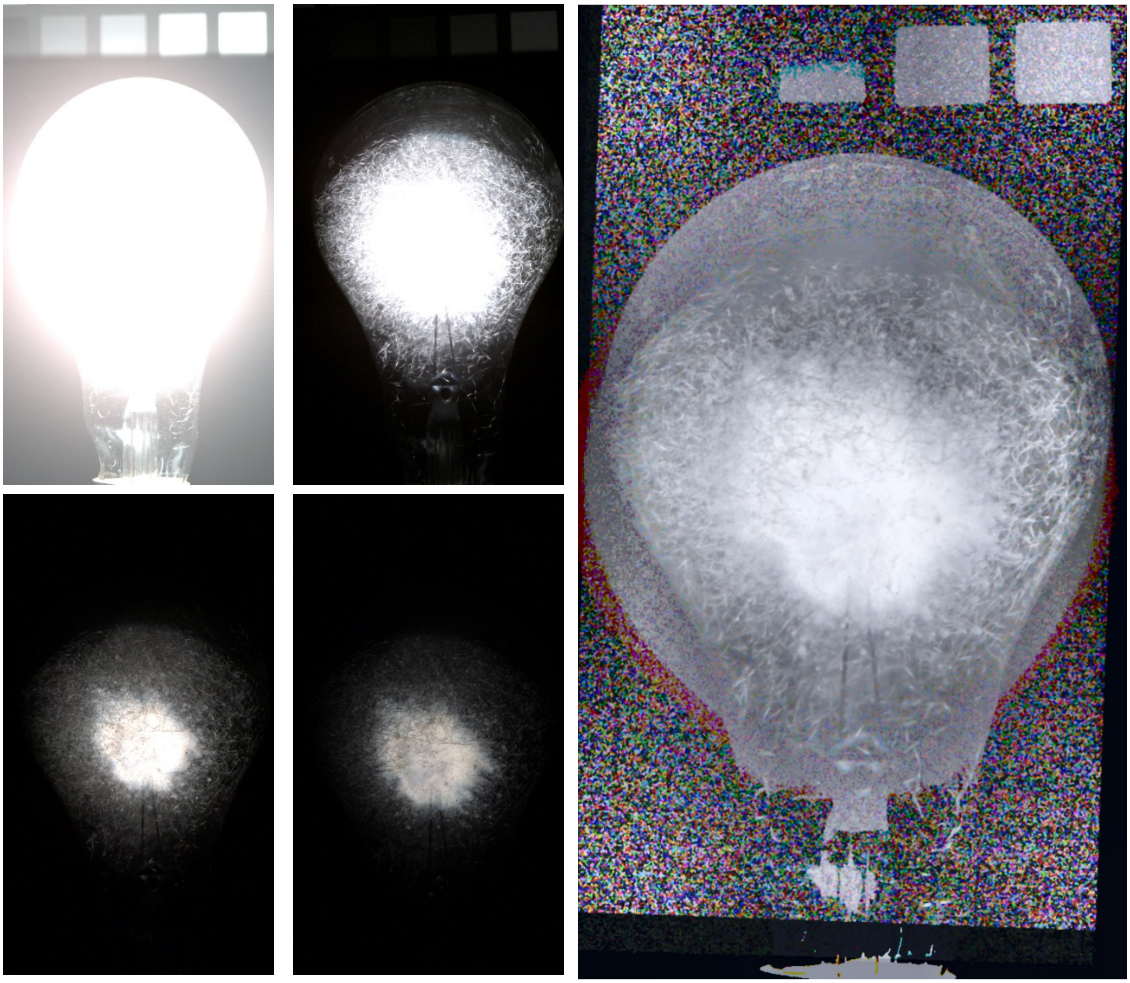


Figure 1-6. Sample of flare and misalignment degradation from small samples for a High Speed HDR video frame combined with a suboptimal weighting function. The 4 original samples are in the left two columns and the tone mapped HDR video frame is on the right hand side.

Chapter 2. Developing a simplified weighting function for high dynamic range frame formation

In the introduction, the first problem that was identified and introduced was the added noise in the dark region of the image. This was caused by the weighting function's inability to reduce the numerical amplification of the noise for the short exposures, numerically amplifying the error on the final formation. In this section of the paper prior art will be reviewed on varying weighting functions designed to try to mitigate this kind of error. Once the review is complete, a novel weighting function will be introduced that significantly reduces the error relative to the standard triangle weighting function and is computationally inexpensive. The new weighting function will be compared to prior art by synthetically creating a logarithmically increasing radiant exitance scene, parsing out the varying images from a camera acquisition model, and forming new HDR images based on the parsed out images. The HDR image's errors can then be compared to the original baseline to demonstrate how the errors propagate and the effectiveness of the new function.

Overview of the current weighting function

Since the original inception of digital high dynamic range imagery in 1995, there have been several proposed weighting functions designed to reduce the overall error associated with high dynamic range image formation. They can be grouped into three categories; proportional to the acquired digital value, proportional to the slope of the camera response function for the given digital value, or proportional to a noise estimate on the estimated radiant exitance. A review of most of the weighting functions, mathematically described, can be viewed in Table 2-1. All nomenclature has been changed from the original papers to allow for a consistent comparison.

In 1995 Mann and Picard not only introduced the concept of HDR image formation, but also introduced the first weighting function [30]. Mann and Picard defined a weighting function that would weigh the acquired digital values proportional to the rate of change of the logarithm of the camera response function. This was done to reduce the effect of quantization and make the error appear uniform. In 1997 Debevec and Malik [32] introduced the first weighting function that was proportional to the acquired digital value. The weighting function was a piecewise function that takes the form of a hat function centered in the middle of the analog to digital conversion range. This was done with the rationale that the data at the beginning of the

range would be degraded by a low signal to noise ratio and data at the other extreme became nonlinear and subject to flare degradation.

In 1999 Robertson *et al* [37] introduced a weighting function that was similar to a Gaussian curve assuming the sensor sensitivity would be highest around the center of the camera response curve, and the extremes would present little to no usable information. Also in 1999 Mitsunaga [33] first introduced the concept of incorporating a camera error model by creating a weighting function that is a first order approximation of the signal to noise ratio. In 2001 Tsin *et al* [38] created the first statistical characterization to form a weighting function. In general there were three modeled uncertainty terms. The first term modeled the thermal noise, the second term incorporated quantization error and amplifier noise, and the third term modeled the contribution from shot noise.

In 2005 Ward [39] proposed a modification to Mitsunaga's weighting function by multiplying it with a broad hat model to reduce the relative importance of the pixel values at the extremes of the acquisition range. In 2006 Kirk [40] introduced the first weighting method that is exactly the reciprocal of the square of the error. This was determined to be the optimal form of the weighting function to minimize the error on formation. In 2007 Akyüz [41] proposed a modification to Ward's weighting function. Instead of the weighting function being a function of the digital value, the weighting function should be a function of the estimated radiant exitance.

In 2010 Granados [42] and Hasinoff [43] introduced weighting functions that are designed to minimize the error on formation by accurately modeling the total variance in the system. To accomplish this, a recursive scheme was developed to accurately model the shot noise. Granados takes into account fixed pattern noise, read noise, shot noise, and dark current noise. Hasinoff's weighting method includes shot noise, read noise, and quantization error.

Table 2-1. Various weighting functions

Author	Weighting Function[1]
Mann and Picard [30]	$w(z) = \frac{1}{f'[Log(z)]}$
Debevec and Malik [32]	$w(z) = \begin{cases} z - z_{min}, & for\ z \le \frac{1}{2}(z_{min} + Z_{max}) \\ z_{max} - z, & for\ z \le \frac{1}{2}(z_{min} + Z_{max}) \end{cases}$
Robertson *et al* [37]	$w(z) = e^{-16\left(\frac{z^2}{{q_s}^2} + \frac{1}{q_s} + \frac{1}{4}\right)}$
Mitsunaga and Nayar [33]	$w(z) = \frac{f^{-1}(z)}{f'(z)}$
Tsin *et al* [38]	$w(z) = \frac{\Delta t}{\sqrt{g\Phi {\sigma_s}^2 + {\sigma_{dc}}^2 + f'(z)^2 {\sigma_q}^2}}$
Ward *et al* [44]	$w(z) = f^{-1}(z) f'(z)[1 - \left(\frac{2z}{q_s} - 1\right)^{12}]$
Kirk and Anderson [40]	$w(z) = \frac{\Delta t^2}{f'(z)^2 \left(\sigma_r + \sigma_q\right)^2}$
Akyüz and Reinhold [41]	$w(z) = f^{-1}(z) f'(z)[1 - \left(\frac{2\Phi}{q_s} - 1\right)^{12}]$
Hasinoff *et al* [43]	$w(z) = \frac{{\Delta t_j}^2}{\Phi \Delta t_j + {\sigma_r}^2 + {g_j}^2 {\sigma_q}^2}$
Granados *et al* [42]	$w(z) = \frac{{\Delta t_j}^2 g^2 {a_i}^2}{g^2 \Delta t_j (a_i \Phi_i + 2\overline{\mu_D}) + 2{\sigma_r}^2}$

[1] nomenclatures changed from original papers for consistency

Introduction to a simplified weighting function

The weighting function that can be used to minimize the error introduced from a weighted average has been shown to be $1/\sigma_t^2$ [40] [42], where σ_t is the standard deviation of the total error associated with the formation of the individual images. Subsequent weighting functions have been developed by Tsin, Krik, Hasinoff and Granados to define σ_t in a manner such that the weighting function is able to accurately reduce the error associated with the formation of the

final HDR image. The five most significant sources of error that affect the signal are reviewed by Healey [3]. They are defined as shot noise, reset noise, dark current, read noise, and quantization error.

All sources of uncertainty associated with the image formation can be considered independent, meaning their covariance is equal to zero and therefore not a function of one another. The total variance, σ_t^2, can be found by adding the various contributors in quadrature

$$\sigma_t{}^2 = \sigma_s{}^2 + \sigma_{rst}{}^2 + \sigma_{dc}{}^2 + \sigma_{rd}{}^2 + \sigma_q{}^2 \quad 2\text{-}1$$

where σ_s^2 is the shot noise variance, σ_{rst}^2 is the reset noise variance, σ_{dc}^2 is the variance arising from dark current noise, σ_{rd}^2 is the read noise variance, and σ_q^2 is the quantization noise variance. Each noise term natively has different units. The units associated with shot noise is typically in photons, read noise in electrons or Volts, dark current in electrons or ADU, quantization noise in ADU and read noise in ADU. In order to have a meaningful comparison all the noise terms must have consistent units. To do this a camera model will be used to convert their respective units into their equivalent recreated radiant exitance values.

First a simplified model of the image sensor must be defined. Starting from a Lambertian source, the photon flux, φ, can be defined as

$$\varphi = \frac{1}{hc}\int_{\lambda_2}^{\lambda_1} L(\lambda)d\lambda \quad 2\text{-}2$$

where h is Plank's constant, c is the speed of light, L(λ) is the spectral radiance, and λ is the wavelength. The subscripts in the integral indicate the range over which the spectral radiance is confined within. Following the derivation by Fiete [45], the contribution from the optics and integration time must be considered. The total attenuation from the aperture will be proportional to $\pi/4(f/\#)^2$, where f/# stands for the f number of the lens, and the integration time is multiplicative. Multiplying (2-2) by the quantum efficiency, η, for an exposure value setting *j,* at pixel location *i,* the total charge accumulation, s_{ij} can be defined as

$$s_{ij} = \frac{\pi g_j A_j \Delta t_j}{4\left(f/_{\#}\right)^2 \left(\frac{s_2}{s1} - 1\right)^2 hc} \int_{\lambda_{i,min}}^{\lambda_{i,max}} \eta(\lambda_i)\tau_{optics}(\lambda_i)\left[h_{sys}(\lambda_i) * L_{scene}(\lambda_i)\right]\lambda_i d\lambda_i \quad 2\text{-}3$$

where *g* is the sensor gain, A_j is the area of the pixel, Δt_j is the exposure time, s_2 is the distance to the imager, s_1 is the distance to the object, τ_{optics} is the transmission of the optics and filters

being used, h_{sys} is the point spread function of the imaging system, * indicates a convolution, and L_{scene} is the spectral radiance of the scene. Since the objective is to compare the relative effects of the different sources of noise on the final image formation, (2-3) can be simplified with several assumptions.

Assuming the lens and associated filters in use were designed to mitigate any wavelength variations over the passband, τ_{optics} can be considered a constant and come out of the integral and will be a value from 0-1 for the system optical attenuation. For an actual lens chromic aberrations can be seen in images, particularly the edges of the lens, but for the sake of the analysis it will be assumed to be minor and negligible. The second assumption that needs to be made is the geometric distortion due to the lensing is negligible, and the low pass filtering effects indicated in the point spread function will be negligible. With these assumptions the scene's spectral radiance in the direction of the lens, the total radiant exitance, will be equal to the imager's spectral irradiance at all locations on the imager.

The quantum efficiency of the sensor will be estimated by the maximum quantum efficiency for a camera's color channel and the associated wavelength will be assumed monochromatic for that channel. If the camera is a monochrome camera the mean quantum efficiency will be used, and a monochromatic input assumed at the center of imager's spectral passband. Finally, assuming the object being imaged is much farther from the lens than the sensor, the ratio s_2/s_1 approaches 0 and can be neglected. The final assumption that must be imposed is the radiant exitance level is low enough to only produce enough charge to just saturate the sensor, and no energy is communicated to neighboring pixels. With these assumptions in place, the charge accumulation at the sensor can be reduced to a simple linear equation as

$$s_{ij} = \frac{\pi g \tau_{optics,j} A_j \Delta t_j \eta_j}{4(f/\#)_j{}^2 hc} \lambda_i \Phi_i \qquad 2\text{-}4$$

where Φ_i is the radiant exitance. For convenience, the term $(\pi\tau_{optics})/(4(f/\#)_j{}^2)$ will be called the system physical attenuation, $A_{phy,j}$. The estimated scene radiant exitance can be found from a reported digital value, Z_{ij}, using

$$\Phi_{ij} = \frac{Z_{ij} N_{well,j} hc}{q_s A_j \Delta t_j A_{phy,j} \eta_j \lambda_i g_j} \qquad 2\text{-}5$$

where N_{well} is the deep well capacity of the sensor, and q_s is the number of discretization steps from the camera's analog to digital converter. It is convenient to define the maximum resolvable radiant exitance, $\Phi_{max,j}$, that a camera can resolve by setting Z_{ij} in (2-5) equal to the number of quantization steps which yields

$$\Phi_{max,j} = \frac{N_{well,j} hc}{A_{phy,j} A_j \Delta t_j g_j \eta_j \lambda_i} \quad 2\text{-}6$$

The mean number of photons, γ_j, collected by the sensor over the integration time can be expressed as

$$\gamma_j = \frac{A_{phy,j} \Delta t_j A_j \lambda_i \Phi_i}{hc} \quad 2\text{-}7$$

Shot noise arises from the quantum nature of light. The energy arrives at the sensor site in packets with a mean value and some fluctuation that is proportional to the number of events, and can be modeled as a Poisson distribution. The variance of the shot noise is equal to the mean therefore the standard deviation is equal to the square root of the mean value. Converting (2-7) to ADU then substituting the result into (2-5), the shot noise at the sensor, σ_s, in units of W/m^2 becomes

$$\sigma_s = \sqrt{\frac{hc}{A_{phy,j} \Delta t_j A_j \lambda_i} \Phi_i} \quad 2\text{-}8$$

Reset noise can be a limiting factor on the sensor's effectiveness. A charge amplifier is typically used to convert the acquired charge to a voltage. The sensor capacitor must be reset before each packet is measured to a reference voltage. There will be thermal noise on the reference voltage and it will cause an error on the determination of the exact charge collected by the sensor site. The standard deviation of the reset noise is

$$\sigma_{rst} = \frac{\sqrt{kTC}}{q} \quad 2\text{-}9$$

where k is the Boltzman constant, T is the temperature in Kelvin, C is the capacitance of the photodetector, and q is the number of electrons per coulomb, $1.602*10^{-19}$. The total number of electrons must be converted to units of W/m^2 to see how the reset noise affects the overall estimate of the scene radiance. The reset noise in W/m^2 is defined as

$$\sigma_{rst} = \frac{\sqrt{kTC}}{qA_j\Delta t_j}\left[\frac{g_j hc}{\eta_j \lambda_i}\right] \quad 2\text{-}10$$

Typically, modern sensors will use a correlated double sampler (CDS) or similar to mitigate reset noise since it can be very high relative to the measurement signal. A CDS will take two samples and report the difference. The first sample is taken from the reference voltage after reset and again after integration with the difference being the acquired signal, nearly eliminating this kind of error. Dark current is the current generation due to the temperature of the silicon lattice of the imager without the presence of any sensor irradiance. An empirical relationship [46] for dark current, d_c, in electrons/pixel/s, can be defined as

$$d_c = 2.5 * 10^{15} A I_d T^{1.5} e^{-\left(\frac{E_g}{2kT}\right)} \quad 2\text{-}11$$

where I_d is the dark current measured at 300K, and E_g is the bandgap energy at the operational temperature of T. It can be difficult to find these specifications or to get a physical measurement in nA of the dark current. Therefore a direct measurement of the dark current in units of ADU will be used. This can be accomplished by taking several measurements of a dark frame, an acquisition with the lens cap on. This process is stated in detail in the Appendix. There are two components to dark current; a mean value defined by (2-11) and the photon shot noise which is proportional to the square root of the dark current signal. The total dark current noise, σ_{dc}, represented in W/m^2 can be expressed as

$$\sigma_{dc,j} = \frac{N_{well,j} hc}{A_j \Delta t_j A_{phy,j} q_s \eta_j \lambda_i g_j}\left(\overline{\mu_{dc,j}} + \mu_{dc,j}\right) \quad 2\text{-}12$$

where $\overline{\mu_{dc}}$ is the mean offset caused by the dark current and μ_{dc} is the dark current noise in ADU. It is common practice to perform a dark frame subtraction before an image is acquired. This is sometimes referred to as a session reference and subtracts the charge accumulation caused by the dark current offset. Some cameras have shaded pixels on the sensor to use as a reference if a full imager session reference is not possible. However, there will still be the shot noise associated by the dark current from the flow of charge even with a session reference. Referring to (2-11) the only way to reduce this form of noise is to increase the band gap energy or decrease the temperature of the sensor.

Read noise comprises of Johnson-Nyquist noise and flicker noise. Johnson-Nyquist noise is caused by the thermal agitation of the charge carried on the source follower and is broadband in spectrum. Flicker noise has a power spectrum that is equal to the reciprocal of the frequency

and is caused by the pixel readout rate. The faster the rate, the lower the noise, until ultimately the flicker noise is below the thermal noise. Since readout rates for even standard cameras can exceed 30 MHz, only the Johnson-Nyquist noise will be considered, and can be defined as

$$\sigma_{JN} = \sqrt{4kT\Delta f R} \tag{2-13}$$

where Δf is the bandwidth, and R is the resistance. The operational frequency and the follower resistance is not readily available information on most cameras. However, it is possible to measure the read noise in ADU per Healy's guidance reviewed in the Appendix. The read noise can be converted into W/m^2 in a similar manner as the dark current utilizing

$$\sigma_{rd,j} = \frac{N_{well,j} hc}{A_j \Delta t_j q_s A_{phy,j} \eta_j \lambda_i g_j} (\mu_{rd,j}) \tag{2-14}$$

where μ_{rd} is the measure standard deviation in ADU.

The final form of error is quantization error. The radiant energy estimate associated with the quantization error is typically treated as an error with constant variance. However, the quantization error will only have a constant variance if the total span does not change between exposures. In the proposed system the physical attenuation or exposure time will change for all Wyckoff layers. The quantization error associated with the discretization of a given signal is governed by a uniform distribution. The standard deviation of the quantization error can be defined by

$$\sigma_q = \frac{1}{\sqrt{12}} \frac{\Phi_{max,j}}{q_s} \tag{2-15}$$

Equations (2-7, 2-9, 2-11, 2-13, and 2-15) can be substituted into (2-1) to find the total magnitude of the error in units of W/m^2. Each error component has some dependency on one or more of the variables listed in Table 2-2. There are a total of 10 parameters that can modify the ideal weighting function. In the formation of HDR images or video, several methods have been proposed, discussed at length in the next chapter, to capture images or form a video stream while maintaining one or more of these variables. All of these parameters must now be considered when forming the simplified weighting function.

Table 2-2. Common Ranges of relevant terms and units.

Term	Range of Values	Units
g	1→10	e^-_{out}/e^-_{in}
q	$1.602*10^{-19}$	coulomb/e^-
C	10^{-13}-10^{-15}	Farads
q_s	2^8,2^{10}, 2^{12}, 2^{14}, 2^{16}	ADU
Z	0→q_s	ADU
η	0.1→0.6	e^-/s / γ/s
λ	$300*10^{-9}$→$700*10^{-9}$	m
A	$4*10^{-12}$-$9*10^{-10}$	m^2
N_{well}	10000→150000	e^-
hc	$1.9864*10^{-25}$	J*m
k	$1.3806*10^{-23}$	J/K
t	273→330	K
A_{phy}	$1*10^{-8}$→1	-
μ_{dc}	0→0.0008*q_s	ADU
μ_{rd}	0→0.0012*q_s	ADU

The simplified weighting function will assume a CDS or similar method is used to remove the reset noise and render it negligible. The dark current and read noise (2-12) and (2-14) will be combined yielding

$$\sigma_{ADU} = \frac{N_{well,j} hc}{A_j \Delta t_j q_s A_{phy,j} \eta_j \lambda_i g_j} \sqrt{(\mu_{rd,j}{}^2 + \mu_{dc,j}{}^2)} \qquad 2\text{-}16$$

To simplify further, the camera response function, f^1(Z) can now be defined. Using the same image formation ideology as Debevec, the camera response function can be defined as

$$f^{-1}(Z) = \Phi \Delta t A_{phy} g \qquad 2\text{-}17$$

where f^{-1}(Z) can be further defined by substituting (2-5) into (17) resulting in

$$f^{-1}(Z_{ij}) = \frac{hc N_{well,j}}{A_j \eta_{ij} q_s \lambda_i} \qquad 2\text{-}18$$

The camera response function is in units of J/m^2/ADU, and is an indicator of how much energy is required to advance to the next ADU value. Charged coupled devices and complementary metal-oxide semiconductors have a linear camera response function except at the

extremes of the acquisition range where saturation affects the linearity. Consumer level cameras typically down sample the acquired data to an 8 bit color space using a nonlinear function to enhance the image's visual appeal. Methods such as Debevec's, or Mitsunag's can be used to find the camera response function and linearize the output.

Comparing the remaining error terms to one another allows for the determination of which term is the dominant factor, and under what circumstances. There are a total of three error terms remaining. The first comparison takes place between the standard deviation arising from the read and dark current noise, σ_{ADU}, and the quantization error. If the error from the read and dark current noise is significantly larger than the quantization error the condition

$$\frac{{\sigma_{ADU}}^2}{{\sigma_q}^2} \gg 1 \qquad 2\text{-}19$$

must be true. Substituting (2-15) and (2-16) into (2-19), then simplifying using (2-18) yields the condition that the read and dark current noise will contribute significantly more degradation to the final image formation when compared to the quantization error when

$$12\left({\mu_{rd,j}}^2 + {\mu_{dc,j}}^2\right) \gg 1 \qquad 2\text{-}20$$

Utilizing Table 2-2 and (2-20), the sum of the ADU noise will be some percentage, δ, of the number of quantization levels

$$\delta \gg \frac{1}{12q_s} \qquad 2\text{-}21$$

which indicates δ must be greater than 0.03% for an 8-bit system and 0.002% for a 12-bit system for the quantization error to be ignored. This situation occurs when the camera has a native dynamic range over 70 dB for an 8-bit system or 93 dB for a 12 bit system. This is uncommon outside of specialized use cameras and indicates the ADU noise will be, for most cases, a dominant factor relative to the quantization error. The second comparison is between the shot noise and the quantization error:

$$\frac{{\sigma_s}^2}{{\sigma_q}^2} \gg 1 \qquad 2\text{-}22$$

Substituting (2-8) and (2-15) into (2-22) yields a percentage, P, of the discretized range where the shot noise will dominate

$$P \gg \frac{100N_{well}}{12g\eta {q_s}^2} \tag{2-23}$$

Referring to Table 2-2, the general range the right hand side of (2-23) will have is O[0.001%–10%]. This indicates there are conditions when the quantization error can contribute to the total error but only for a very small percentage of the acquirable range.

The quantization error contribution has been shown to be small for most cases relative to shot noise, read noise and dark current noise. The new weighting function can be formed where

$${\sigma_t}^2 \approx {\sigma_s}^2 + {\sigma_{ADU}}^2 \tag{2-24}$$

Therefore, knowing the optimal weighting function, w_{ij}, is the reciprocal of the total variance, the full simplified weighting function can be defined as

$$w_{ij} = \frac{\left(\Delta t A_{phy,j} g_j\right)^2 N_{well}}{f^{-1}\left(Z_{ij}\right)^2 \left(Z_{ij} q_s \eta g_j + N_{well}\left({\mu_{rd,j}}^2 + {\mu_{dc,j}}^2\right)\right)} \tag{2-25}$$

The full simplified weighting function can be reformed into a non-dimensional form by normalizing the weighting function by the maximum value it would take on, scaling all weights from 0-1. The maximum value will be found when the digital value is at a minimum, 0, the physical attenuators are at a minimum, the gain at a maximum, and the exposure time at a maximum, the supremum of the set $\Delta t A_{phy} g$. The non-dimensional weighting function, w^*_{ij}, can be defined as

$$w^*_{ij} = \left(\frac{\Delta t A_{phy,j} g_j}{max[\Delta t A_{phy} g]}\right)^2 \left(\frac{1}{1 + \frac{Z_{ij} q_s \eta g_j}{N_{well}\left({\mu_{rd,j}}^2 + {\mu_{dc,j}}^2\right)}}\right) \tag{2-26}$$

where max[·] indicates the maximum of the set. The first multiplying factor scales the weighting function by the magnitude of the attenuation and gain parameters. The second multiplying factor scales the magnitude based on the ratio of the shot noise contribution and read noise and dark current noise ratio to the camera's sensitivity. Provided there is a significant attenuation for a large span, the first term will drive the overall magnitude of the weighting function, since the second term is a scaled value from 1 and to a value proportional to the ratio of the errors, which is typically no more than one order of magnitude with high bit depth A/D systems and large deep well capacity sensors. The simplified weighting function can then be reduced to

$$w^*_{ij} = \left(\frac{\Delta t A_{phy,j} g_j}{max[\Delta t A_{phy} g]} \right)^2 \quad \text{2-27}$$

with little concern for altering the final formation in most cases. Both weighting functions can be seen in Figure 2-1 utilizing the camera parameters for a Vision Research Phantom 711 as the high speed camera in use. The simplified weighting function can be adequately estimated by (2-27) when the ratio of the photons per ADU is less than the noise floor or when the attenuation factors are much greater than the ratio of the errors. This indicates it will be effective in the darker regions of the image when there is little attenuation, or effective over the whole range when there is a lot of attenuation.

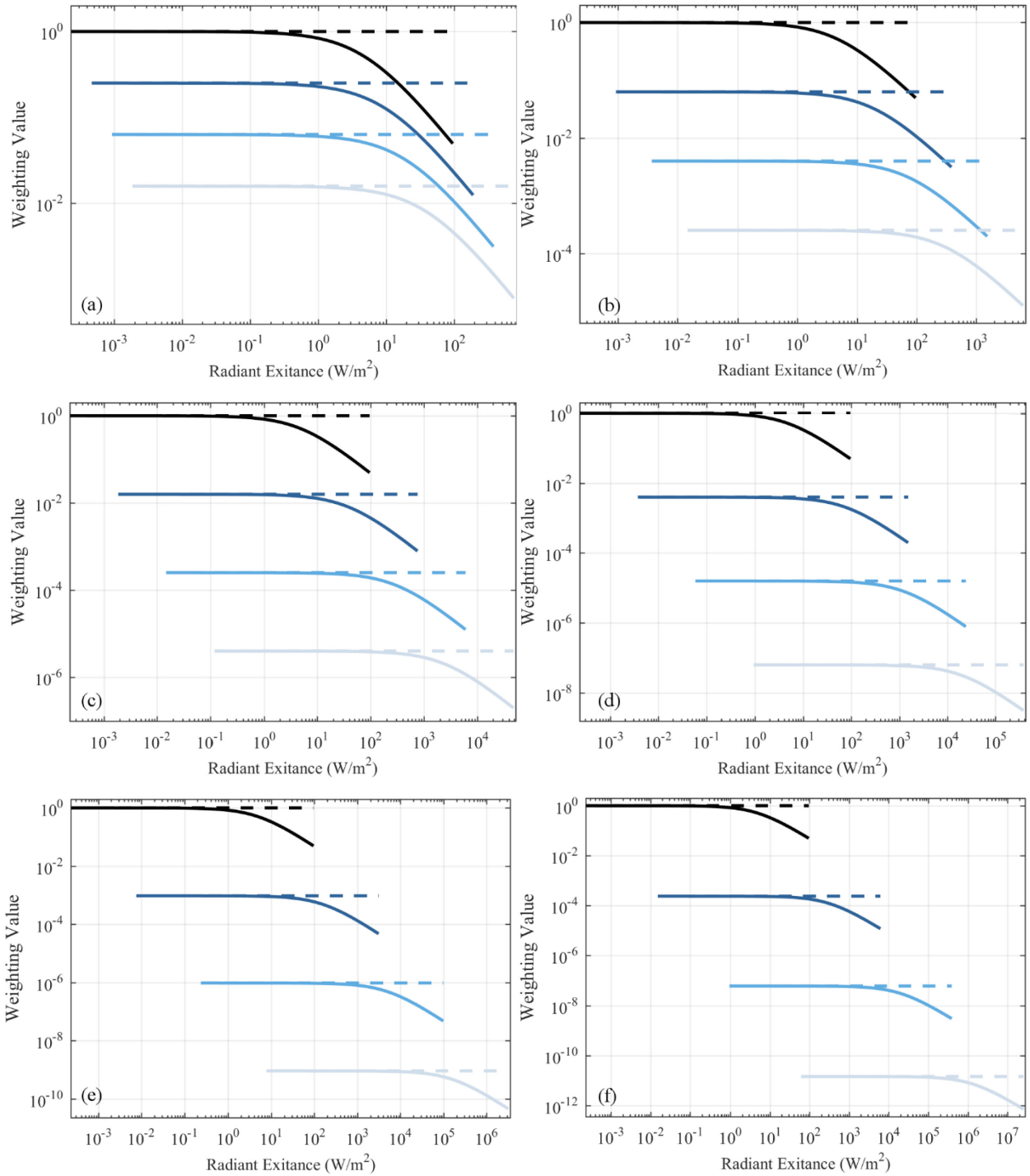


Figure 2-1. Dashed lines created utilizing (2-27) solid lines created with (2-26). Weighting function for four Vision Research Phantom 711 cameras equally spaced with (a) 1 stop (b) 2 stops (c) 3 stops (d) 4 stops (e) 5 stops (f) 6 stops

Generating synthetic images

Throughout the literature associated with testing high dynamic range imagery formation or tone mapping there is an inconsistency in the methodology used to demonstrate the error associated with the formation of the final image [31] [41]. Typically there is some reference scene that has been staged and regions within the image compared. A ground truth image will be created acquiring many image captures and forming the HDR image with the motivation that the many images will tend to the actual value. Observing Figure 1-5, it is clear that there are different sources of noise depending on what the scene is leading to an unexpected bias between comparisons in the analysis.

Another method that has been employed to generate a ground truth image is the creation of a synthetic scene through the use of programs [31] such as Blender^{Tm}, Mayia^{Tm}, or 3D Studio Max^{Tm}. These programs are ray trace programs that recreate what a scene would look like if full radiometric reconstruction took the place of a given synthetically created geometric scene including the object's material diffusivity, transparency, caustics, transmission impedance etc. These are very advanced programs that require significant time and skill to use, however produce very accurate results. Despite the accurate results, the synthetic scene can change from those that utilize this method. Another problem is the difficulty involved to reproduce other authors' work since the scene would have to be recreated exactly geometrically as well as the exact material properties and lighting sources used for all the components viewed in the scene.

In order to compare the various weighting functions with an unbiased methodology, a simulated scene needed to be created, and synthetic images generated so a ground truth can be known. The weighting functions to be tested are not a function of the spatial frequency, only the amplitude at each pixel site, therefore any pattern that spans the range of the acquisition system can be used for the simulation.

The image chosen will consist of vertical strips with increasing amplitude along the abscissa. Each successive strip will increase in amplitude logarithmically to span the full range of the camera system without undue redundancy. It is assumed a multi imager system is used to record the HDR video. The cameras are looking at the same field of view through a succession of optics such as a beam splitter with some level of transmittance to achieve the desired Wyckoff layers. The exposure time is held constant for each camera and only the optical train is used to attenuate the incident light. The process of creating the images is graphically depicted in Figure 2-2, and described in detail in the following subsection.

Forming the simulated images

A simulated radiant exitance field of monochrome light must be created to test the various weighting functions. A succession of vertical bars, each being 10 pixels wide and 20 pixels high was chosen. Each bar will contain 200 pixels at a given mean radiant exitance level for the respective processing method to estimate the actual value.

Once the radiant exitance field is created, the shot noise associated with each pixel can be calculated using (2-8). The radiant exitance field with shot noise, $\psi_{s,i}$, can be found by

$$\psi_{s,i} = \sqrt{\frac{hc}{\Delta t A \lambda}\Phi_i} \cdot randn(\cdot) + \Phi_i \qquad 2\text{-}28$$

where $randn(\cdot)$ is a normally distributed random number scaled from -1 through 1. The radiant exitance image with added shot noise needs to be synthetically acquired by the virtual cameras. The charge accumulation from a given scene for a radiant exitance level is modeled in (2-4) and yields a value for the charge accumulation at the sensor site.

Next the read and dark current noise is added to the simulated charge accumulation found from (2-4). The total charge at each sensor site, e^-_{tot}, is defined as

$$e^-_{tot} = S(\psi_{s,i}) + \frac{N_{well,j}}{qs}\sqrt{{\mu_{d,j}}^2 + {\mu_{d,j}}^2}randn(\cdot) \qquad 2\text{-}29$$

where $S(\psi_{s,i})$ is the charge accumulation from the shot noise image. If the dark current noise and read noise were found empirically and sensor gain is not unity, then the standard deviation values must be divided by the sensor gain as an added step. The data will be discretized into bins depending on the quantum step equivalent value. If the charge accumulation is greater than N_{well} the simulated digital value is clamped to the maximum value allowed by the A/D to simulate saturation. The data is quantized into their respective images, adding the quantization error to the simulation. The four images in Figure 2-2 were created assuming the use of four Vision Research Phantom 711 cameras with a four stop spacing, neutral density filters, and a lens with an f/# of 5.7. The Phantom 711 cameras have a 12 bit A/D, 23300 e^- deep well capacity, and nearly linear quantum efficiency of 0.32 per the manufacturer's data sheets. The camera response function is linear unless modified by the user through software. The measured read noise variance was 28 ADU and the measured dark current noise variance was 11 ADU.

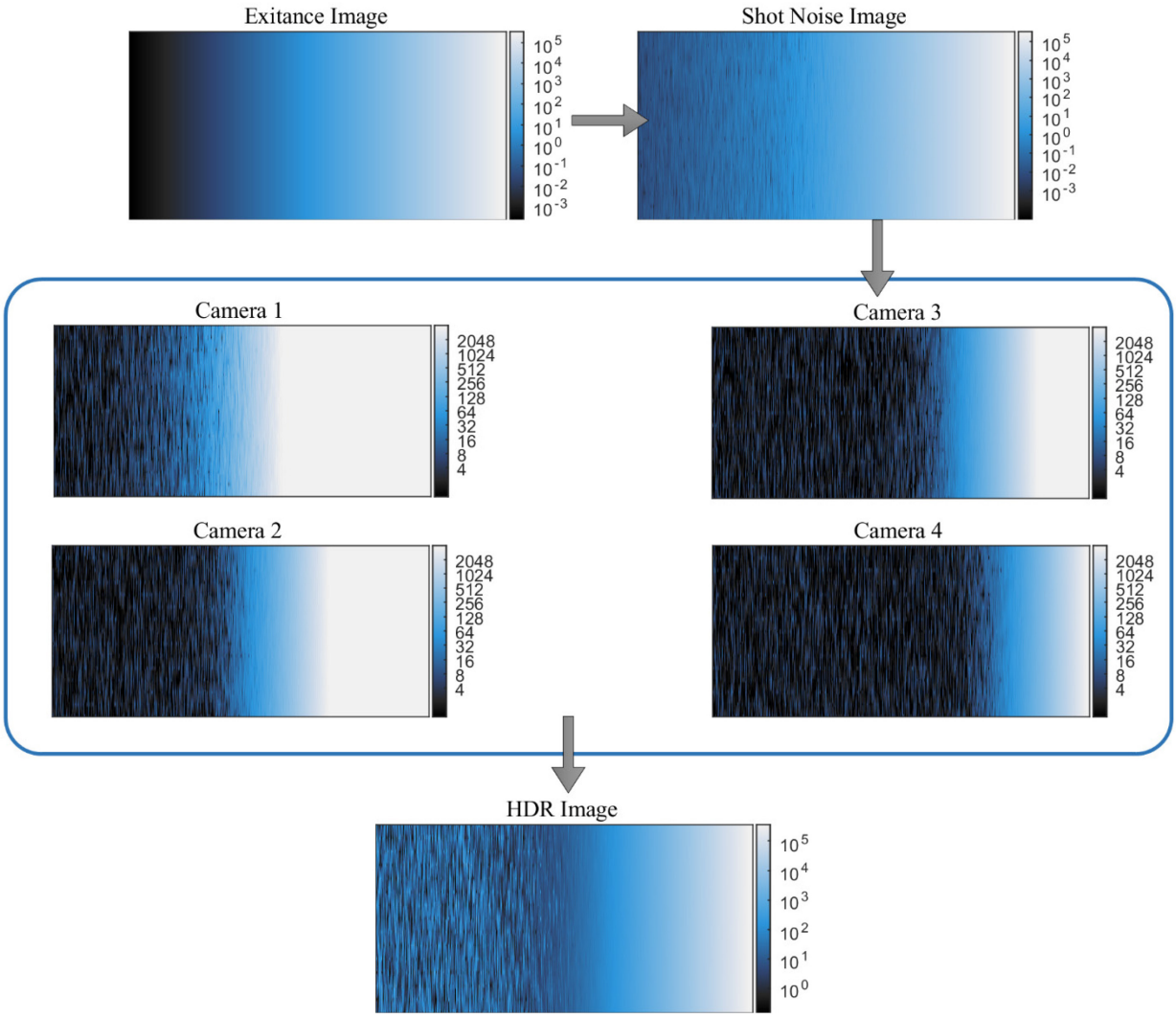


Figure 2-2. First the radiant exitance scene is created, then shot noise added. Data is discretized based on camera settings and an HDR image can be formed.

Forming the HDR image

Forming the HDR image is performed by utilizing a weighted average of the radiant exitance estimates from each frame as the maximum likelihood estimator that will optimally reduce the error associated with centrally distributed probability density functions. The nomenclature utilized by Debevec will be maintained for constancy with prior art. If the camera response function is nonlinear, common in consumer level cameras, it can be found by utilizing methods such as the ones introduced by Debevec or Mitsunaga and can be used in lieu of (2-18).

The final radiant exitance estimate, $\Phi_{HDR,i}$ from the series of acquired Wyckoff layers can be defined as

$$\Phi_{HDR,i} = \sum_{j=1}^{J} w_{ij} \frac{f^{-1}(Z_{ij})}{\Delta t_j A_{phy,j} g_j} \left(\sum_{j=1}^{J} w_{ij} \right)^{-1} \quad 2\text{-}30$$

where J is total the number of cameras or images used to form the different Wyckoff layers.

Results

All of the weighting methods introduced in Table 2-1 were utilized in the formation of independent HDR images. The images were created utilizing (2-30) and from the previously described synthetically created scene utilizing 200 logarithmically spaced radiant exitance columns, detailed in the prior section. Prior art utilizes either a change in exposure time or a change in sensor gain in order to generate the various images. The idealized video system uses varying optical attenuation; therefore, the weighting functions in Table 2-1 were slightly modified to accommodate this change. In order to account for the change in acquisition methods, for all weighting functions included in Table 2-1 that are a function of the exposure time, the quantity Δt_j was replaced with $\Delta t_j A_{phy,j}$. If the change is not made, then the weighting function will remain the same for each attenuation level, and the weighting functions will not work as originally intended.

Comparison of the different weighting functions

The exposure time was held constant for all cases at 15 μs, chosen as a realistic exposure time used in high speed videography. The architecture of the high speed camera system utilizes four high speed cameras each with their own lens. Each camera would be placed with the appropriate optics train such that they are all equidistant to the source being viewed. It is assumed that the cameras will be arranged in a way so there is no parallax or spatial misalignment with the acquired images. A constant aperture setting is assumed for each camera lens defined as an f/# of four. The f/# can be modified to enact the attenuation, however in a real system to negate the effect of varying depth of field, and dissimilar bokeh, the f/# should be held constant.

The attenuation required for the development of the Wyckoff layers will be produced from the use of neutral density filters and attenuation from beam splitters such as the system proposed

by Tocci [47]. Several scenarios were tested ranging from an attenuation equivalent of one, two, four, and six stops where one stop is defined as $1/2^{st}$ times the incident energy, where st is the number of stops. This corresponds to a multiplicative attenuation allowing the camera system to span a significant dynamic range. In all test cases the saturation value was given a weighting function of zero.

In order to compare the different weighting sets the mean error was found for each bin of identical mean radiant exitance levels from each image. This allows for a comparison from the recreated HDR image using all of the described weighting methods and the known ground truth for every radiant exitance level for each of the four attenuation levels, seen in Figure 2-3. It is assumed that any dark current offset would be subtracted out before the recombination takes place in an actual system removing any initial bias. The mean square error is defined as

$$ME = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\Phi_{HDR,i} - \Phi_i\right)^2} \qquad 2\text{-}31$$

where n is the number of points in each vertical strip. The mean error is an error metric equivalent to the standard deviation plus any bias error caused from the quantization of the radiant exitance field. The new weighting function proposed in (2-26) and (2-27) are plotted in Figure 2-3 with a continuous bold line along with the equal weighting case. The weighting functions that are proportional to the pixel digital value, Mann, Debevec, and Robertson, are designated with a long dash line. The weighting functions that are proportional to the image response function, Mitsunaga, Ward and Akuz, are designated with a dash dot line. The weighting functions that are proportional to the system error, Tsin, Kirk, Hasinoff, and Granados are designated using a short dashed line. The four line charts show the mean error on the ordinate and the radiant exitance on the abscissa.

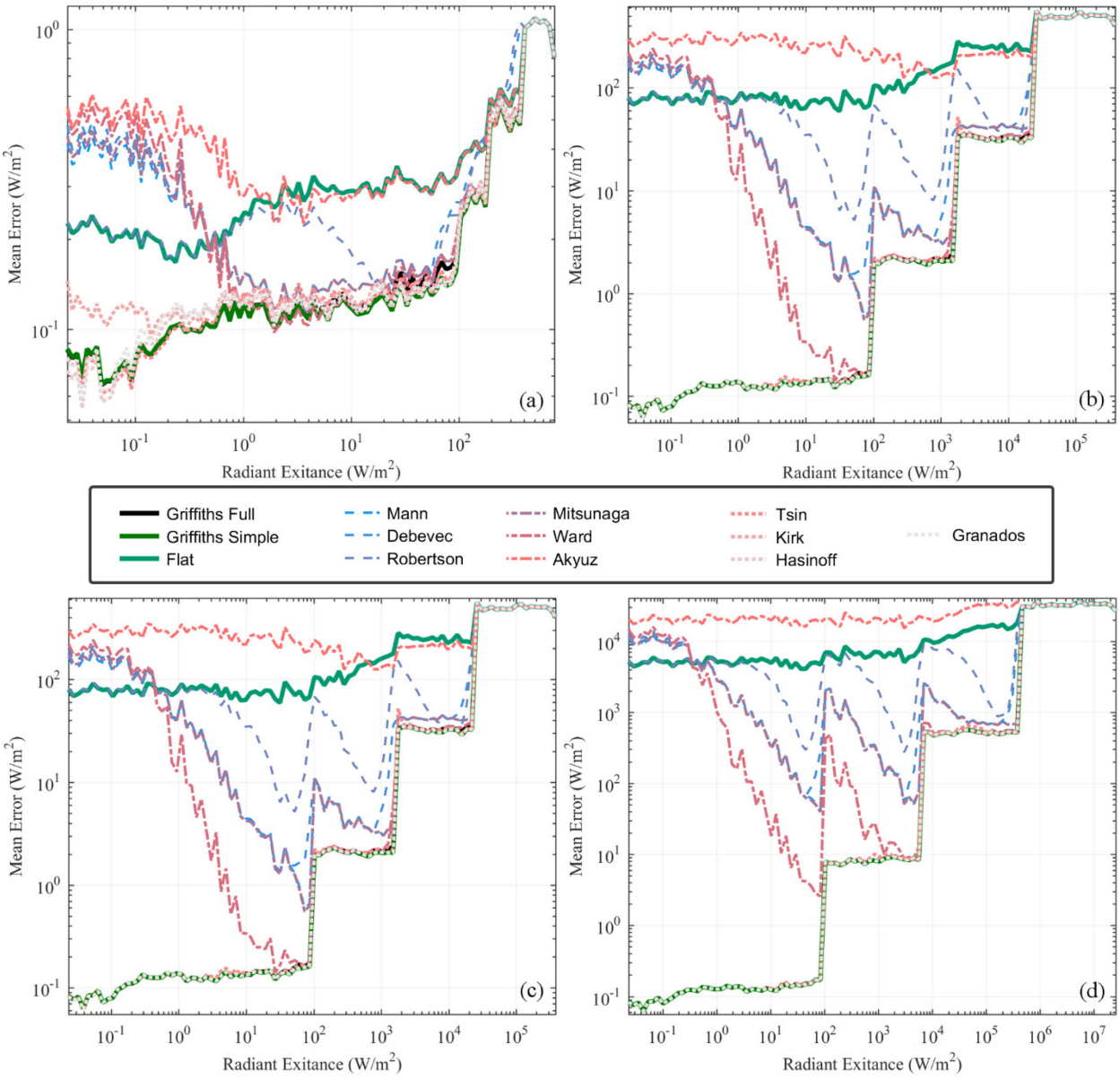


Figure 2-3. Mean error for (a) 1 stop spacing (b) 2 stop spacing (c) 4 stop spacing (d) 6 stop spacing.

The top left figure, Figure 2-3 (a), is a comparison of the mean error for all 13 cases spanning the full 90 dB dynamic range of the simulated system. The top right line graph, Figure 2-3 (b), spans 108 dB dynamic range when using the simulated two stop system. The bottom left line graph, Figure 2-3 (c), spans 144 dB dynamic range, created with a four stop spacing system. The bottom right figure, Figure 2-3 (d), spans an extreme dynamic range of 180 dB, or 30 stops, utilizing a six stop spacing between cameras.

Throughout the simulated range the Granados and Hasinoff weighting function exhibited the lowest error when compared to the other weighting functions. The weighting methods that

introduced the least amount of error were proportional to the error inherent in the acquisition of the imaging system and not a function of the acquired pixel value, observed by an almost identical trace on Figure 2-3. In the low radiant exitance range, the images or frames with the greatest amount of attenuation severely degrade the estimate on the final image formation, the same phenomena that was observed in Figure 1-5 in the introduction. This error addition is caused by the read and dark current noise from the low level output of the highly attenuated exposures being given too much weight in the image formation process and degrading the final image formation.

When the weighting function is only proportional to the digital value or the slope of the camera response function, such as in Ward, Robertson, Mitsunaga, and Debevec, the magnification of error becomes more significant. When the attenuation level increases, the error becomes orders of magnitude greater in the lower exitance ranges when not using an optimized weighting function seen in the six stop spacing data. To further communicate this, Figure 2-4 demonstrates the difference in the signal to noise ratio of the various weighting methods to the Granados weighting function.

This figure was produced by simulating the radiant exitance image with 10,000 logarithmically spaced mean radiant exitance value columns. Negative values indicate lower signal to noise ratios in the resultant image when compared to the Granados weighting function, and a positive value would indicate better performance. In all cases, the flat weighting function, Mann, Debevec, Robertson, Mitsunaga, War and Akyuz methods underperformed relative to the Granados function. When only a small amount of attenuation is present, the Kirk weighting function slightly underperforms the Granados weighting function in the low radiant exitance range due to the lack of the inclusion of the dark current noise source.

For all cases the proposed weighting function reduced the error to within a maximum deviation of 2.3 dB of the Granados weighting function across a dynamic range set spanning 90 dB to 180 dB. This demonstrates how the proposed non-recursive weighting function performs relative to the recursive weighting functions, effectively reducing the noise in the image without the need for recursive iterations. The total error seen in Figure 2-3 very closely tracks the measured error by the Grandos and Hasinoff weighting functions. The largest discrepancy between the proposed weighting function and the Granados weighting function can be seen when there is very little attenuation, Figure 2-4 (A). In this case the proposed method has a maximum deviation of 2.15 dB for the full weighting and 2.22 dB for the simplified weighting function

proposed in (2-27). With a 2 stop spacing the maximum deviation was 2.33 dB and 1.25 dB, 1.95 dB and 0.31 dB for a 4 stop spacing, and 0.53 dB and 0.08 dB for a 6 stop spacing respectively.

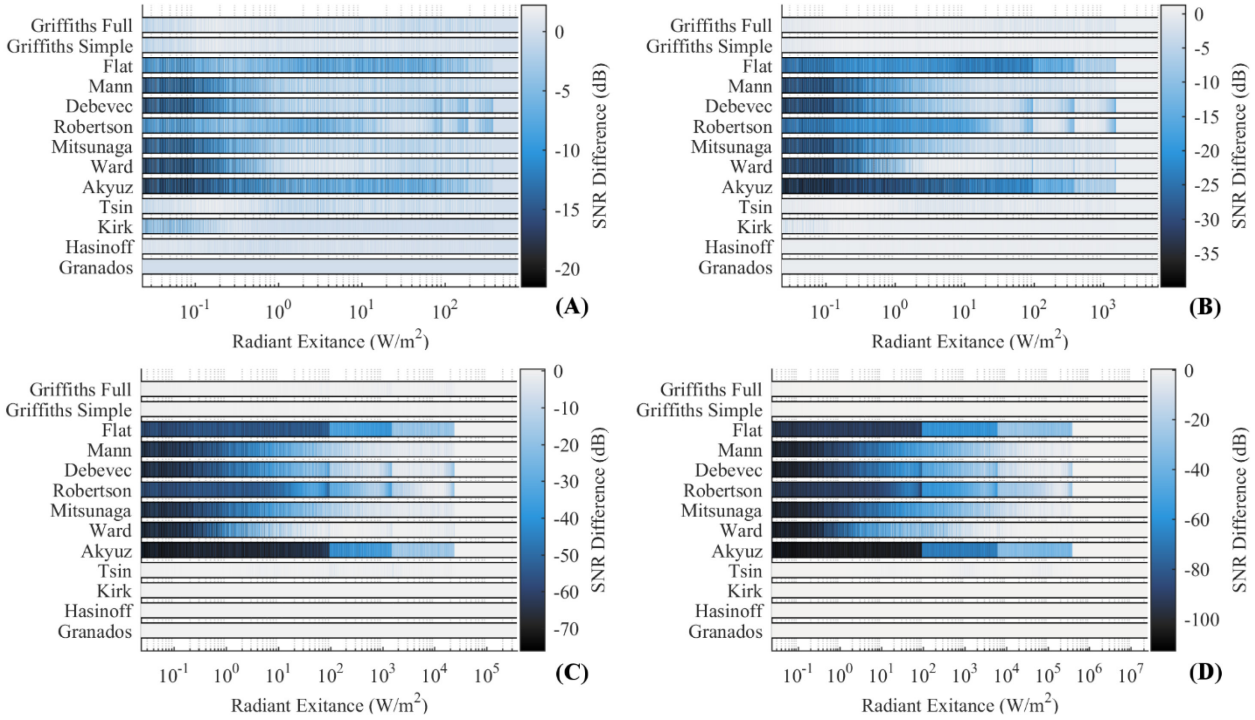


Figure 2-4. Ribbon chart depicting the amplitude of the difference of the signal to noise ratio in dB from the indicated weighting function to the Granados weighting function for; (A) 1 stop spacing with four Vision Research Phantom 711 cameras (B) 2 stop spacing (C) 4 stop spacing (d) 6 stop spacing

Conclusion

Multiple weighting methods were compared utilizing synthetically created images that simulate the output from a series of Vision Research Phantom 711 high speed video cameras. The images were created from a simplified camera model with additive noise. The noise parameters that were included consist of shot noise, read noise, dark current noise, and quantization error. It was assumed there is no fixed pattern noise in the synthetic camera and reset noise is mitigated in the electronics at the sensor level. The shot noise and signal level were assumed to be from a monochrome source at logarithmically varying predefined radiant exitance levels that would span the total range of the designed acquisition system. The mean error was calculated and compared utilizing a variety of different weighting functions and dynamic ranges spanning from 90 dB to 180 dB. It was found that the most effective weighting methods were proportional to the reciprocal of the variance of the total uncertainty introduced by

the acquisition system. The proposed weighting methods both were shown to effectively mitigate the error in the formation of an HDR image without the need of a recursive scheme to estimate the actual radiant exitance with only a slight error cost. It was also shown that a simplified weighting function can effectively be used to reduce the error on formation with knowledge of only the physical attenuation to the sensors, integration time, and sensor gain, greatly simplifying its application.

The final demonstration of the effectiveness of the simplified weighting function can be seen in Figure 2-5. On the left hand side of the figure, the original HDR bulb that was shown in Figure 1-6 can be viewed. On the right hand side is the identical bulb, created using the identical formation method, but utilizing the new weighting function. The difference in the baseline noise becomes apparent as a reduction in the random scattering of colors around the bulb. However there are still errors that are caused by flare degradation and misalignment. A flare degradation can be seen surrounding the bulb and the misalignment error can be seen in the color patches above the bulb.

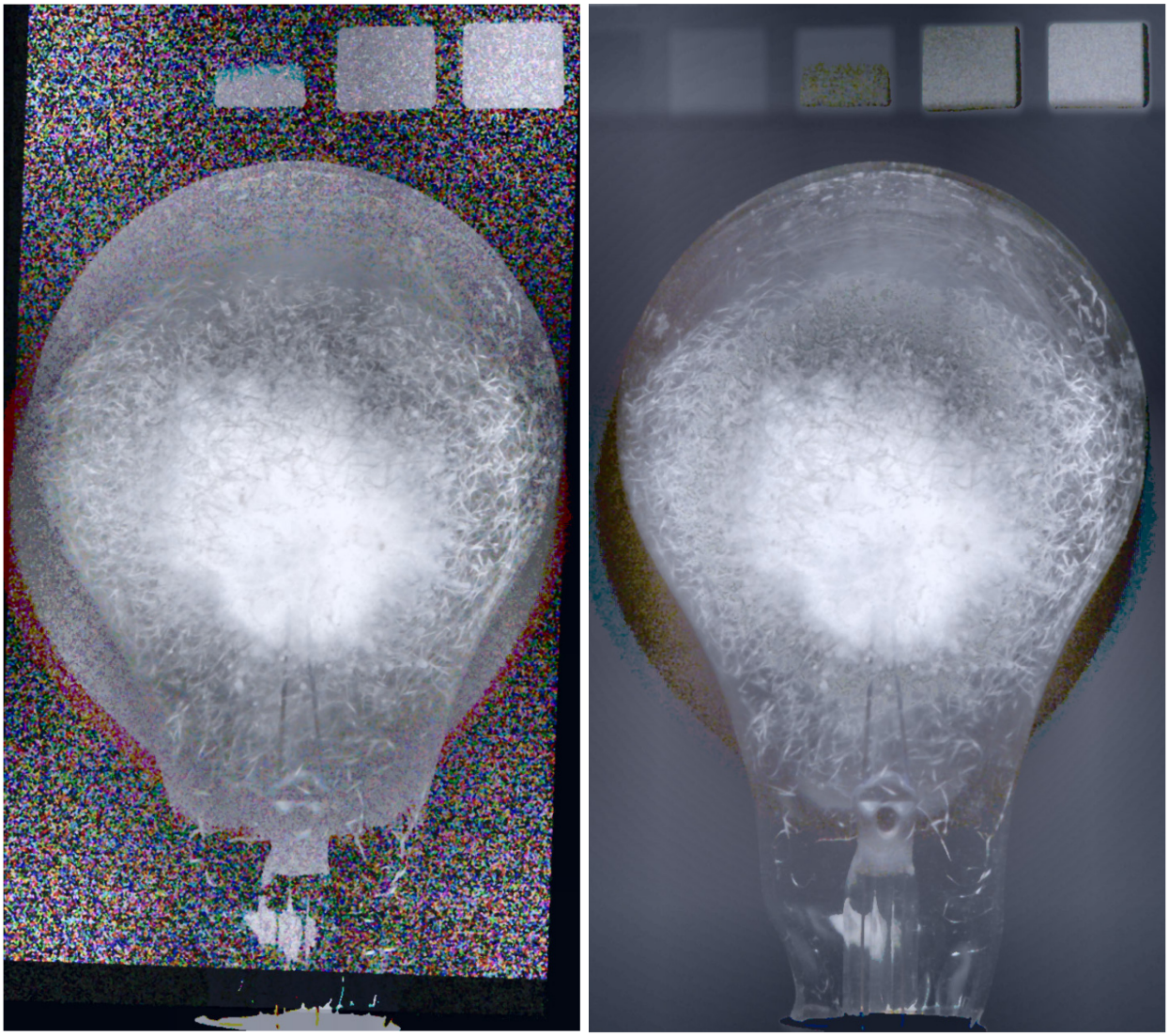


Figure 2-5. Left. Original bulb with standard formation method and hat weighting function. Right. Same formation method, using newly introduced weighting function. The baseline noise is drastically reduced, but the flare and misalignment error is still present.

Chapter 3. The Error Minimization Method

High dynamic range imagery has been demonstrated as a methodology of multiple image fusion to surpass the dynamic range limit of a single sensor. Utilizing an optimal weighing function, the symmetrically distributed errors associated with the image formation, given enough samples, tend to the expected value. However, in the presence of non-symmetrically distrusted error, image degradation, in the form of a bias error, occurs from the uncorrelated information.

Uncorrelated noise sources can arise from the source content or the imaging system. The source content must be assumed temporally invariant over the duration of the exposures, if they are not acquired in parallel, sometimes leading to ghosting artifacts. The radiant exitance estimation uncertainty stems from multiple sources such as baseline sensor noise, image misalignment, uncorrelated lens distortion from multi imager and lens systems and veiling glare. The non-symmetrically distributed errors have been shown to limit the total dynamic range that a series of images can reproduce in a given image. Introduced in this chapter is a stochastic approach to help reduce these artifacts in the presence of a minimal number of samples, increasing the effective contrast within a high dynamic range image without significant image degradation.

Introduction

High dynamic range imagery is the fusion of several images exposed with different exposure values to record the different estimated radiant exitance levels within a given scene. Typical high dynamic range imagery is formed utilizing exposure bracketing. Exposure bracketing is the process of changing the exposure value of an image capture to acquire both the highlights and the shadow detail in an image sequence by increasing and decreasing the exposure settings.

The extension to high dynamic range video from still photograph was natural, however presented new challenges. When viewing scenes that are not static, ghost artifacts appear in the reconstructed image from the dissimilar scene in each exposure. Several different methods and architectures have been developed to acquire the frames synchronously [47, 48] to avoid the ghosting problem. However, each method adds optical components to the imaging chain, adding to image degradation. Image degradation mainly takes on two forms of error; a misalignment

error due to the multi-imager's alignment, and added veiling glare due to the added intraocular reflections and scattering [49].

The simplest form of a synchronous high dynamic range system will comprise of a plurality of image sensors and matching beam splitters, connected to a lens or multiple lenses aligned along an optical axis, as discussed in [48, 50], demonstrated in Figure 3-1. An alternative configuration consists of a plurality of sensors all viewing the same scene aligned, similar to a stereo vision system [47, 51] then register and morph the pertinent information to avoid parallax. The third configuration utilizes a single sensor and requires a varying intensity masking to capture the scene [52].

The first and second configuration will have to realign the image sensors mechanically [53]. Due to the size of a pixel and the tolerance required to setup the camera system, it is reasonable to assume there will be some level of misalignment due to manufacturing tolerances. The severity of the misalignment is typically bound by the modulation transfer function of the optical system. The modulation transfer function, defined as the square of the modulus of the Fourier transform of the point spread function [54], defines the limit of the spatial frequency response of the camera system. There are two primary ways of handling misalignment error; treat it as a ghost image for removal, or employ a super resolution methodology [50] to use all of the information at each spatial location for an estimate of a new image projected on an artificial sensor.

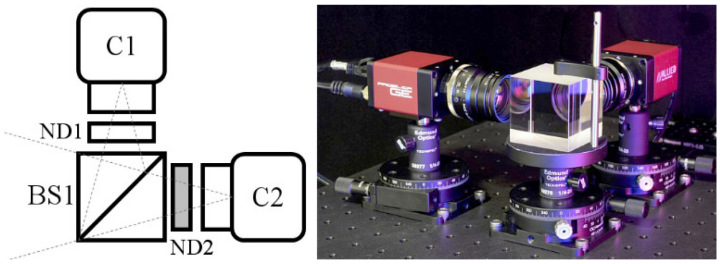


Figure 3-1. Left. Schematic of a two camera HDR video system comprised of two cameras, C1 and C2, two neutral density filters, ND1 and ND 2 and a beam splitter. Right. Implementation of the schematic with two Prosilica GE cameras and 40mm prime lens.

In this work, all non-symmetrically distributed errors will be treated as a pseudo ghost image since there is no temporal change; there is only a spatial deviation from misalignment or degradation from veiling glare. There is a significant body of literature describing the identification and mitigation of ghosting in high dynamic range images [55, 56, 57, 58, 59]; an excellent review has been assembled by Srikantha [60].

The majority of these methods develop a ghost map and negate the pixels that are considered a ghost image relative to a reference image. This process works well when there are a significant number of samples, however Srikantha indicates only a few methods work in the presence of a minimal number of samples, defined as less than five. In the spirit of minimizing the number of number of sensors needed for synchronous HDR video or image capture, the number of sensors must be reduced and the exposure difference maximized. This condition leads to very few samples, typically two, in which there is correlated data available to detect image abnormalities, reducing prior methods effectiveness.

It will be shown in Chapter 5 that the total dynamic range is limited to the number of quantization levels, assuming the deep well capacity of the sensing element is greater than the number of quantization levels. When the dynamic range is significantly increased for a limited number of cameras, the maximum spatial frequency attainable by a single camera becomes limited by the glare spread function [61].

The first source of non-symmetrical distributed error in creating a multi imager based camera is the accurate alignment of the sensors. Aggarwal *et al* [48] were the first to address this problem with a split aperture system. Tocci *et al* [47] developed a multi imager video system where they had to find a method to handle pixel misalignment. A new method has been developed to assume the imagers are misaligned and use that misalignment for the development of a super resolution image [50]. The proposed high speed high dynamic range system will have similar challenges regarding pixel misalignment. The system is composed of current off the shelf components that are built to a specific tolerance. The sensor positioning for each of the high speed cameras also varies relative to the manufacturing tolerances of the camera and the mounting and alignment system. Due to the many variations in the manufactured parts and optics, any system that is created would have to be aligned as close as possible mechanically then refined through software. The high speed high dynamic range system will be demonstrated in an environment that is subject to significant physical perturbations, such as blast waves and debris. It is therefore reasonable to assume that the difficulty in perfectly aligning the cameras and keeping them aligned throughout the test would be impractical, and a method to handle misalignment should be developed.

The second source of non-symmetrical distributed error stems from the glare spread function of the system, which describes the scatter of light from the intraocular reflections and scattering from the surface texture of the lensing elements [54]. The medical community has

been developing methods to remove veiling glare for improved contrast in X-Ray images [62, 63]. These methods typically involve the characterization of the glare spread function, then removal via a blind deconvolution. This method is computationally effective, however assumes the glare spread function is isoplanatic. Harvey [54] shows this assumption is invalid, and the scattering is dependent on both the angle of the light incident to the lensing element as well as the location of the light on the lens relative to the optical axis.

Reinhard [28] describes a glare removal method where the glare spread function is estimated as a third order radially symmetric polynomial. A glare image is produced from the estimated glare spread function and acquired radiance information. The estimated glare magnitude is subsequently subtracted from the estimated scene radiance to form the final high dynamic range image. The informational content caused by glare tends to exhibit itself as a low pass filter, however as Harvey has demonstrated, can be very complex to model in its entirety.

The glare is caused by either dirt on the lens causing a scattering of the light or by internal reflections of the lens components. A modern telephoto lens can comprise of over 20 individual glass elements. Every time there is an impedance mismatch between elements there will be some level of reflectance and transmittance of the incident light. Therefore a lens consisting of many elements is subject to added flaring compared to a lens composed of less elements. The angle of the bright light source relative to the optical axis will also alter how the lens flaring will appear. The shape of the flare will also be altered by the number of blades used in the aperture [64].

When synthetically rendering a flare, the magnitude of the light source will be known; the geometry of the optics, spacing of lens elements and construction of the aperture can be simulated to form the flare image. With the development of a high dynamic range image or video frame, the high dynamic range image would have to be directly measured without degradation to know where the bright source originates from. Using the current methodology to form the high dynamic range image, the degradation will already be present making this a challenging task with a limited number of samples. Reinhard's approach proves to be an effective fix for simple glare aberrations since it is a measure of the spatial frequency of the glare degradation. Then using an inverse filter reduces the identified spatial frequency in the image whether it is associated with glare or not.

Talvala [65] introduced a glare reduction method by using a mask in front of the lens. The masked areas could then be used to estimate the glare contribution to the final image. This method proved to be effective, however is not currently practical to use with a video system.

Due to the spatial and angular dependency of the glaring, it becomes improbable to calculate a new glare spread function on a per frame basis for removal, particularly in the presence of image misalignment. Instead a new method of statistical glare and misalignment degradation removal is proposed by purposefully introducing different glares at each sensor by using different lensing. This will force an uncorrelated glare spread function on each image. The uncorrelated information will have to be identified, and then can subsequently be removed. This is the foundation of the proposed formation method entitled the error minimization method.

The Error Minimization Method

Image misalignment and veiling glare will both be treated as a pseudo ghost image in the HDR video frame formation. Described in the introduction, the error tolerance from misalignment is highly dependent on the modulation transfer function of the system. If the spatial resolution is limited due to diffraction or optical scattering then there will be a statistically insignificant difference in the recorded radiant exitance for a slight offset of the camera system, even in the presence of a step function. Some misalignment can be statistically irrelevant so long as the misalignment is less than a statistically discernable frequency defined by the modulation transfer function of the system.

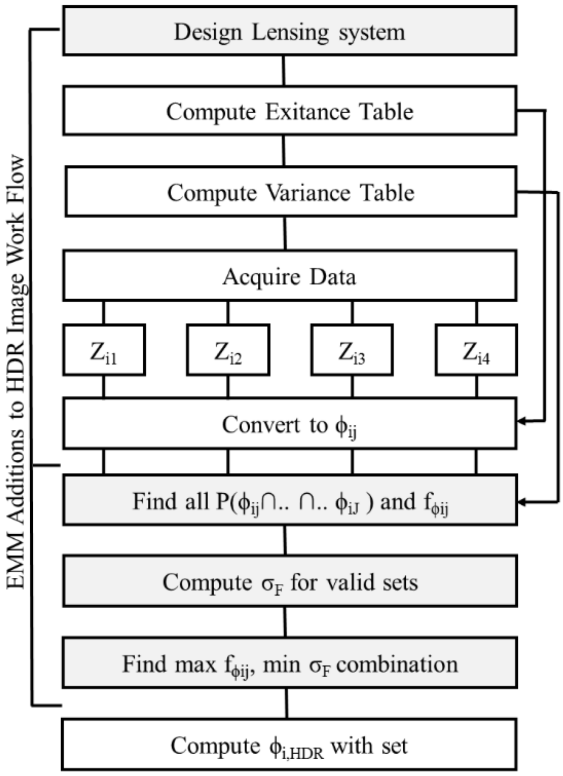


Figure 3-2. Flowchart of the proposed error minimization method. The grayed areas are the added steps when compared to the standard formation process.

In this chapter a camera uncertainty model will be used to estimate what range of possible mean radiant exitance values could have resulted in each imager's response. Then the probability of the actual mean being confined to a range determined by other cameras out to a predefined confidence interval is calculated. If there is a probability of intersection between the various sets of imagers, the imager set that has the highest probability of originating from the same source with the most samples and the lowest error will be used for the final image formation. An overview of the process can be seen in Figure 3-2.

Introduction to the Error minimization method

The error minimization method (EMM) is a procedure to detect statistical abnormalities with a limited number of samples based on known prior distributions. The procedure detects bias errors inconsistent with known zero mean stochastic errors originating in the image acquisition chain. In doing so, the procedure can prevent the inclusion of data points that have degraded information originating by either misalignment or glare.

The first step is to carefully choose what lensing is desired for the HDR system. The EMM will work regardless of lensing, however it will be shown to be more effective if dissimilar lenses are utilized later in this chapter. The dissimilar lensing makes veiling glare dissimilarities more prevalent in the various captures allowing the glare degradations to present themselves as large bias errors and subsequently ignored. A lookup table can be produced for every pixel value that will be acquired. The error estimate for a given pixel can be computed using (3-9). The reduction in error associated with acquiring the pixel information can be calculated using (3-21). The data is acquired with the capture system and converted using the lookup tables found prior to capture.

The exitance estimate and associated error bounds are then used to find the span of radiance exitance values that could have instilled the acquired value out to some confidence interval. A tighter confidence interval, for example 68%, will result in less potential camera sets that can be used for the final formation, and less inclusion of potentially dissimilar data. This has a tradeoff of reducing bias error but increasing the uncertainty by not including more samples in the final formation. First the uncertainty associated with image formation will be reviewed and the process for forming the exitance and radiance tables discussed.

Creating the Exitance and Variance table.

The image formation process was described in depth in 0, therefore only the resultant equations that are pertinent to this chapter will be reviewed. Recalling from (2-5), the radiant exitance, $\Phi_{ij,}$ using the linear camera model can be defined as

$$\Phi_{ij} = \frac{Z_{ij} N_{well,j} hc}{q_s A_j \Delta t_j A_{phy,j} \eta_j \lambda_i g_j} \qquad 3\text{-}1$$

where Z_{ij} is the digital value, N_{well} is the deep well capacity, hc is the product of Plank's constant and the speed of light, q_s is the number of quantization steps, A is the active pixel area, Δt is the exposure time, η is the quantum efficiency, λ is the wavelength, and A_{phy} is defined as

$$A_{phy} = \frac{\pi \tau_{optics}}{4\left(f/_{\#}\right)^2} \qquad 3\text{-}2$$

where A_{phy} represents the optical attenuation to each camera. Following the generally utilized literature nomenclature the camera response function, $f^{-1}(Z_{ij})$, can be defined as

$$f^{-1}(Z_{ij}) = \frac{Z_{ij} hc N_{well,j}}{A_j \eta_{ij} q_s \lambda_i} \quad 3\text{-}3$$

and is in units of Joules/m^2. The radiant exitance in (3-1) using (3-3) can be rewritten as

$$\Phi_{ij} = \frac{f^{-1}(Z_{ij})}{\Delta t_j A_{phy,j} g_j} \quad 3\text{-}4$$

The error from the read noise and dark current noise, σ_{ADU}, can be modeled as

$$\sigma_{ADU} = \frac{f^{-1}(Z_{ij})}{Z_{ij} A_{phy,j} \Delta t_j g_j} \sqrt{{\mu_{rd,j}}^2 + {\mu_{dc,j}}^2} \quad 3\text{-}5$$

where μ_{rd} and μ_{dc} are the read and dark noise standard deviation, respectively, in analog to digital units. The photon shot noise is an arrival process and modeled by a Poisson distribution. The photon shot noise at the sensor site can be estimated by

$$\sigma_{ss,ij} = \frac{hc}{\lambda \Delta t_j A A_{phy,ij}} \sqrt{\frac{Z_{ij} N_{well}}{\eta g q_s}} \quad 3\text{-}6$$

Utilizing the camera response nomenclature (3-6) can be rewritten as

$$\sigma_{ss,ij} = \frac{1}{\Delta t_j A_{phy,ij}} \sqrt{\frac{f^{-1}(Z_{ij}) hc}{g \lambda A}} \quad 3\text{-}7$$

The quantization error is the error that arises from the discrete nature of the sampling mechanism. The quantization error, σ_q, can be modeled by a uniform distribution, and translated into an error on the radiant estimate by

$$\sigma_q = \frac{1}{\sqrt{12}} \frac{f^{-1}(Z_{ij})}{Z_{ij} A_{phy,j} \Delta t_j g_j} \quad 3\text{-}8$$

The total error, σ_t, on the estimate of the radiant exitance can be found by summing the independent sources in quadrature.

$$\sigma_t = \sqrt{{\sigma_{ss}}^2 + {\sigma_{ADU}}^2 + {\sigma_q}^2} \quad 3\text{-}9$$

In order to combine the images a software based alignment is applied to all of the images. There are a number of ways to perform image alignment for HDR images. The most common is the median threshold alignment method due to its resilience in false alignment in the presence of amplitude differences [44]. Another method is feature matching utilizing a scale invariant feature transform to determine the proper morphology. This method requires similar features be seen in the various exposures, therefore exposures of extreme difference may not align properly. In the extreme cases a manual method may have to be used if other methods fail. The alignment information can be found by running a calibration on the camera system after it is setup. The alignment data is stored in matrix form as an affine transform to pan, zoom, rotate, and shear the images as needed. Next, the exitance table and associated variance table can be made.

The process to perform this task is graphically depicted in Figure 3-3. First, it must be determined if the camera has a linear response or not. If the camera does not have a linear response, a method such as Debevec's can be used to linearize the data. A scaling parameter will have to be used if the output high dynamic range image will have radiometric units. In general, it will not affect the formation of the image if the data is not in radiometric units, however some tone mapping methods need radiometric or photometric units for proper display remapping. If the camera is a linear camera then the exitance table can be formed using knowledge of the camera and optical components in the optical train. Once the camera response function is known the total variance can be found for each Wycoff Layer.

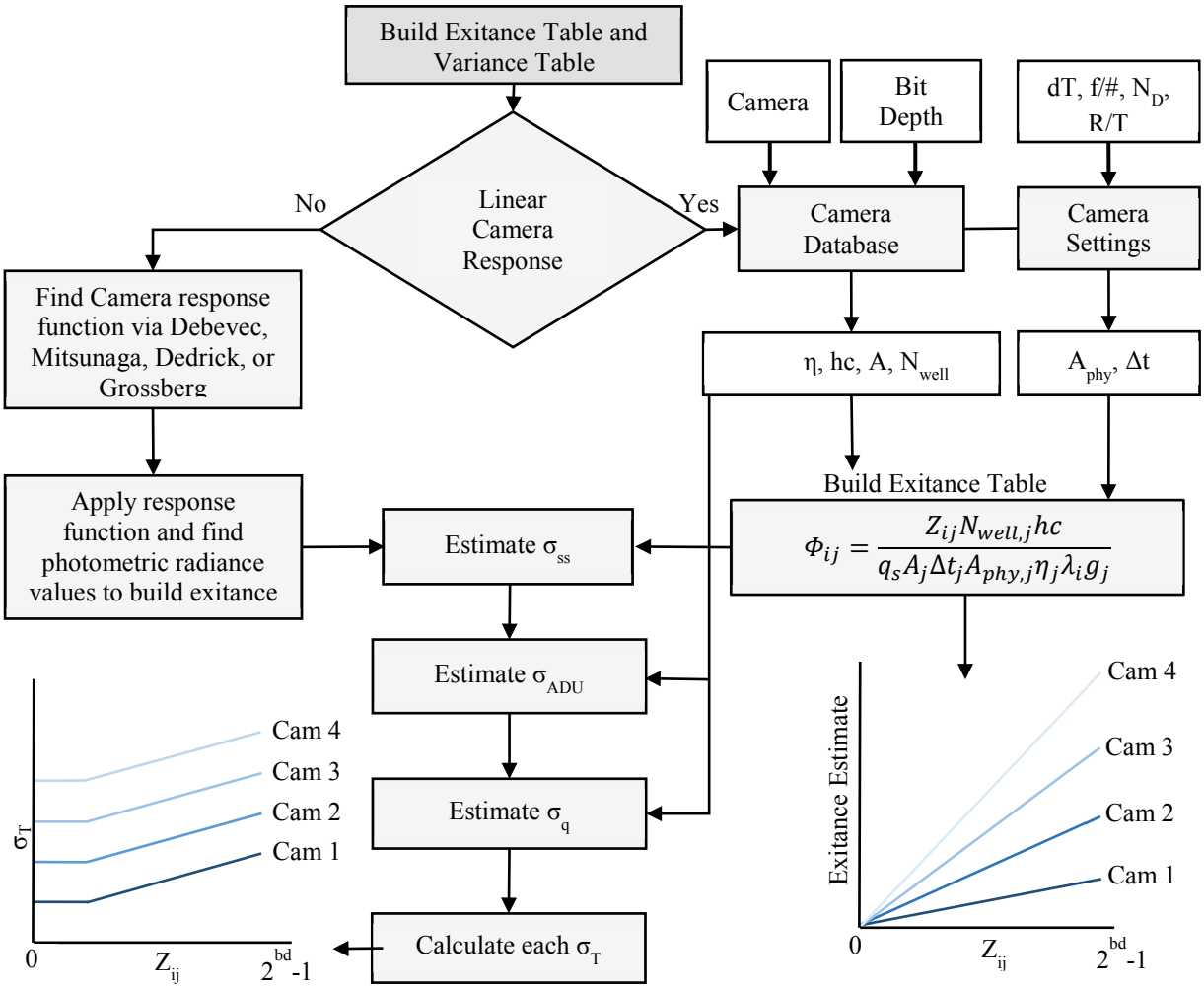


Figure 3-3. Process for forming the exitance and variance tables.

Joint Probability estimation

In order to statistically determine if a given point is an outlier within a data set there needs to be enough samples of the same phenomena to define the mean value and a confidence interval in which the points lies. For a given number of data points [66], an outlier cannot be defined outside of ξ standard deviations

$$\xi < \sqrt{n + \frac{1}{n} - 2} \qquad 3\text{-}10$$

where n is the number of samples in the data set due to the lack of data. In the case of two images with viable data, the value of ξ is 0.71, for three cameras the value of ξ is 1.15 and four cameras the value of ξ is 1.5 standard deviations from the mean. This means for 2 cameras a

confidence interval greater than 52% and 3 cameras a confidence interval greater than 75% cannot be achieved due to the lack of samples. To work around the problem of a limited population size, the image formation process is considered.

The image acquisition process begins with a photon flux, where the total number of photons collected at a pixel site can vary for any given increment in time. The total number of photons at each site in the camera system will vary for a given scene photon flux due to the attenuators in the system, therefore the uncertainty on the exitance estimate will vary for each sensor. An estimate of the probability of the acquired data points stemming from the same source must be derived. Comparing the probability density function associated with each estimate reveals the total probably of intersection between the estimates, reveling an indication if the values could have come from the same source, or if one or more of the values is corrupt by non-symmetrically distributed error not accounted for in (3-9).

First assuming that each sensor's measurement value, Z_{ij}, is independent, the probability of intersection for some set of imagers J, and an imager j in that set, is given by

$$P\left(Z_{ij} \cap .. Z_{iJ} \forall Z_{ij} \in J\right) = \prod_{j \in J} \mathrm{P}(\mathrm{Z}_{ij}) \qquad 3\text{-}11$$

Since the primary error sources asymptotically approach a normal probability density function under a many photon assumption, the probability of intersection using the standard normal deviation given by (3-9) can be defined by

$$P\left(\phi_{ij} \cap .. \phi_{ij+1} \, \forall \phi_{ij} \in J\right) = \prod_{j \in J} \frac{1}{\sigma_{t,ij}\sqrt{2\pi}} exp\left[-\frac{\left(\phi - \overline{\phi_{iJ}}\right)^2}{2\sigma_{t,ij}}\right] \qquad 3\text{-}12$$

where ϕ is a range of all radiant exitance values, and $\overline{\phi_{iJ}}$ is the true mean radiant exitance estimate from the acquired digital value Z_{ij}. The probability of the observed values, defined in (3-12), originate from the same source assuming the observed mean at the sensor is the same as the actual mean predicted from the scene. However, a problem arises when the observed value is not the expected value, and (3-12) is no longer valid.

To work around this problem the process of formation is considered, and the probability of the mean being contained within a range is calculated. It is known a stimulus, ϕ_{ij}, caused the current recorded value. Therefore the range of exitance values that could have created the recorded value can be defined by

$$\phi_{ij} \in [\phi_{ij} - c_i\sigma_{t,ij}, \phi_{ij} + c_i\sigma_{t,ij}] \quad 3\text{-}13$$

where c_i defines in standard normal units the confidence interval the set of potential mean radiance values is allowed to fall within. It is also known, given the number of standard normal units, what the total probability of the true mean falling within the defined range is from the Z statistic. However, there is only one sample per imager that estimates the exitance level at the defined location, meaning there are zero degrees of freedom on the estimate. Due to lack of degrees of freedom, the probability the observed value is the mean value is treated as equal everywhere within the defined range in (3-13). The probability density function then takes on uniform probability density function spanning the range identified in (3-13), and the magnitude is defined by the ratio of the range width and confidence interval, defined as

$$P(\phi_{ij}|r_{c_i}) = \begin{cases} \dfrac{r_{c_i}}{2c_i\sigma_{t,ij}} & \phi_{ij} \in [\phi_{ij} - c_i\sigma_{t,ij}, \phi_{ij} + c_i\sigma_{t,ij}] \\ 0 & otherwise \end{cases} \quad 3\text{-}14$$

where r_{c_i} is the cumulative density function value associated with the number of standard normal units selected. The cumulative joint probability, $f_{\phi_{ij}}$, can be defined using some set of imagers, J, and the probability density function defined in (3-14) as

$$f_{\phi_{iJ}} = \int_0^{\phi} \prod_{j \in J} P(\phi_{ij}|r_{c_i})\, d\phi \quad 3\text{-}15$$

This concept is visually demonstrated in Figure 3-4 under varying conditions. Looking at the top left graph, (a), two exitance estimates are given by the dark blue circle and the dark gray circle. The probability density function associated with each measured value is shown as the dark line trace in either blue or gray for their respective sample estimates assuming the measured value is exactly equal to the mean value. However, with only one estimate from each sensor per unit time, it is not known if the measured value is the expected value or on one of the tails of the distributions. Therefore two other distributions are indicated with light lines of similar hue representing the extremes within the given confidence interval that the value could arise from.

The shaded region is the set of potential true mean radiant exitance values that could have resulted in the observed value. On the top left graph there are two probability density functions that overlap and indicate there is a probability that the observed values came from the same source and whose cumulative joint probability is represented by the integral of the product of the area of the overlapping shaded region. By comparing the top left, (a), and top right, (b), graphs

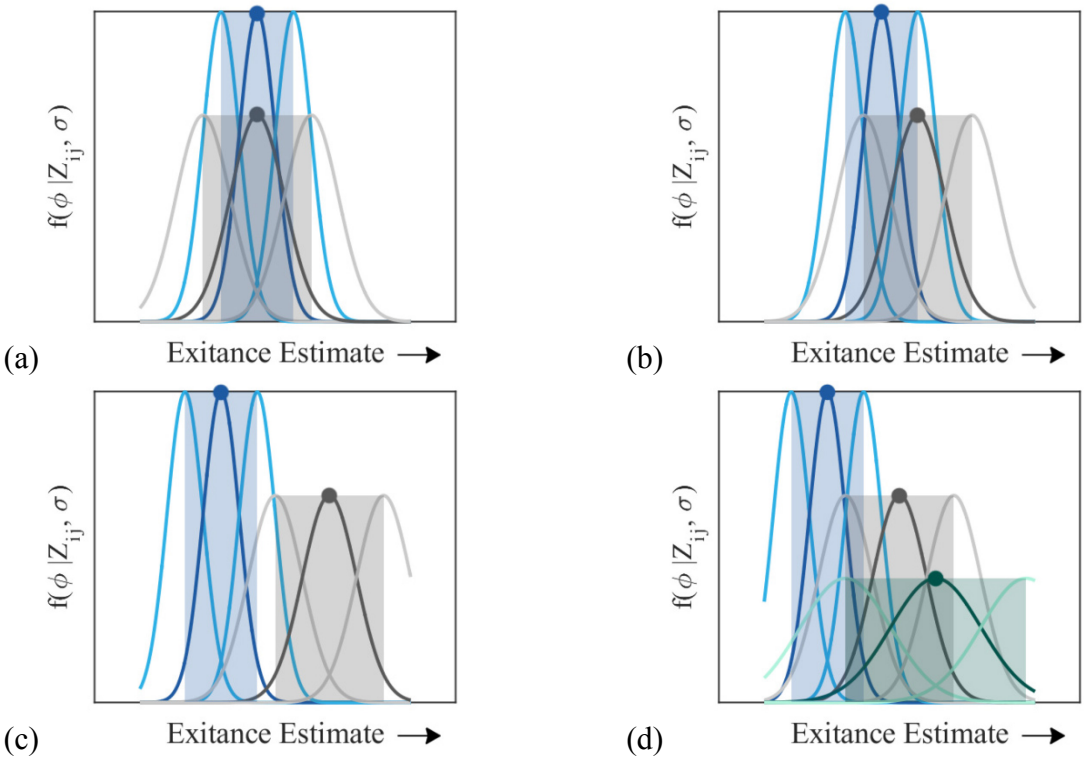


Figure 3-4. (a) Probability of intersection when observed values are identical with 95% confidence bounds. (b) Probability of intersection when observed values are different (c) No probability of intersection. (d) Probability of intersection with multiple varying observations, however small relative to the top right and top left scenarios.

the result from (3-15) would indicate there is a better chance the samples in the top left graph came from the same source than the top right graph. Comparing now the scenario in the bottom left graph, (c), there is no chance that samples could stem from the same source, therefore they should be treated as originating from different sources. The bottom right, (d), is an example how this methodology can be extended to any number of images, in this particular case three, where there is a probability, however small, that the three observations came from the same source.

Choosing the best set of cameras

The last step in determining the best camera set to use for the formation of the image is to form a rule set based on the ideology of minimizing error. The maximum likelihood estimator for normally distributed data is the arithmetic mean. Debevic [32] has shown a weighted mean should be used for the final formation of HDR images. The final radiant exitance estimate can be defined by

$$\Phi_{i,HDR} = \sum_{j \in J} w_j \Phi_{ij} \sum_{j \in J} (w_j)^{-1} \quad 3\text{-}16$$

where $\Phi_{i,HDR}$, is the HDR image. For the error minimization method, all exposures will be weighted equally until the final set of cameras is determined. Following the analysis by Ku [67] for determining the cumulative error on a measurement, first it is assumed that each measurement has an actual value and an associated error yielding

$$\Phi_{ij} = \overline{\Phi_{ij}} + \epsilon_{ij} \quad 3\text{-}17$$

where ϵ_{ij} is symmetrically distributed error and $\overline{\Phi_{ij}}$ in the actual mean. From (3-16) we can define the final radiant exitance value as

$$\Phi_{i,HDR} = f(\Phi_{iJ}) \quad 3\text{-}18$$

where Φ_{iJ} is an indicator of all the cameras used in the defined sets for pixel i. Taking the Taylor series expansion [68] of (3-18) yields

$$f(\Phi_{iJ}) = f(\overline{\Phi_{i,J}}) + \frac{\partial \Phi_{i,HDR}}{\partial \Phi_{iJ}} \epsilon_{iJ} + \frac{1}{2!} \frac{\partial^2 \Phi_{i,HDR}}{\partial {\Phi_{iJ}}^2} {\epsilon_{iJ}}^2 + \ldots \quad 3\text{-}19$$

Since the offset from the mean on the measurement is assumed to be small, only the first order term is kept. The variance can be approximated by rearranging 3-19 and by definition squaring the mean

$$mean\left(f(\Phi_{iJ}) - f(\overline{\Phi_{i,J}})\right)^2 = mean\left(\frac{\partial \Phi_{i,HDR}}{\partial \Phi_{iJ}} \epsilon_{iJ}\right)^2 \quad 3\text{-}20$$

Using 3-16, the partial derivative term is equal to one. The perturbation squared, ${\epsilon_{ij}}^2$ is defined as the variance on the estimate for each J. The total error on formation, σ_F, can be defined by

$$\sigma_F = \frac{1}{N} \sqrt{\sum_{j \in J} {\sigma_{T,j}}^2} \quad 3\text{-}21$$

where N is the number of cameras in the set of J. From (3-21) there ceases to be an error reduction with added samples when the ratio of the number of samples used between set one, N, and set two, M, when

$$\frac{M}{N} < \frac{\sum_{J2} {\sigma_{T,J2}}^2}{\sum_{J1} {\sigma_{T,J1}}^2} \quad 3\text{-}22$$

where J_1 is symbolic of the set of N imagers and J_2 is symbolic of the set of M imagers. The final task in the algorithm is to choose the camera set that minimizes the value stemming from (3-21) while maximizing the value from (3-15). This process can be done through a simple maximum and minimum search of the set of values. Once the set of imagers is chosen for the final formulation the final value for the pixel is determined using the standard weighted average, defined by (3-16).

In the event that faster processing is desired, a slightly modified process can be used. Instead of calculating the cumulative joint probability, the condition of intersection can be directly found if the range of the sets intersect. If there exists a probability that the cameras information stemmed from the same source, then those cameras are grouped together in the same pool. A discrimination can be formed using 3-21 to decide which camera set to use for final formation, however due to the weighting function the numerical error will already be reduced by including the large variance estimates. Therefore the pool with the most estimates within the same set can be used in the final formation.

In the event that no cameras values intersect, than the camera with the highest digital value from a pool that is below a saturation value and above a baseline noise threshold should be used. This event will happen when viewing something very dark relative to the brightest resolvable value or when the brightest resolvable value occurs and all other captures are at or near saturation.

Demonstration of the EMM Method

A scene was created to specifically test a scenario where there is a very bright source of light just adjacent to a comparatively dark scene to induce a significant flare bias. A halogen light source was positioned to radiate the edge of the lens to purposefully increase the veiling glare and flaring artifacts. The halogen light source, a Dolan-Jenner MI-150, was placed behind an extended Macbeth color chart diagrammatically depicted in Figure 3-5. The front face of the color chart was illuminated with a diffuse LED light array. An image was captured of the scene with a Prosilica GE 1900C GigE camera with the halogen light source turned off. The LED array's intensity was increased until the front face of the color chart was easily resolvable by the longest exposure setting. A series of ten images were recorded and averaged to record the dark frame noise. Next a series of ten images were acquired and averaged to create a low noise image reference image. The ground truth image was formed by subtracting the dark frame reference

from the low noise image and converting the resulting data to their associated radiant exitance values and subsequently their luminance values seen in Figure 3-5 spanning nearly the full 72 dB dynamic range of a single 12 bit sensor. The quantum efficiency of each color channel was used from the Kodak KAI-2093 datasheet as [0.37 0.34 0.30] at a center wavelength of [620 nm 540 nm 460 nm] for red, green, and blue color channels respectively.

The extended Macbeth color chart was placed in front of the halogen light source and care was taken to avoid reflected light sources by placing black felt lined cloth around the test area. With the color chart being optically opaque, it is assumed in future analysis any difference from the acquired images and the magnitude of the ground truth image is due only from either veiling glare or aperture diffraction based glare. From a comparison of the bias created on the color chart from the different formation methodologies conclusions can be drawn as to which method is better at reducing the bias error in the final image formation.

Demonstration of bias mitigation

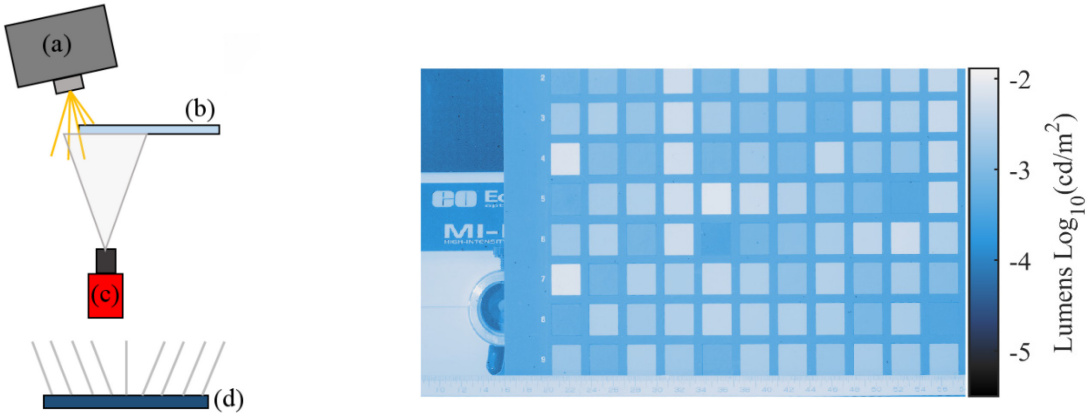


Figure 3-5. Schematic of the test setup whereby (a) is the halogen light source, (b) is the Macbeth color chart, (c) is the Prosilica camera, (d) is the diffuse LED array. Right. Estimated luminance with the halogen source turned off.

A dynamic range of 160 dB range was selected to ensure a significant flare source and tax the limit of a four camera system. At this range there will only by about a 20 dB overlap in resolvable range from each camera, severely limiting the number of samples the methodology has to utilize to recreate the scene. The spacing for each camera was designed from the optimal camera spacing that will be introduced in Chapter 5 and defined as

$$A_{opt} = \frac{1}{N-1} Log_2 \left[\frac{q_s}{10^{DR/20}} \right] \quad 3\text{-}23$$

where DR is the dynamic range desired and q_s the number of quantization levels in the sensor. From (3-23) the attenuation required from each camera setting for a 160 dB range is 4.86 stops per camera or a factor of 0.0345. The current system was utilizing a single camera, therefore a change in exposure time was used to alter the dynamic range instead of optical attenuators. Subsequently a dark frame reference was captured for each exposure setting and subtracted before the final image was formed.

The long exposure time used to form the low noise image was at 4.982 seconds using a Nikkor 55mm fixed focal length lens set to an f/# of 8. To ensure a 160 dB dynamic range the exposures utilizing (3-23) were set to 204, 5918, 171748, and 498246 microseconds respectively. The camera was set to a 204 μs exposure time and the halogen source's amplitude increased until a near saturation condition occurred on the imager at the light source. This would ensure a step of slightly less than 160 dB at the boundary of the source and color chart would exist, inciting significant flare degradation. For all exposures key points were selected and an image alignment was performed. A four by four median filter was used to remove the effect of fixed pattern noise, seen as salt and pepper noise in the images. The resulting images after alignment, filtering, and dark noise subtraction can be seen in Figure 3-6. The images in Figure 3-6 were created using Matlab's built in histogram equalization tone map functionality with 6x6 tiles, a saturation parameter of 1.2 and a fixed lightness adjustment to span the full 8 bit range from the 12 bit acquisition. Each column of images were acquired using different lenses all with an f/8. The Kodak KAI-2093 CCD that is in the Prosilica camera has an anti-blooming drain with an attenuation factor of 100x, however was not sufficient to handle the photon flux from the created scene. Therefore there is a significant blooming artifact as well, seen as a vertical bar originating from the halogen source in all four source images

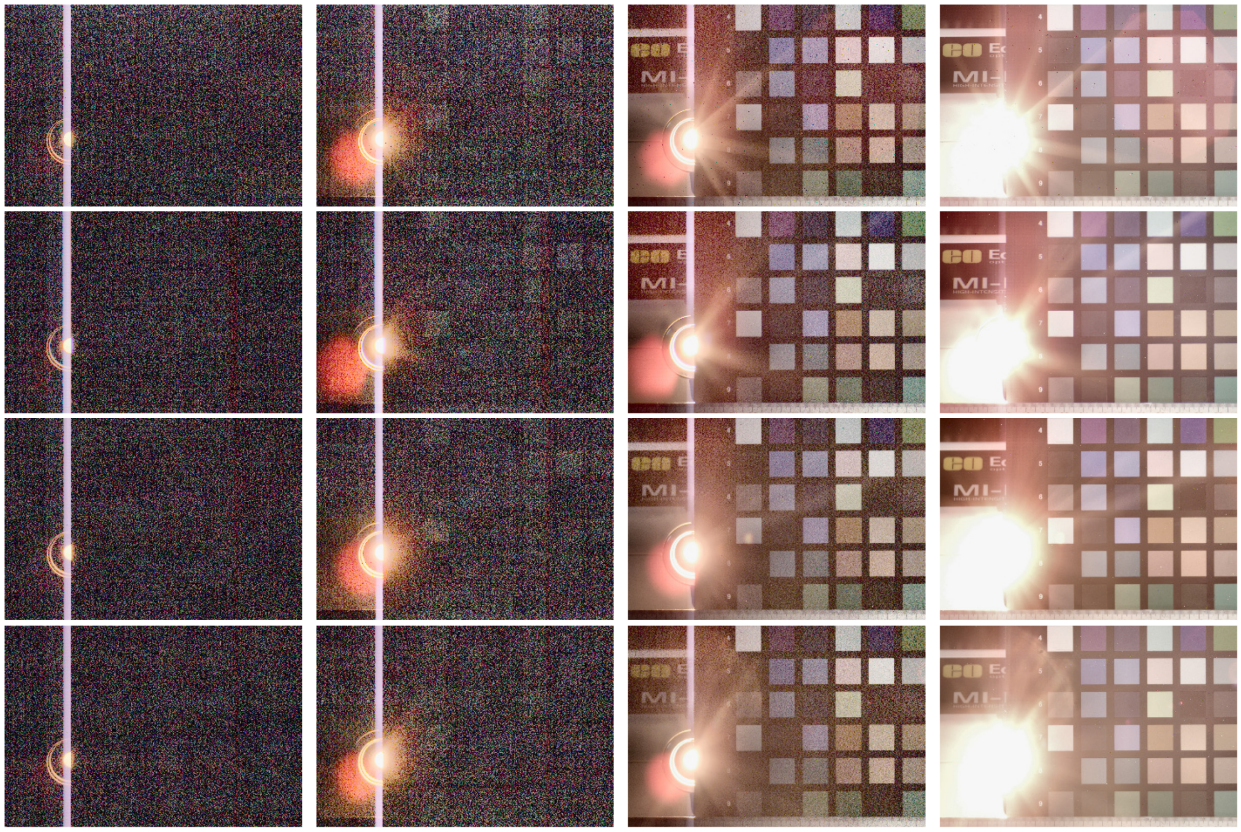


Figure 3-6. Each column has the same exposure time, however each image in each row was taken with a different lens to insight different levels of flare degradation.

The subsequent series of acquired data was formed into an HDR image using either the standard weighted average method or new error minimization method, utilizing the weighting function introduced in 0, [69], and tone mapped in an identical way as in Figure 3-6. To prevent significant bias, the extreme ranges needed to be ignored and given a near zero (10^{-30}) value weighting. The range of data that each algorithm was allowed to operate on for processing was clamped to a range of two standard deviations above the baseline noise, above the zero digital value, and 2% cutoff from the maximum bit depth value. The reconstructed data can be seen in Figure 3-7 using both methods of reconstruction for visual comparison. In Figure 3-7 top row, the Nikkor lens was used for all four exposure times. An increase in contrast can be observed in the EMM created image when compared to the standard formation image, as shown by the black bars on the color checker chart and the black label on the halogen source.

The second row of images was created using different types of lenses for each capture. In order to retain a valid comparison to the first data set, the longest exposure image used the same Nikkor fixed focal length lens. The subsequent exposures, in order from long to short, a fixed focal length Navistar 55mm lens set to an f/# of 8, an Nikon AF variable focal length 24-85 mm

lens set to 55mm an f/# 8, and a Nikon Aspherical variable 28-200mm focal length lens set to 55mm and f/8. Each of these lenses has a different number of lens elements and aperture blades, creating different flaring patterns on each exposure seen in Figure 3-6. Utilizing the varying lens system with the EMM, the effect of the aperture blade diffraction is minimized, seen in Figure 3-7 bottom right. The large red inner lens reflection seen overlaid on the color checker encompassing rows two through four and columns three through five is also removed. The red reflection on the lower left corner of the image is reduced. However even with the various lensing that feature was present in all images seen in Figure 3-6 but different sizes, which is why some, but not all of it has been corrected.

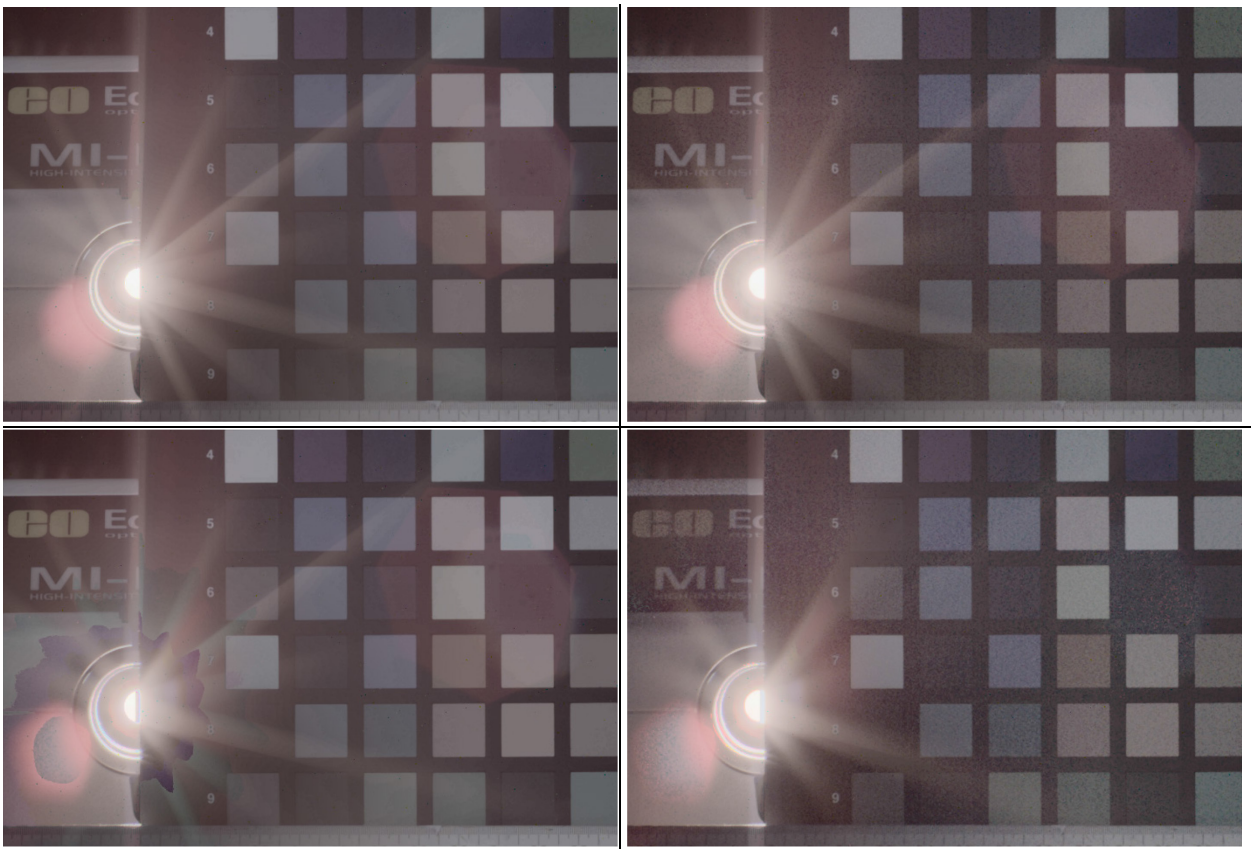


Figure 3-7. (Top Left)Tone mapped HDR image from the standard reconstruction method using the Nikkor lens only. (Top Right) Tone mapped HDR image from the EMM reconstruction method using the Nikkor lens only. (Bottom Left) Tone mapped HDR image from the standard reconstruction using 4 four different lenses at the same f/# and focal length. (Bottom Right) Tone mapped HDR image from the EMM reconstruction using 4 four different lenses at the same f/# and focal length.

Care has to be taken when choosing what lens to use; some combinations end up emphasizing the flare rather than mitigating it. Ideally the flare patterns will project out at different angles and be thin. In this sense each could be removed as they would be different for each frame.

Results

The data in Figure 3-7 and the baseline data was converted to CIE XYZ D65 illuminant color space and the luminance channel extracted and scaled. In order to compare the different formation methods the percent bias error from each luminance channel was found for each pixel. The percent bias error, ε_b, is defined by

$$\varepsilon_b = \frac{\Upsilon_{i,HDR} - \Upsilon_{i,ref}}{\Upsilon_{i,ref}} * 100 \qquad 3\text{-}24$$

where $\Upsilon_{i,HDR}$ is the estimated luminance from the multi-channel reconstructed HDR image, and $\Upsilon_{i,ref}$ is the estimated luminance from the multi-channel reference image shown in Figure 3-5. The bias error was only found for the region of the image that encompassed the color checker chart. Other regions were not used because there would have been a reflected light source degrading the comparison. The resulting percent bias error images can be seen in Figure 3-8.

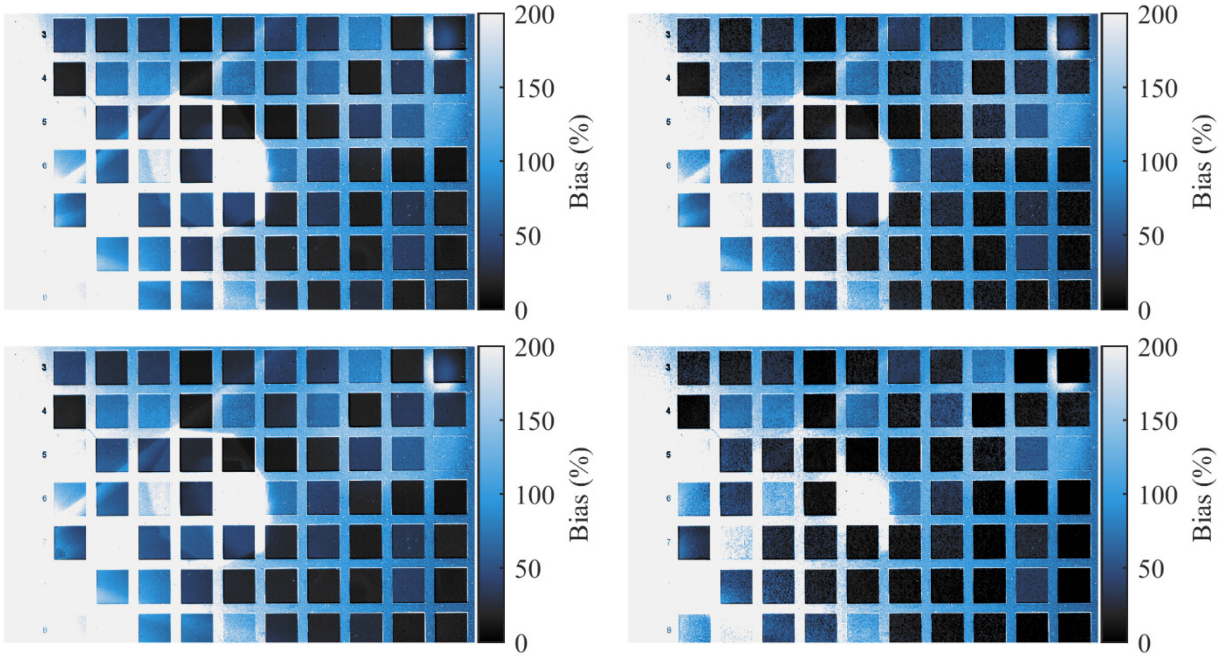


Figure 3-8. (Top Left) Percent bias error using the single lens standard formation. (Bottom Left) Percent bias error using the multi lens standard formation. (Top Right) Percent bias error using the single lens EMM formation (Bottom Right) Percent bias error using the multi lens EMM formation.

In the left column of Figure 3-8, the standard formation method bias error is shown. In the right column the error minimization method is used. The top row utilized the same lens and the bottom row utilized varying lenses. There is less error when utilizing the EMM method when compared to the standard method. The advantage of using varying lenses can be seen in the

reduction in the bias error in the regions where there is significant glare degradation. Otherwise, there is little discernable difference whether a varying lens system or a single lens system is used. This notion is further emphasized by Figure 3-9 where the percent error difference from one method to the other method is presented. Figure 3-9 indicates with a black band if the values are within ±2% of one another's bias error. Green values indicate a lower bias error for the multiple camera configuration. The top figure indicates there is some advantage to the multi lens system using the standard formation, but not very much.

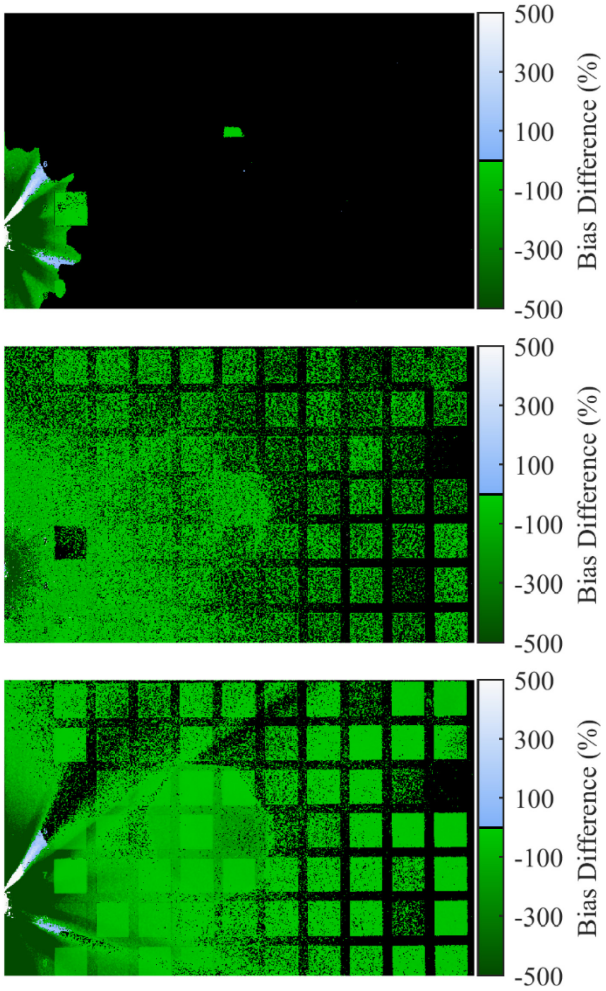


Figure 3-9. (Top) Bias difference from the single lens and multi lens standard formation method where black indicates no difference, green a value closer to the actual and light blue a value further from the actual. (Middle) Percent bias difference from the single lens standard method and single lens EMM method, green indicating less bias in the EMM. (Bottom) Bias difference from the single lens standard and multi lens EMM formed image, green indicating less bias in the EMM.

The middle row of Figure 3-9 indicates a reduction in the glare bias error across a large majority of the image when the single lens error minimization method is compared to the standard formation method. This is why the tone mapped version has more contrast then the

standard version. The bottom intensity chart shows the difference between the single lens standard method and multi lens error minimization method. This indicates significant glare reduction, over 500% near the halogen source, for the error minimization method relative to the standard formation method. The red lens flare prior discussed, can now be seen as a green hexagon in the bottom image in Figure 3-9, demonstrating the reduction of the flare bias. The bias difference plot also indicates a region where the multi lens system performed worse than the single lens system. There are two bands of added aperture diffraction bias. This occurs when the flare pattern between the various lenses are correlated, and end up adding to a higher bias instead of being reduced. Weighing the cost associated with this added error to the significant degradation elsewhere in the image is a tradeoff of the proposed method.

Conclusion

In this chapter a new method for high dynamic range image formation has been introduced. The new method identifies data within a given population with very limited number of samples that is statistically unlikely to originate from a correlated source. This process is completed by estimating *a priori* the variance associated with a measurement including the degradation in the image formation process. With this information, the probability that the data from some set of cameras originate from the same sources is then calculated. The set of cameras that has the highest probability with the lowest propagated formation error is then used to form the final image.

Several different scenarios were shown and compared utilizing both the EMM and standard formation methodology. In all cases the EMM reduced the error over a larger portion of the image than was adversely affected.

One final demonstration on the effectiveness of this method, the HDR flash bulb from the prior chapter will again be compared and can be seen in Figure 3-10. The main objective of the EMM is to remove the non-symmetric errors and this can be seen in two areas in this image. Comparing the right hand side figure with the left hand side figure shows the abrupt discontinuity that surrounds the bulb is no longer present. This error was caused by significant flaring in the high sensitivity image. Observing the color patches on the top of the image, the abrupt discontinuities surrounding the right edge have abated. This harsh discontinuity in the image was caused by an image misalignment.

The error minimization method can come at a cost of added random noise. This is an expected result since the processing routine is basically ignoring the high signal to noise ratio information in lieu of the low signal to noise ratio data due to the identified bias error and favors the low signal to noise ratio data without the bias error. However, it will be shown in the next section how tone mapping can further improve the image appearance now that all sources of error have been reduced. The information that is in the area with added noise is orders of magnitude darker relative to the bulb. The tone mapping that was used in these images is Matlab's built in tone mapping histogram equalization. It is a spatially varying mapping method that brightens all shadow detail and darkens highlight detail. In the next section a perceptual tone mapping will be introduced to improve the overall image appearance.

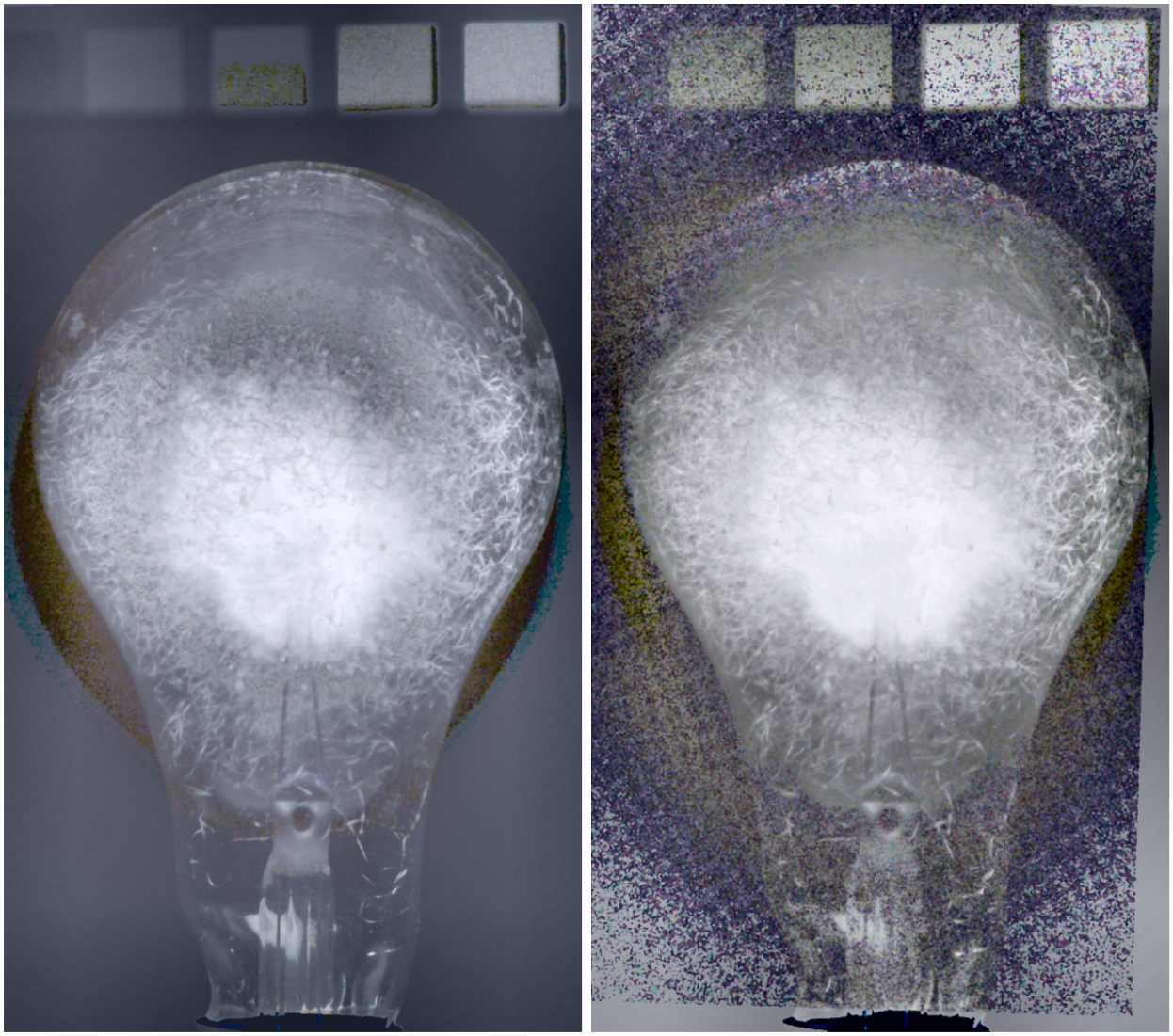


Figure 3-10. Left. Standard formation method and new weighting method. Right. Image formed using the EMM method with c_i=0.25.

Chapter 4. Tone mapping inspired by the human vision system

In this document a simplified weighting method has been introduced and a new stochastic process for the formation of HDR imagery. The next step toward being able to see what detail is within an HDR image is to tone map the HDR data into the display space of the media. The range of data collected in a high dynamic range image or video frame can exceed the dynamic range of the display media by many orders of magnitude. This makes the display of the high dynamic range data in a perceptually accurate way a challenge. Typical monitors operate in an 8 bit per channel color space. This means there are 256 shades of gray that are displayable in a scene that could have more than a million shades of gray recorded. Once the high dynamic range information is found, a method must be developed to remap the radiance values into an 8 bit display space for visualization. This remapping of the visual data is referred to as tone mapping.

In this chapter several tone mapping methodologies will be introduced as well as some colorimetric terminology. A new tone mapping methodology will be introduced based on the retinal photo transduction mechanism. The results will be compared to prior arts' methodologies and applications discussed. One difficulty when comparing tone mapping methodologies is defining which is better, as there is no real answer aside from a means tested physiological study. Therefore what is presented herein is a method that for some circumstances and users will produce a desirable effect.

Introduction

Current methods of tone mapping can be broken down into two main types each with subcategories. The two main forms of tone mapping operators are local and global operators. A local operator changes the brightness of a pixel or region of pixels based on their relative brightness when compared to its neighboring pixels. A global operator scales the brightness value of the full image based on a calculated or defined background brightness.

The subcategories break down into physiological or psychological models. The hue and brightness that the human eye acquires is not always what is perceived. A current model such as Pattanaik's [70] uses a multi-scale observer model, others such as Hood [71] use a sigmoidal compression, in either spatial variant or invariant forms. Reinhard and Devlin [72] use a modified sigmoidal operator based on the human vision system that includes chromic adaptation.

Chromic adaptation is the change in a color's appearance due to the amplitude at which it is radiating. There are other versions such as Mantiuk [73] that adapt the visual output based on the expected perceived contrast of the viewing media such as a projector or LCD. The Tumblin Rushmeier [74] method modifies the luminance map to remap the colorimetric information. Lischinski [75] developed a user based adjustment mechanism that allows for the adjustment of photographic parameters in defined luminance regions. This is similar to a traditional dodge and burn operation used in film.

Durand and Dorsey [76] introduced the use of a bilateral filter. The bilateral filter decomposes the image into a detail layer and a brightness layer, compress the brightness layer and then recombines the two. Ahan [77] developed the Retnix model, which adapts the local contrast as a function of the perceived contrast. Moroney [78] introduced the CIECAM color model to adapt the perceived color based on the background illumination. Kim [79] introduced a global algorithm that adapts to linear and logarithmic layers of brightness. This method was introduced to maintain a consistent look across many different ranges of images and is suitable for video formation. Fattal [80] introduced a method that scales the Laplacian of the image. Fattal's tone map operator basically decreases the magnitude of the gradient in the images to reduce the overall luminance range. Aydin [81] developed a stable temporal tone mapping method for videos that reduces the flicker common in tone mapped videos. Flicker is caused by slight changes in the background luminance and the subsequent drastic remapping on a per frame basis. This looks highly unnatural since the human eye cannot adapt globally that fast, and so Aydin proposed a temporal response as well as a spatial response to mitigate this problem.

In this section a tone map operator will be introduced that mimics the human visual systems adaptation mechanism. The radiant exitance estimate from the HDR image will be mapped into the estimated luminance. Then the pupil attenuation and retinal response will be estimated to get a predicted perceived brightness. The perceived brightness will be reapplied to the HDR information and color balanced in both a global and spatially varying process.

Basic modeling of the human eye in varying lighting conditions

As light first enters the cornea and passes toward the pupil, there will be some attenuation due to the dilation of the pupil. The pupil will react to the incident light and change the eyes effective aperture. The light will continue to pass through the aqueous humor, which is the fluid within the eye, until it impacts the retina. The retina is covered with millions of photoreceptive

cells called rods and cones. When the light interacts with the photosensitive cells a cascading chemical reaction takes place [82], which ultimately leads to a change in potential at the ganglion cells. The ganglion cells are connected to the nervous system and send the information onto the brain for processing.

The first step to model the human vision system is to model the pupil and how it reacts to light. The pupil responds to light by contracting and dilating. The diameter of the pupil can change from 2 mm to 8 mm on average depending on the retinal irradiance. This can account for a change in the retinal irradiance by a factor of 16. Holladay [83], Crawford [84], Moon and Spence [85], and Groot and Gebhard [86] proposed the following relations for the luminance and pupil diameter, $D(Y_b)$

Holladay $$D(Y_b) = 7\, e^{-0.1007 {Y_b}^{0.4}} \quad (4\text{-}1)$$

Crawford $$D(Y_b) = 5 - 2.2\, tanh[0.61151 + 0.447\, \mathrm{Log}_e(Y_b)] \quad (4\text{-}2)$$

Moon $$D(Y_b) = 4.9 - 3\, tanh[0.4\, Log_e(Y_b) - 0.00114] \quad (4\text{-}3)$$

Groot $$D(Y_b) = 7.175\, e^{-0.00092(7.597 + \mathrm{Log}_e(Y_b))^3} \quad (4\text{-}4)$$

where Y_b is the background luminance. A unified formula has been constructed to account for age, monocular vs. binocular vision, luminance and field of view. This formulation has been proposed by Watson and Yellot [27] and is primarily based off of the Stanley and Davies [87] data. The Stanley *et al* formula defines the effect of the pupil response to the field of view and is defined by

Stanley and Davies $$D(Y_b, a, e) = 7.75 - 5.75\left(\frac{(Y_b ae/846)^{0.41}}{(Y_b ae/846)^{0.41} + 2}\right) \quad (4\text{-}5)$$

Watson $$D(Y_b, a, y, y_0, e) = D(Y_b ae) + (y - y_0)[0.02132 - 0.00952 D(Y_b ae)] \quad (4\text{-}6)$$

where a is the area subtended in deg^2, e is the number of eyes used to view the scene, y is the age of the individual and y_0 is mean reference age. The reference age was not reported by Stanley, however Watson concluded an age of 28.58 years was most probable. The equations can be viewed graphical in Figure 4-1 with an age of 30, 60 degree field of view with binocular vision.

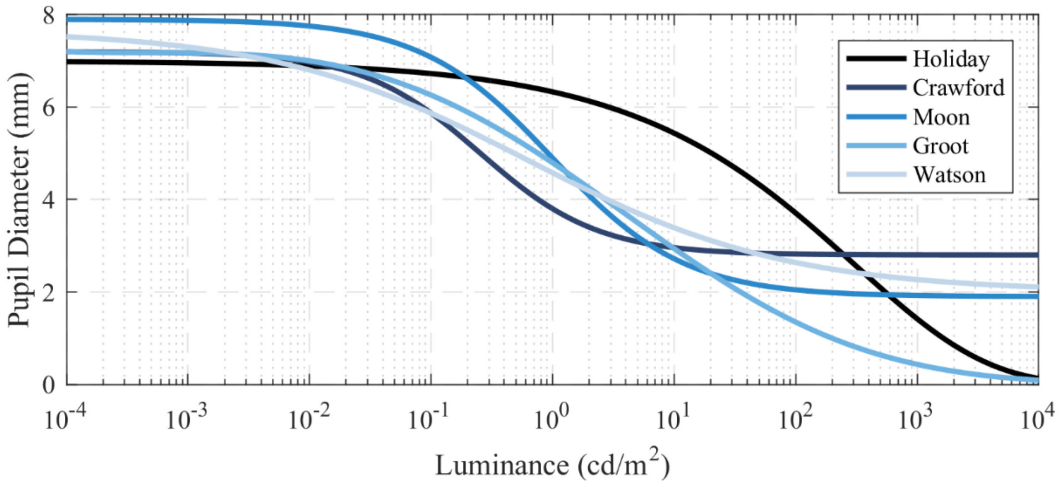


Figure 4-1. A plot showing the predicted pupil dilation based on ambient luminance.

Watson's equation is only valid for ages greater than 20 and less than 83 years old, modification need to be made for other age ranges. The corresponding retinal irradiance can be found by calculating the attenuation factor from the pupil dilation and multiplying by the pupil irradiance. The pupil acts like an aperture, therefore the attenuation is a function of the ratio of the maximum diameter to the minimum diameter. The retinal irradiance, E_r, can be defined by

$$E_r = E_p \frac{D^2}{D_{max}^2} \qquad 4\text{-}7$$

where E_p is the pupil irradiance, D is the estimated pupil diameter found from (4-6), and D_{max} is the maximum pupil dilation. This notion of attenuation is expressed on the line drawings seen in Figure 4-2. The light blue line is representative of no attenuation from the pupil as a function of luminance. The dark blue line shows the retinal Luminance after adaptation. The difference between the two lines indicates how much light is being attenuated simply from the mechanical aperture of the human eye. Observing Figure 4-2 there is a maximum attenuation of about a factor of 12, or 3.5 stops. Therefore there must be another mechanism that allows for the human vision system's total dynamic range.

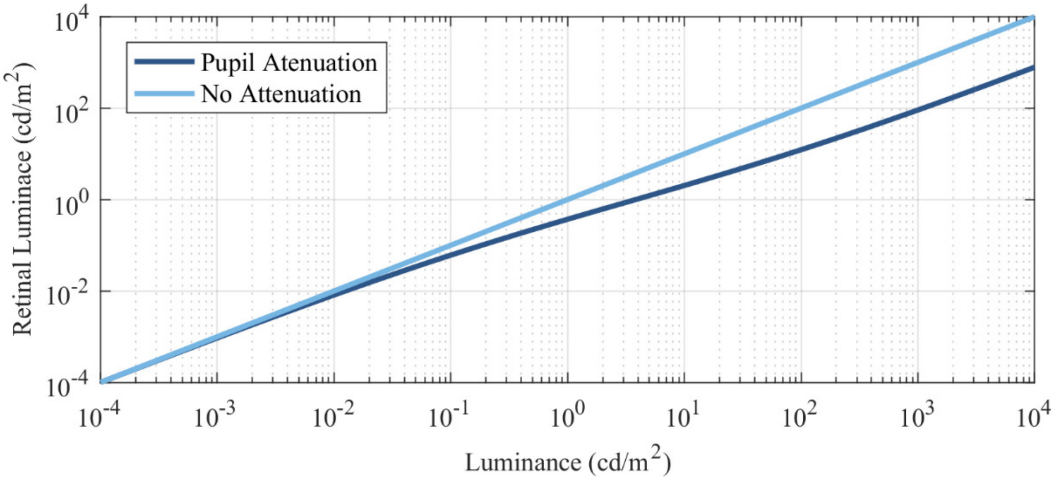


Figure 4-2. Attenuation of light from pupil dilation as a function of luminance.

Continuing the transduction path, light passes through the pupil, lens and vitreous body before reaching the retina where it is reflected off a pigmented layer of cells before being absorbed by the photoreceptors. The human retina has two forms of receptors, rods and cones. Rods are sensitive to low light and are responsible for vision from twilight to very dark [88]. Cones are less sensitive and are responsible for vision in daylight to moonlight.

The neural receptors for rods are intermingled with its neighboring rods reducing the ability to discriminate detail, but increasing their collective sensitivity. Marr [89] has shown how their response very closely resembles the output of difference of Gaussian filters of varying standard deviations. It is estimated that there are about 125 million rod cells in the human eye and encompass the majority of our vision. There are about 6 million [90] cones in the human eye primarily in the fovea. Cones have three different variations each of different lengths tuned to absorb light in the range of 564-580 nm, 534-545 nm, and 420-440 nm. These photoreceptors are activated in different visual ranges, the ranges are the photopic and scotopic range, and the range where they overlap is the mesopic range.

Light is absorbed by the rod and cone photoreceptors through a photochemical reaction. Light interacts with rhodopsin and goes through a cascading chemical process that ultimately inhibits the flow of sodium ions [91], hyperpolarizing the rod toward the potassium ionic charge, and depolarizes the neural inhibitors. This causes the bipolar cell to depolarize and causes an increase in neuro transmitter sent to the ganglion cells, observed as an increase in potential in an electroretinogram. The reaction breaks down photosensitive pigments and temporarily renders them insensitive, a process called bleaching [92]. Therefore, light adaptation can be considered a

process of depletion and regeneration of the photo pigments that is proportional to the incident light. This was first modeled by the Naka-Rushton [93] equation in 1963 as

$$\frac{R}{R_{max}} = \frac{I^n}{\mathrm{I}^n + {\sigma_L}^n} \qquad 4\text{-}8$$

where R/Rmax is the normalized cellular response, I is the light intensity, and σ_L is the light adapted semi-saturation constant, n is a constant that is used to fit the sigmoidal operator and is typically reported as 0.74 [94]. A good description of this process is reviewed by Evans [95]. The Naka-Rushton equation indicates the photoreceptor response given some quanta of light for a given previous background state.

The background state determines the light adapted semi saturation value. The half saturation constant is found during dark adaptation and is reported as 3.1-3.4 $\log_{10}$ Trolands [96] by Valeton [97] *et al* and 4 $\log_{10}$ Trolands by Boynton [98]. Valeton *et al* stated the light adaptation had contributions from both the photo bleaching mechanism and cellular adaptation. They defined the light adapted semi saturation parameter, σ_L, as

$$\sigma_L = \sigma_D \sigma_x \sigma_\beta \qquad 4\text{-}9$$

where σ_D is the half saturation constant, σ_x is the cellular adaptation parameter, and σ_β is the reciprocal of the fraction of pigment available at the bleached state. The values that Valeton state for each one of these terms can be seen in Table 4-1.

Table 4-1. Adaptation information from Valeton.

$Log_{10}(Y_b)$	$Log_{10}(\sigma_L)$	$Log_{10}(\sigma_x)$	$Log_{10}(\sigma_\beta)$
-	3.2=σ_D	-	-
2	3.5	0.3	0
3	3.904	0.7	0.004
4	4.4	1.1	0.1
5	5.2	1.1	0.9
6	6.3	1.1	2

Forming the new tone map operator

Using the data presented in Table 4-1, an empirical relation can be found for the product of the cellular adaptation mechanism and the percent of photo pigment that is available as a function of the retinal luminance. It is assumed there is a finite amount of photo pigment and

therefore must have a maximum and a minimum. This assumption is supported by Valeton's data since after six decades, the value of σ_β indicates 100% of the photo pigment has been utilized. It is also assumed the function should asymptotically approach the extremes, since this tendency is also seen in Valeton's table. One of the best functions to fit this condition is a sigmoidal function. To help ensure asymptotic behavior, several additional points were added for three decades on the low extreme that mimic the dark adaptation level, and the fit function was allowed to extrapolate where the upper asymptote would occur since it is not known. Utilizing (4-9) and integrating the sigmoidal function, the light adapted semi saturation parameter, σ_L, was defined as

$$\sigma_L(Y_b) = 10^{\sigma_D + o_v + \alpha\left[1+erf\left(\frac{Log_{10}(Y_b)-o_h}{h\sqrt{2}}\right)\right]} \qquad 4\text{-}10$$

where o_v is a vertical offset, o_h is a horizontal offset, h is a parameter that adjusts the slope of the concavity of the function, and α scales the amplitude. Equation (4-10) is an extension of a cumulative density function of the Gaussian probability density function, and was chosen because it is a sigmoidal operator, whose parameters implications are commonly known. Using a nonlinear regression, the values for o_v, o_h, h, and α were found and are listed in Table 4-2.

Table 4-2. Cellular Adaptation function coefficients.

Parameter	Value
α	3.959
o_v	-0.02856
o_h	6.729
h	2.696

with an r^2 value of 0.9981. Using these values, a prediction for the light adapted semi saturation parameter can be estimated for any given background retinal radiance by

$$\sigma_L(Y_b) = 10^{7.13+3.96\left[erf\left(\frac{Log_{10}(Y_b)-6.73}{3.81}\right)\right]} \qquad 4\text{-}11$$

Since a significant amount of data fitting, and assumptions were used in the formation of (4-11) only three significant figures were kept. A plot of the new light adapted semi saturation function exhibited in (4-11) can be seen in Figure 4-3 with Valeton's experimental values from Table 4-1 superimposed as circles.

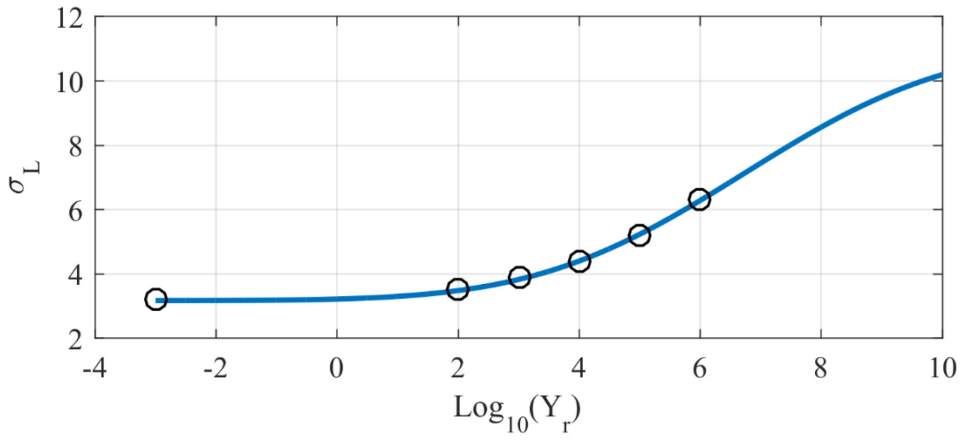


Figure 4-3. Line graph of the empirically derived function to find the light adapted semi saturation parameter. The circles are representative of the data reported by Valeton.

Substituting the value of the light adapted semi saturation value into the Naka-Rushton equation would allow for the prediction of how the eye would respond to a given background retinal luminance. The modified Naka-Rushton equation becomes

$$\frac{R}{R_{max}} = \frac{{Y_r}^n}{{Y_r}^n + \sigma_L(Y_b)^n} \qquad 4\text{-}12$$

where Y_r is the retinal luminance, Y_b is the background, or adaptation retinal luminance, and n is the experimentally found constant of 0.74. Using the modified Naka-Rushton equation (4-12) with the semi-saturation function (4-11) yields the response curves seen Figure 4-4 for various background luminance levels.

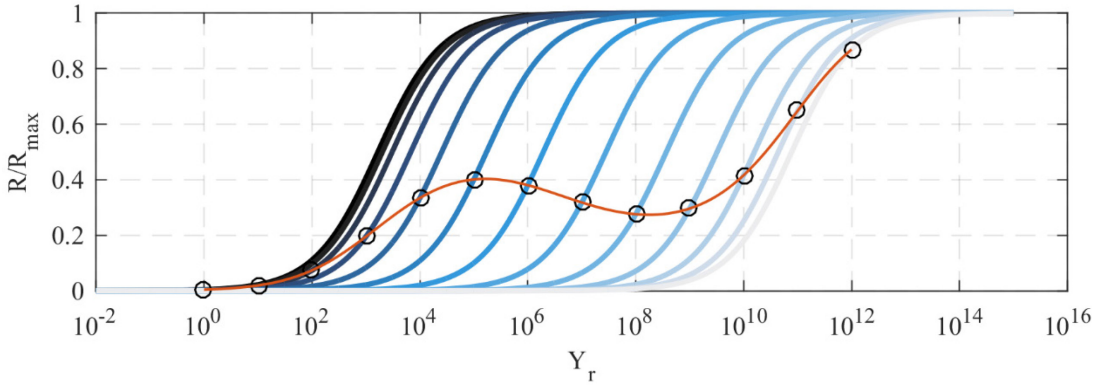


Figure 4-4. Predicted normalized output from the modeled adaptation mechanism. The black circles indicate the magnitude of the background luminance for the given trace. The red line indicates the normalized response for each possible background curve.

Figure 4-4 demonstrates the adaptation mechanics as light increases in intensity. Observing the response curves associated with low background luminance levels, there is a

significantly greater change in the response to brighter stimulus when compared to darker stimulus. The opposite is observed when the background luminance is very high. There is a small change in the retinal response when there is a brighter stimulus, but significant change in the response when a darker stimulus relative to the background is present.

Equation (4-12) can be used to scale the brightness of a given scene or region with knowledge of the retinal luminance and background luminance. The system will scale and adapt in a nonlinear fashion by a change in the pupil size and a change in the light adapting cellular mechanisms. The brightness scaling however does not yield colorimetric scaling.

In the retina, color vision is possible because of the cones. Cones can be broken down into three subgroups based on their physical dimensions called long, medium, and short. The long cones are tuned to a wavelength of 564 nm, the medium are turned to 534 nm, and the short are tuned to 420 nm. Rods are tuned to 500 nm and are active at low levels of illumination levels. This is why humans do not have color perception in low level light.

In the early 1930's, Wright [99] conducted a series of experiments to map out the tristimulus response by showing test subjects an array of colors and asking them to match the displayed color to a matching color. From this data, and utilizing metamerism, the phenomena that different tristimulus inputs can result in the appearance of the same color, the CIE 1930 equivalents was formed. Looking at the result from the Wright data, it can be observed that a negative red would have to be added in order to fully recreate our vision system's color acquisition ability. Utilizing an equivalent red, green, and blue response the full color spectrum can be recreated to be perceptually correct, despite originating from different wavelengths. The result is the CIE 1931 RGB color matching function where all colors are now additive.

The RGB color space was mapped to a triangular prism called the XYZ color space and a normalized version called the xyY color space. Typically the color accuracy of a device is displayed on an isoslice of this color space with the ordinate and abscissa originating from the x and y channels. The Y scale in the XYZ color space is an estimate of the photopic luminance from the acquired RGB color space and what will be used to map the acquired scene radiance into the perceptual luminance. In order to measure the hue of a given color, another more intuitive color space was created called the Hue, Saturation, Lightness (HSL) color space. The HSL color space is a cylindrical color coordinate system where hue is the color angle and saturation is the radial offset from the cylinder's central axis, and represents how much of that color is present. Lightness is how bright that combination appears and is adjusted by offsetting

the location of the hue/saturation plane from the base of the cylinder. The HSL color space is unique since it decorrelates the lightness of a given scene from its hue and saturation, unlike the standard RGB color cube. This will allow for the remapping of the apparent brightness from the modified Naka-Rushton equation into a color space where the hue and saturation can be preserved.

Scaling the output

A new sigmoidal tone mapping operator has been introduced based off of the Naka-Rushton equation, combined with a fitted function for the light adapted semi saturation parameter derived from Valeton's data. The retinal response can be used in conjunction with the pupil dilation to come up with a tone mapping process that scales the perceptual luminance captured in an HDR image to map the information into the display space of the monitor.

The process of tone mapping using this function has several steps and several variations. The general scheme can be seen in the flow chart depicted in Figure 4-5. The first step is to convert the high dynamic range RGB information into the XYZ color space and HSL color space. The luminance channel, the Y channel from the XYZ color space, is scaled into physical units if the original image is not already recorded in units of radiant exitance. Using the camera optics and sensor dimensions the pixels corresponding to user defined field of view can be selected. If none is selected then a 2 degree field of view is used to mimic the fovea. The fovea is the region in the eye that allows humans to discern fine detail. The median value from this pool of pixels is found. The median is used instead of a mean to prevent a single small very dark or very bright region in the image from biasing the overall look of the data.

The background luminance value is used in conjunction with (4-6) to find the estimated pupil dilation. The retinal luminance is then found with the estimated pupil dilation from equation (4-7). All of the data in the image can be scaled according to (4-12) to find the relative brightness. Before the lightness channel can be used to replace the L channel in the HSL data, a histogram equalization is performed to stretch the data from 0 to 1 should the data not already span the full range of an 8 bit output display. Finally, a gamma curve is applied to match the output gamma of the monitor or print media that the image will be displayed on. The saturation of this image will need to be scaled when converting from the HSL to RGB color space [100]. This is done by multiplying the S channel by a scale factor from 0 to 1, typically around 0.7.

The new HSL image can now be converted back into an RGB image and converted to an unassigned integer with 8 bits per channel for final display, recording, or printing.

A similar methodology can be used to convert the global mapping methodology into a spatial variant tone mapping methodology. An identical process can be carried out, but instead of scaling all of the pixels by the estimated background luminance, the image can be broken down into a series of windows. Here, the center pixel is scaled based on some number of surrounding pixels defined by the user. This process can be carried out for all pixels in the image. First, a base image must be made. In order to avoid abrupt changes, a background luminance image is convolved with a Gaussian kernel to blur the image. Then the mean background illumination can be calculated. The rescaling of the center pixel happens in an identical way to the spatial invariant method except on a much smaller window. This process is carried out for every pixel in the image, and a new image can be formed. This allows the darker regions to appear brighter and the brighter region to appear darker.

The spatial variant method allows for greater visualization in the shadow and highlight regions, however has the side effect of reducing the overall contrast of the image. A way to restore the contrast is by stretching the output of the new image to the full range of the display

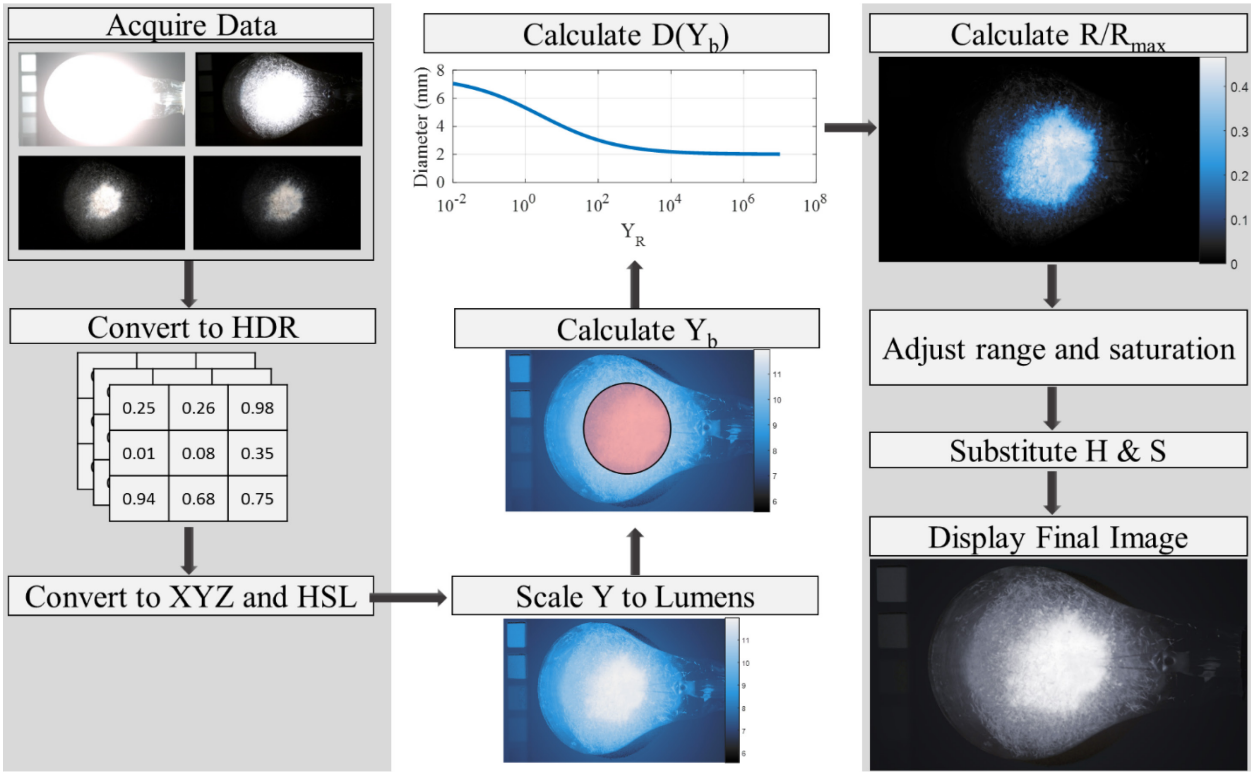


Figure 4-5. Process of creating and tone mapping an HDR image from a video sequence.

space. This is typically done by forming a histogram of the original data and plotting the cumulative sum of the histogram. The pixel brightness value associated with the bottom 0.5 percent and top 0.5 percent are found and the rest of the pixels are scaled linearly to the full range of the display space. This process restores the contrast while eliminating the extreme outliers of either bright or dark, to allow for a higher contrast image.

Comparison of results and extensions

In order to demonstrate the new tone mapping methods capability, a series of commonly referenced images in the public domain [101, 102, 103] were tone mapped with nine different methods, seen in Figure 4-6 through Figure 4-11. Different images were selected to demonstrate the strengths and shortcoming of the various methods for different spatial frequencies and total dynamic ranges.

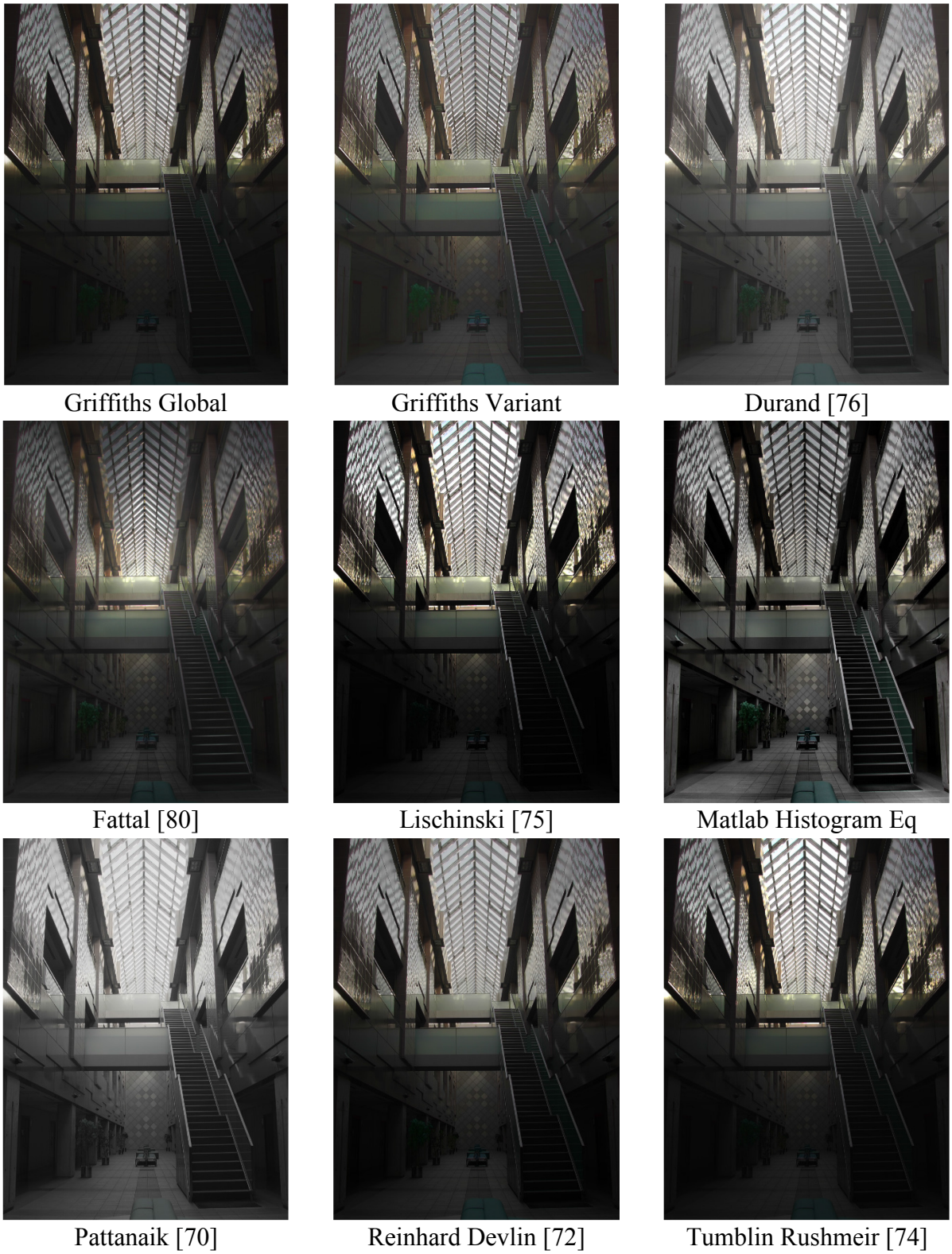
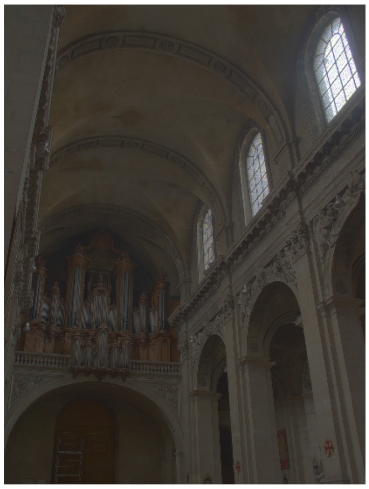


Figure 4-6. "Atrium Morning" image, tone mapped using various methods.

Griffiths Global

Griffiths Variant



Durand [76]

Fattal [80]

Lischinski [75]

Matlab Histogram Eq

Pattanaik [70]

Reinhard Devlin [72]

Tumblin Rushmeir [74]

Figure 4-7. "Church" image, tone mapped using various methods.

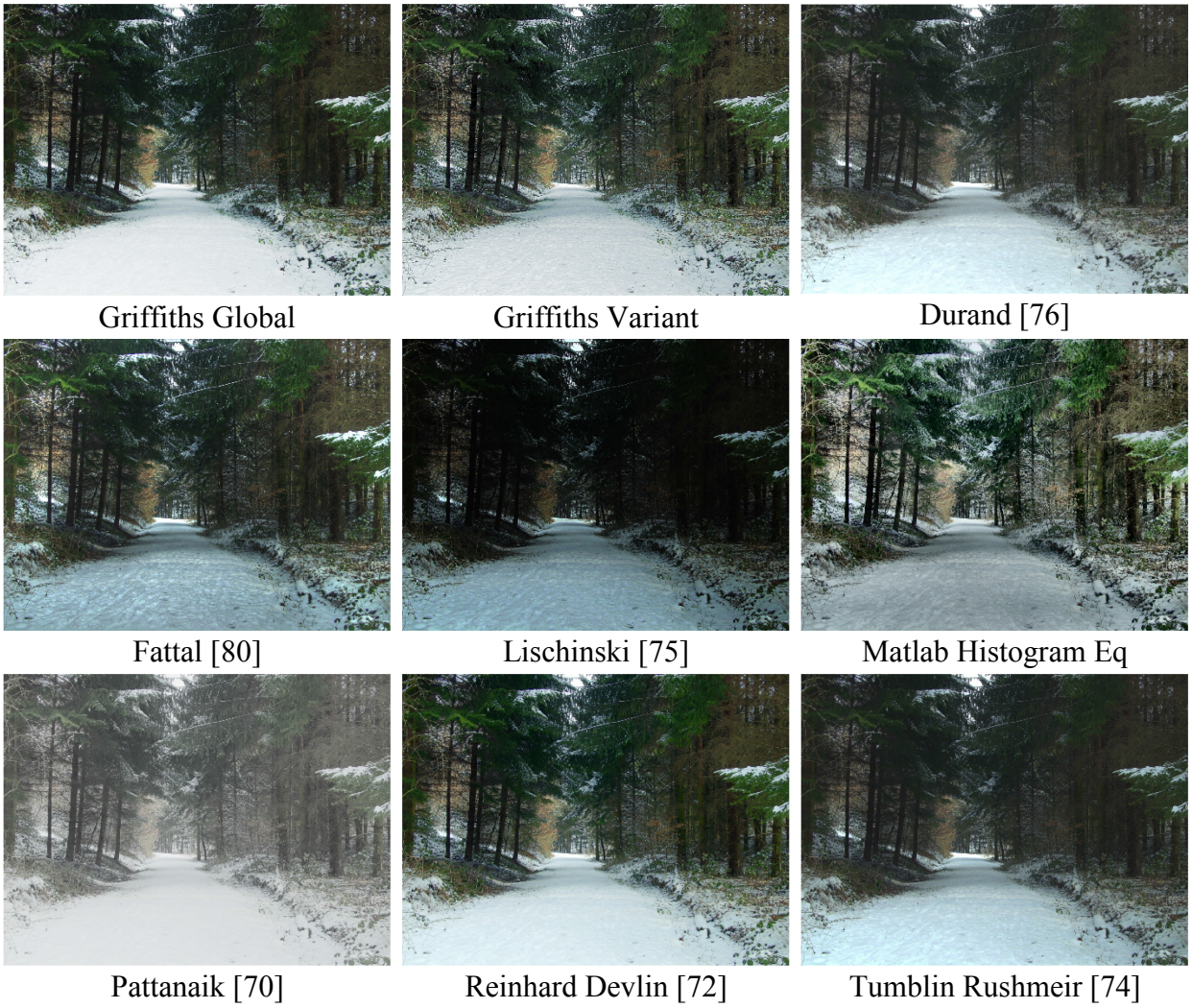
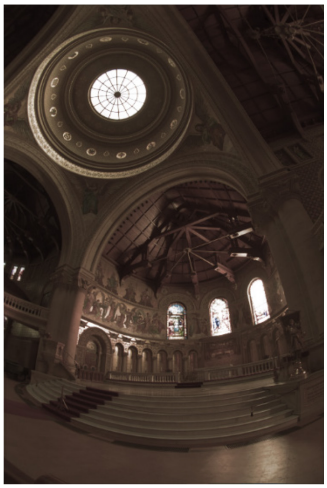


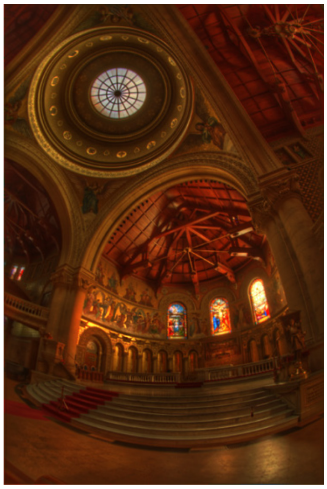
Figure 4-8. "Forest path" image, tone mapped using various methods.



Griffiths Global

Griffiths Variant

Durand [76]

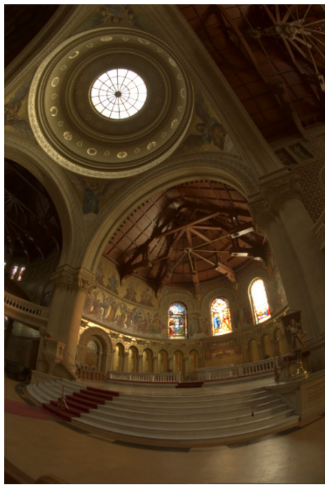


Fattal [80]

Lischinski [75]

Matlab Histogram Eq

Pattanaik [70]



Reinhard Devlin [72]

Tumblin Rushmeir [74]

Figure 4-9. "Memorial" image, tone mapped using various methods.

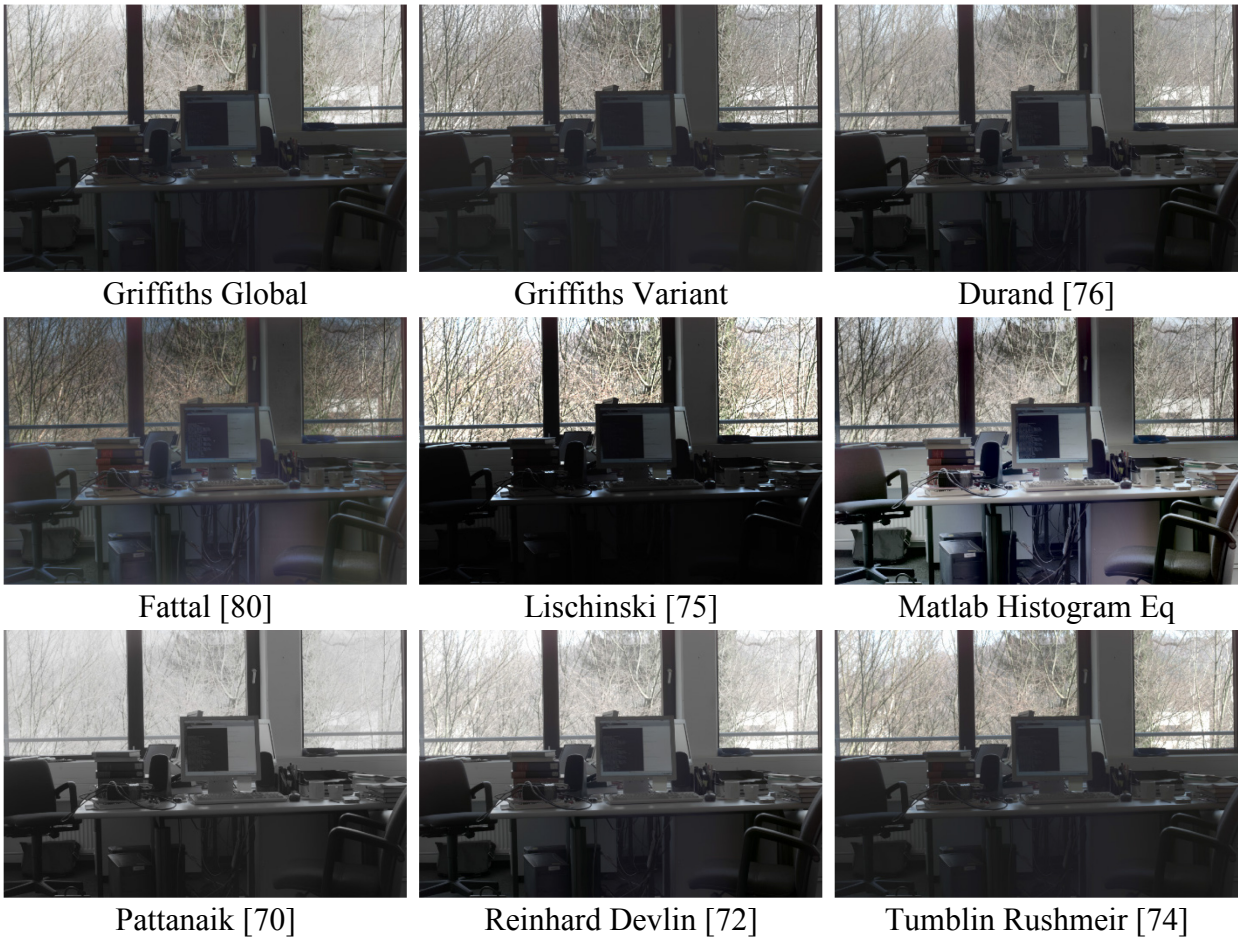


Figure 4-10. "Office" image, tone mapped using various methods.

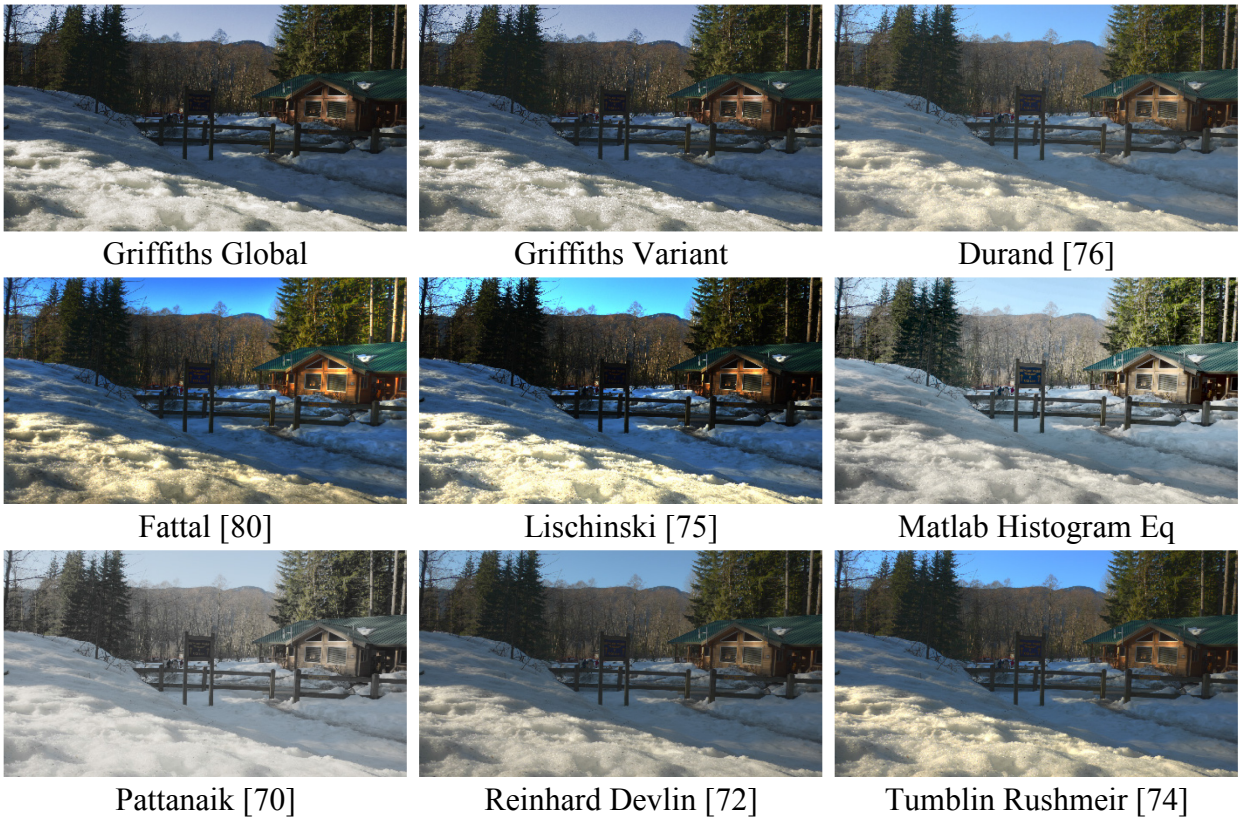


Figure 4-11. "Park" image, tone mapped using various methods.

Conclusions

In this chapter a new tone mapping methodology has been introduced. A rescaling of the measured radiant exitance is proposed by scaling the estimated radiant exitance from the HDR image into the photopic luminance. Once the luminance is known the retinal dilation is predicated, and the retinal irradiance calculated. Once the retinal irradiance is known the retinal response can be calculated from the empirical relation found for the light adapted semi saturation parameter.

The normalized perceptual brightness from the modified Naka-Rushton equation can then be scaled to the 8 bit output and adjusted using histogram equalization and gamma mapping. This process can be done either for a spatial variant or global methodology. The proposed method was then compared to seven other tone mapping methods for six different scenes.

In general the Fattal tone map operator exhibited the greatest amount of contrast while preserving the most detail. The Pattanaik tone mapping operator tended to exhibit the least amount of color saturation regardless of the scene. This showed to be a good attribute for the

memorial image, however, outdoor scenes such as the forest scene are left looking like there is little contrast. The Tumblin Rushmeir tone map operator was able to maintain a good balance between color saturation and contrast for most scenes. Matlab's built in tone map operator transforms the RGB to XYZ then scales the luminance channel in a log base 2 scale. It converts the data back into a RGB scale and, performs a histogram equalization and linearly remaps the data into an 8 bit value. The built in Matlab tone mapping method tends to work well when all the detail in the scene needs to be preserved, however tends to produce unnatural looking result.

The introduced tone map methods were able to recreate each scene in an accurate way. In general the global method will show a more perceptually realistic view, however it can crush shadows or highlight detail depending on the background illumination. The spatially variant method will dodge the shadows and burn the highlights to reveal more detail in each image at the expense of contrast, and perceptual accuracy. This can be seen in Figure 4-9, the memorial scene where more detail is observed in the stained glass windows in the spatially variant methods. A similar effect can be seen in the windows of the church scene. In the "Atrium Morning" scene the shadow detail is brighter making it easier to see, as well as more color information in the windows of the ceiling. In the "Park" scene more contrast can be observed in the snow, as well as more detail in the horizon trees. The office scene is similar in both images, however there is more detail in the window area around the trees.

One final comparison will be made to demonstrate how the tone mapping method combined with the error minimization method can improve the overall appearance of the image. Seen in Figure 4-12 are the same high dynamic range images created using the Error Minimization Method introduced in Chapter 3. On the left hand side of the figure is Matlab's histogram equalization tone map operator and on the right side is the proposed perceptually based tone map operator. The background luminance has been scaled to produce a visually appealing image. The main difference is the data that appears noisy is all at a very similar luminance level. When a tone map operator is used that perceptually scales the image, the bright data is given more contrast and the small perturbations in the dark regions are crushed since the change would be barely perceptual. This has the result of producing a realistic looking image, without the flare and misalignment degradations.

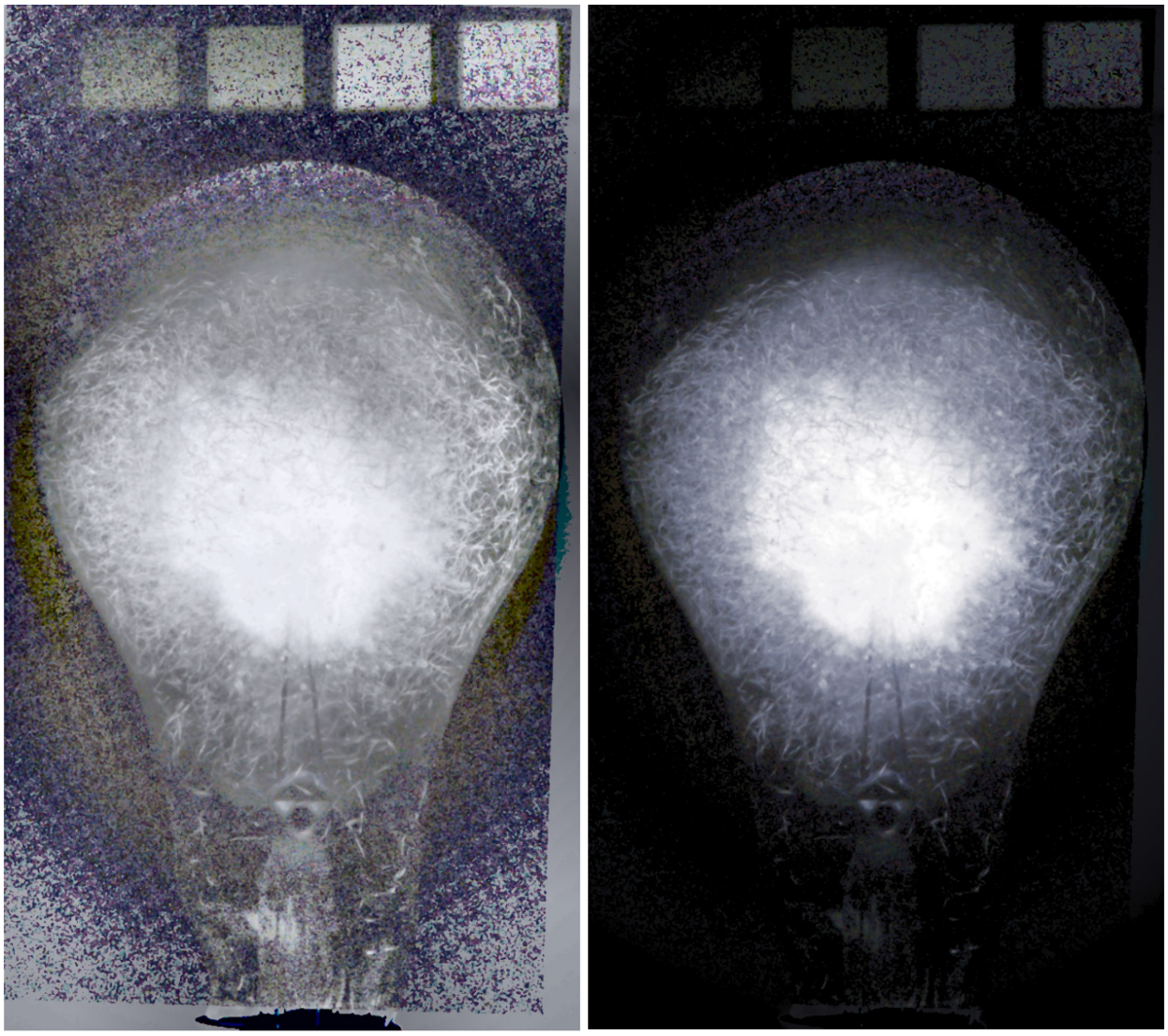


Figure 4-12. Left. Image from Chapter 3 using Matlab's histogram equalization tone map method and EMM. Right. The same HDR data from the EMM method tone mapped with the proposed spatial varying tone map operator.

Chapter 5. Developing a high speed high dynamic range video system

High speed digital video is a commonly used measurement tool. It is used in the automotive industry, manufacturing industry for quality control, government for qualitative and quantitative assessment of functional items, and more recently, in the entertainment and sporting industry. The dynamic range of a digital camera is limited to bit depth of the analog to digital converter and deep well capacity of the sensor. The dynamic range of an explosive event can exceed the camera's native dynamic range by many orders of magnitude, often in excess of 120 dB causing loss of data from sensor saturation. A method has been developed to utilize current off the shelf synchronized high speed cameras to film events consisting of extremely high dynamic range events. This chapter will describe the process for the design and implementation of an optimized system composed of any number of cameras, and demonstrate the capabilities of a constructed system to capture a video. Several examples will be given both in a controlled laboratory setting and a non-laboratory setting. The first set of demonstrations is of a strobe bulb spanning 169 dB range (28 stops) at 79,000 pps using three and four high speed cameras. A second set of demonstrations will be presented from several fielded experiments, proving the system's ability to expand beyond the controlled laboratory environment.

HDR video background

Researchers have developed ways to capture and estimate photometric radiance information in excess of a single image's static dynamic range by statistically recombining multiple captures. One assumption of this type of process is the scene is invariant to change such as movement of objects within the frame during all exposures. When change does occur, it leads to ghosting within the image due to the differences in the image series. If there are enough exposures to statistically define the static objects, the objects in motion can be statistically removed [104]. This method fails if the object in motion needs to be recorded, or if there are not enough samples to remove the ghost objects. This has led to significant challenges in the development of High Dynamic Range (HDR) video, some of which have been addressed in the previous chapters.

Several researchers have formed different solutions such as rapid bracketing [105], split aperture imaging [48], a Fresnel [47] based imaging system, and randomized integration time

[106]. Frame bracketing is an effective method to capture HDR video; however, the method requires a camera to run significantly faster than the desired output video's playback, limiting low light sensitivity. Frame bracketing also reduces the maximum speed at which the camera can operate. A 256x256 video filmed at 100,000 pps with three bursts would have to operate at a rate over 19 GHz. If the camera has a 12 bit image, the data storage rate would be greater than 29 GB/s, currently making this type of system difficult to produce, however plausible with continual advances in solid state memory.

The Fresnel based system utilizes multiple images through the use of a Fresnel beam splitter that parses the light to the various sensors. This system allows for commercial off the shelf components to be used; however, this will not work if the source of light to be viewed is polarized. In 2011, Tocci *et al* [47] showed how a system comprised of a field programmable gate array and a pair of cameras could be used for the creation of real time tone mapped HDR video. By having the cameras observe the same field and selecting which related pixels from the camera pairs had the most accurate information, an aggregate of the selections could be used to display a welding process. This method allowed for the direct visualization of the weld, melt pool, and materials being welded in their demonstration.

That same year, NASA used six high speed cameras to video the STS-134 launch [107] and five high speed cameras and one thermal camera to visualize the STS-135 launch. NASA had a camera shelter with 6 high speed cameras running at 200 pps set at six different exposure times. The cameras were oriented to form a 3 x 2 grid. The subjects being filmed were sufficiently far relative to the camera spacing such that parallax error was not a problem, and the first digital high speed HDR video known to the author was formed. However, NASA's system would be subject to parallax errors if it were to be used to capture an event closer to the lens. Also the use of six high speed cameras makes this a cost and time prohibitive setup for daily use and implementation.

Introduced in this paper is a permutation of the HDR video system developed by NASA merged with an extension of the system introduced by Tocci. In order to reduce the number of cameras while maintaining the desired dynamic range capabilities, an optimization of the image formation pipeline will reveal the optimal attenuation per camera needed for the desired dynamic range and noise tolerance. Several authors have proposed different exposure spacing [42] [108], this method will reveal the ideal spacing for this application. The optimal camera spacing

coupled with error minimization method for video frame formation will allow for less cameras to be used and still span the needed dynamic range while producing visually appealing results.

System Physical Architecture

The high speed HDR system developed by NASA had the cameras staged side by side. A Vision Research Phantom 711 camera has a physical width of 144 mm. The parallax introduced from the minimum camera offset from the dissimilar foreground and background would inhibit the successful combination of the differently exposed images when viewing objects close to the cameras relative to their lens's focal length.

Tocci *et al* demonstrated the viability of using a multi sensor system utilizing beam splitters to capture differently exposed video frames and remove the parallax error. Tocci *et al* utilized a Fresnel beam splitter to split the incident light to the varying sensors with different attenuation levels. Using this as a foundation, a modified version was assembled on an optical breadboard to allow for the use of high speed cameras and non-polarizing optics. A schematic of the generalized setup can be seen in Figure 5-1.

Non-polarizing cube beam splitters were utilized to minimize ghosting inherent in plate beam splitters and distortions when viewing polarized sources. Cube beam splitters were chosen for their physical strength since the setup would be subject to blast overpressure and other physical trauma inherent to sensors in proximity to energetic testing. Neutral density filters can be placed either in between the cube beam splitters or directly on the lens. Care should be taken to ensure the optics have anti-reflective coatings and minimal ripple in the passband to prevent color shifts. The full optical path consists of the beam splitter, neutral density filter, potentially varying apertures if needed, and sensor gain modification should a different kind of camera be utilized.

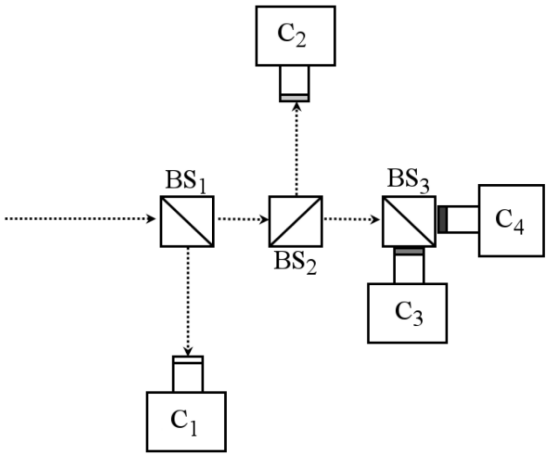


Figure 5-1. Schematic of the HDR video system. Cameras are labeled C1-C4, and beam splitters are labeled BS1-BS3.

Each of the cameras is mounted to a three axis pan tilt mount to allow for fine adjustment and alignment of the camera bodies. The beam splitters are placed on rotating stages to allow for adjustment of each beam splitter. The cameras are aligned by placing a target on the pedestal that each beam splitter sits upon. The cameras are aligned so the center of the sensor focuses on the center of the target on the pedestal. The target is removed once the cameras are aligned, then the beam splitters are placed on the pedestal and rotated so each image coincides with a target point placed in the far field. Once the system is assembled, a hood is placed over the optical path so no spurious light can affect the system.

The advantage of this form of system becomes apparent when the exposure of each camera can be held constant and synchronously triggered to prevent different levels of motion blur affecting the final data. Next the optical path must be modeled including the sensor model to accommodate all the different sources of attenuation, amplification, and noise.

Optimization of the architecture

The HDR camera system can be modeled according to the process depicted in Figure 5-2. Reviewing the derivation from 0 with modifications pertinent to this system, the incident photon flux, φ, enters the leading beam splitter and can be defined as

$$\varphi = \frac{1}{hc}\int_{\lambda_2}^{\lambda_1} L(\lambda)d\lambda \qquad 5\text{-}1$$

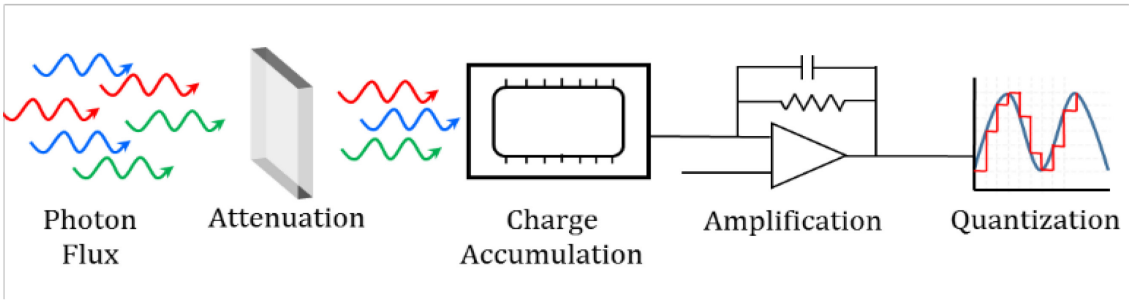


Figure 5-2. Pictorial representation of the system model. The incident photon flux is attenuated, converted to charge, amplified, discretized and stored.

where h is Plank's constant, c is the speed of light and L(λ) is the spectral radiance where λ is the wavelength. Some of the light will be reflected, transmitted or scattered by each beam splitter multiplicatively. The light will be further attenuated by a neutral density filter and collimated by the physical aperture of the lens on the camera, reducing the overall photon flux to the imager. The attenuation for the camera system, A_{phy}, can be defined by

$$A_{phy} = \frac{\pi}{4f/\#_j^2\left(\frac{s_2}{s_1}-1\right)^2} 10^{-N_{D,j}} \begin{bmatrix} R_j \prod_{j=1}^{J-1} T_{R,j-1} \\ \vdots \\ \prod_{j=1}^{J} T_{R,j} \end{bmatrix} \quad 5\text{-}2$$

where f/#, is the f stop number, s_2/s_1 is the ratio of the distance to the object and the focal length, N_D is the neutral density filter value, R_j is the reflectance of the j^{th} beam splitter and T_{Rj} is the transmittance of j^{th} beam splitter with T_0 equal to 1 always. Under the assumption the object being viewed is significantly farther than the focal length the ratio s_2/s_1 tends to 0 and can be omitted from (5-2). Applying (5-2) to the camera model, the charge accumulation for a camera, s, can be defined as

$$s = \frac{A_{phy}\Delta t A}{hc}\int_{\lambda_{min}}^{\lambda_{max}} \eta(\lambda)\, L_s(\lambda)\lambda d\lambda \quad 5\text{-}3$$

where A is the active area of the pixel, Δt is the exposure time, η is the quantum efficiency of the sensor at a given wavelength, L_s is the spectral radiance of the scene, and λ is the wavelength of light.

The next process in Figure 5-2 is amplification. Simplifying (5-3) by assuming monochromatic light, and adding in the sensor gain, the charge accumulation for a camera j and pixel i can be defined as

$$s_{ij} = \frac{\Delta t_j A_{phy,j} g_j A \eta \lambda}{hc} \Phi_i \tag{5-4}$$

where g is the sensor gain, and Φ_i is the radiant exitance. Following Figure 5-2, the charge is then discretized using

$$s_{ij} = \frac{q_s \Delta t_j A_{phy,j} g_j A \eta \lambda}{N_{well} hc} \Phi_i \tag{5-5}$$

where N_{well} is the deep well capacity of the sensor, and q_s is the number of discretization steps from the camera's analog to digital converter. Rearranging (5-4) allows for the radiant exitance to be estimated from a digital value, Z_{ij}, given by

$$\Phi_{ij} = \frac{Z_{ij} N_{well} hc}{q_s A \Delta t_j A_{phy,j} g_j \eta \lambda} \tag{5-6}$$

The maximum radiant exitance that can be acquired can be defined when the digital value is equal to the number of quantization levels, the minimum of the physical attenuation, exposure time and gain, defined as

$$\Phi_{max} = \frac{N_{well} hc}{min[\Delta t A_{phy} g] A \eta \lambda} \tag{5-7}$$

where $\min[\cdot]$ is the minimum of the set. A weighted average is the maximum likelihood estimator used to find the radiant exitance estimate of the scene from the various Wycoff layers. The HDR estimate, Φ_{HDR}, can be found using

$$\Phi_{HDR,i} = \frac{N_{well} hc}{q_s A \eta \lambda} \sum_{j=1}^{J} w_{ij} \frac{Z_{ij}}{\Delta t_j A_{phy,j} g_j} \left(\sum_{j=1}^{J} w_{ij} \right)^{-1} \tag{5-8}$$

where J is the number of cameras used in the system. When a nonlinear camera is being used (5-8) can be modified according to [32]. However, almost all sensors have a linear response before software compression, therefore (5-8) is valid when utilizing the raw camera output.

Shot noise arises from the light traveling in energy quanta and is therefore modeled by a Poisson distribution. With large arrival numbers it can modeled as a Gaussian distribution. The shot noise can be defined as

$$\sigma_{sw,ij} = \sqrt{\frac{hc}{\Delta t_j A \lambda} \Phi_i} \tag{5-9}$$

Figure 5-2 shows the mean radiant exitance gets attenuated before reaching the sensor site. Therefore the shot noise, being proportional to the mean of the radiant exitance, will also be reduced. The shot noise at the sensor sight can be defined as

$$\sigma_{ss,ij} = \sqrt{\frac{hc}{A_{phy,j} \Delta t_j A \lambda} \Phi_i} \tag{5-10}$$

and the read and dark current noise can be estimated by

$$\sigma_{ADU,ij} = \frac{N_{well} hc}{g_j \Delta t_j A_{phy,j} \eta \lambda A q_s} \sqrt{{\mu_{rd,j}}^2 + {\mu_{dc,j}}^2} \tag{5-11}$$

where μ_{rd} is the read noise standard deviation and μ_{dc} is the dark current standard deviation.

The weighting function that has been found to minimize the error on formation is proportional to the reciprocal of the total variance in the system [42], [44] [43]. A simplified version of these weighting functions has been introduced in Chapter 2, a dimensional version can be defined as

$$w_{ij} = \frac{1}{{\sigma_{ss,ij}}^2 + {\sigma_{ADU,ij}}^2} \tag{5-12}$$

where σ_{ss} is the shot noise standard deviation at the sensor site and σ_{ADU} is a combination of the read and dark current noise.

Non-Dimensionalizing the governing equations

In order to have a generalized analysis of a camera system, the governing equations that were just reviewed need to be non-dimensionalized. The non-dimensionalized equations can be used to generate non-dimensional images that can later be recombined to form a non-dimensional HDR image. Then through an analysis of experiments, the ideal attenuation settings can be found by maximizing an objective function that maximizes the range the camera system will

have above a defined SNR. A listing of all the parameters used to non-dimensionalize the characteristic equations can be seen in Table 5-1.

Parameters that have units of ADU are normalized by the number of quantization steps. Parameters that have units of W/m^2 are normalized by the maximum radiant exitance the camera system can measure. Terms that are in units of charge were normalized by the deep well capacity of the sensor. Other terms such as exposure time, gain and physical attenuation were normalized by the reciprocal of the minimum value they can take.

The radiant exitance shall be modeled in as a non-dimensional parameter. The radiant exitance can be defined as

$$\Phi_i^* = Z_{ij}^* A_{phy,j}{}^* g_j^* t_j^* \tag{5-13}$$

and will take on a value in the interval

$$\Phi_i^* \in \left[10^{-\frac{DR}{20}}, 1\right] \tag{5-14}$$

where DR is the dynamic range of the camera system in units of decibels and is defined by

$$DR = 20Log_{10}\left(\frac{\Phi_{max}}{min[\Phi]}\right) = 20Log_{10}(q_s) \tag{5-15}$$

This defines the dynamic range of the acquisition system, not the resolvable range due to the baseline noise floor. The shot noise from (5-10) can be modeled using

$$\sigma_{sw,ij}^* = \sqrt{\frac{\Phi_i^* t_j^*}{\eta^*} \min[A_{phy}] \min[g]} \tag{5-16}$$

where min[A_{phy}] is the minimum value of physical attenuator vector and can be found from

$$\min[A_{phy}] = \frac{q_s \max[A_{phy} tg]}{\min[tg] 10^{\frac{DR}{20}}} \tag{5-17}$$

For the high speed HDR video system, the gain and the exposure time will be held constant across all cameras, therefore (5-17) can be reduced to

$$\min[A_{phy}] = \frac{q_s \max[A_{phy}]}{10^{\frac{DR}{20}}} \tag{5-18}$$

where max[A_{phy}] is the physical constraint of the architecture from the reflectance of the first beam splitter and smallest f/# lens that can be used. Equation (5-16) indicates the shot noise will decrease as the exposure time increases. A best practice would be to increase the exposure time as much as possible to reduce this source of noise.

After attenuation the mean radiant exitance will be reduced, therefore reducing the shot noise at the sensor site. The non-dimensional shot noise at the sensor site is defined as

$$\sigma^*_{ss,ij} = \sqrt{\frac{\Phi^*_i {A_{phy,j}}^* t^*_j}{\eta^*} \min[g]} \tag{5-19}$$

The incident energy is converted to charge via (5-5). The non-dimensional charge accumulation can be defined by

$$s^*_{ij} = \frac{\Phi^*_i}{{A_{phy,j}}^* g^*_j t^*_j} \tag{5-20}$$

The read noise and dark current standard deviation is defined by (5-12) and can be shown in non-dimensional form as

$$\sigma^*_{ADU,ij} = {A_{phy,j}}^* g^*_j t^*_j \sqrt{\mu^{*2}_{rd,j} + \mu^{*2}_{dc,j}} \tag{5-21}$$

Table 5-1. Non dimensional terms defined.

Equation	Description
$t^* = \frac{min[\Delta t]}{\Delta t}$	Exposure time relative to the minimum.
$g^* = \frac{min[g]}{g}$	Sensor gain relative to the minimum.
${A_{phy}}^* = \frac{min[A_{phy}]}{A_{phy}}$	Physical attenuation relative to the minimum.
${\mu_{rd}}^* = \frac{\mu_{rd}}{q_s}$	Ratio of read noise relative to the signal maximum.
${\mu_{dc}}^* = \frac{\mu_{dc}}{q_s}$	Ratio of dark current noise relative to signal maximum.
$Z^* = \frac{Z}{qs}$	The digital value relative to the number to quantization steps.
$\eta^* = \frac{N_{well}}{\eta}$	Number of photon required to saturate the sensor.
$\Phi^* = \frac{\Phi}{\Phi_{max}}$	Radiant exitance relative to the system maximum.
$w^* = \frac{w}{{\Phi_{max}}^2}$	Weighting function relative to the max radiant exitance.
${\sigma_{ADU}}^* = \frac{\sigma_{ADU}}{\Phi_{max}}$	Read and dark current noise relative to the maximum resolvable radiant exitance.
${\sigma_{ss}}^* = \frac{\sigma_{ss}}{\Phi_{max}}$	Sensor level shot noise relative to the maximum resolvable radiant exitance.
${\sigma_{sw}}^* = \frac{\sigma_{sw}}{\Phi_{max}}$	Radiant exitance shot noise relative to the maximum resolvable radiant exitance.
$s^* = \frac{s}{N_{well}}$	Charge accumulation relative to the deep well capacity of the sensor.
$e^* = \frac{e}{N_{well}}$	Total charge accumulation per sensor relative to the deep well capacity.

Optimization

In this section a steered analysis of experiments optimization will be introduced to find the ideal camera settings in an efficient way. This is an optimization process that has been created under the assumption that the defined objective function in a multi parameter space has a global minima or maxima that can be found through a random search criteria.

The total dynamic range, number of cameras, and the camera's dynamic range needs to be known for the analysis. A random guess of the ideal attenuation factors is produced with a random number generator. The maximum and minimum values are fixed by the desired dynamic range and minimum attenuation level. The middle cameras must be confined to the range, $A_{phy,j+1} \dots A_{phy,J-1} \in \left[min\left[A_{phy}\right], max\left[A_{phy}\right]\right]$, to allow for any value including a repeated max[A_{phy}] and min[A_{phy}] for the experiment using A^*_{phy}. A non-dimensional radiant exitance field is created using (5-13 and 5-14) logarithmically spaced. The shot noise will be added to the radiant exitance field utilizing (5-19) by

$$\Phi^*_{s,ij} = \Phi^*_i + \sigma^*_{sw,ij} randn[\cdot] \quad 5\text{-}22$$

The read noise and dark current noise contributions are added to the non-dimensionalized charge accumulation, found in (5-20), is defined by

$$e^*_{ij} = \frac{\Phi^*_i}{A_{phy,j}{}^* g^*_j t^*_j} + \sqrt{\mu^{*2}_{rd,j} + \mu^{*2}_{dc,j}}\, randn[\cdot] \quad 5\text{-}23$$

Since e* is a ratio of the charge accumulation to the deep well capacity, the simulated non-dimensional digital value can be defined as

$$Z^*_{ij} = \frac{\lfloor q_s e^*_{ij} \rceil}{qs} \quad 5\text{-}24$$

where $\lfloor \cdot \rceil$ denotes rounding to nearest integer function. The rounding is needed to simulate the quantization error inherent to the discretization of the data. An indicator value, Ψ, will be used to clamp the digital value from zero to 1 by

$$\Psi(Z^*_{ij}) = \begin{cases} 0, & Z^*_{ij} < 0 \\ 1, & 0 \geq Z^*_{ij} \geq 1 \\ 0, & Z^*_{ij} > 1 \end{cases} \quad 5\text{-}25$$

Values greater than 1 indicate oversaturation of the sensor and an indicator value of 0 will be given. Values less than 0 arise from the symmetrically distributed error in (5-23) when Z* approaches zero, and are non-physical, therefore they will be given an indicator value of 0.

The HDR image in dimensional form is defined by (5-8). The non-dimensional HDR image can be formed and is defined as

$$\Phi^*_{HDR,i} = \frac{\sum_{j=1}^{J} \Psi(Z^*_{ij}) w^*_{ij} Z^*_{ij} A^*_{phy,j} t^*_j g^*_j}{\sum_{j=1}^{J} w^*_{ij}} \quad 5\text{-}26$$

where the non-dimensional weighting function, w^*, can be defined by

$$w^*_{ij} = \frac{1}{\sigma^{*2}_{ss,ij} + \sigma^{*2}_{ADU,ij}} \quad 5\text{-}27$$

with σ^*_{ss} is defined in (5-19) and σ^*_{ADU} defined by (5-21).

In order to optimize the camera system, it is desired to have the maximum radiance range possible above the noise threshold to accurately be able to reproduce the scene. The objective function for a particular iteration k, $\mathcal{O}^k$, to maximize, can be defined as

$$\mathcal{O}^k = \frac{1}{p} \sum_{i=1}^{p} \left[20 Log_{10} \left(\frac{\Phi^*_i}{\left| \Phi^*_i - \Phi^*_{HDR,i} \right|} \right) > \frac{J\zeta - DR}{J - 1} \right] \quad 5\text{-}28$$

where ζ is the base dynamic range in dB, J is the number of cameras being simulated, and p is the number of points used to span Φ^*. The total coverage is calculated for the initial values of A^*_{phy}. The process is iterated several times, typically less than 30 iterations, to build a baseline of guesses and the associated values of $\mathcal{O}$. The pool is then sorted in descending order, so the maximum of (5-28) is indexed first. A search bound that narrows as the error is reduced can then be defined. The range, ι, of the new guess is defined by

$$\iota = \frac{J(DR - \zeta)}{DR(J - 1)} \quad 5\text{-}29$$

with the upper bound for the guess range, g_u, to be

$$g_u = (\iota - \max[\mathcal{O}]) A_{phy,j} + A_{phy,j} \quad 5\text{-}30$$

and the lower bound, g_l, as

$$g_l = A_{phy,j} - A_{phy,j} (\iota - \max[\mathcal{O}]) \quad 5\text{-}31$$

The next new guess of the optimal A^*_{phy} still with the maximum and minimum values set, will be formed by

$$A^{k+1}_{phy,j} = \left(g^k_u - g^k_l\right)\text{rand}[\cdot] + g^k_l \qquad 5\text{-}32$$

where rand is a uniformly distributed random number. An example of a four camera optimization spanned 170 dB range can be seen in Figure 5-3.

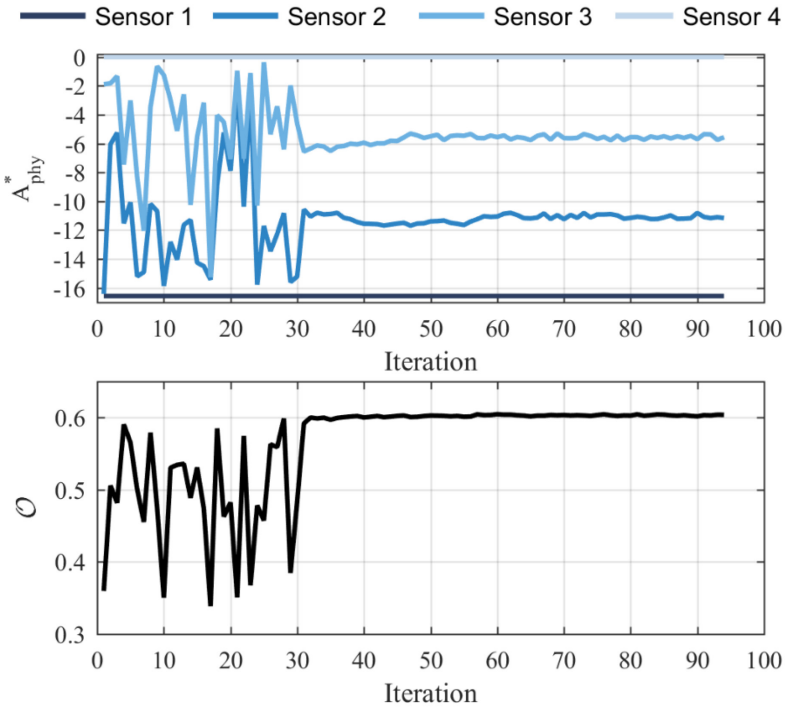


Figure 5-3. 170 dB camera system iterative convergence.

Optimal design of an n camera high speed video system

The optimization routine was performed for three different camera configurations utilizing two, three, and four cameras. Vision Research Phantom 711 cameras were modeled due to their availability to the author. The setup of the four camera system can be seen in Figure 5-4. The value of max[A_{phy}] was defined by the physical constraints imposed by the optics on hand. The first beam splitter's reflectance value of 0.45, and the smallest f/# the 200 mm Cannon lens could take on, 2.8, defined max[A_{phy}] as 0.0451. The Vision Research Phantom 711 cameras utilize a 12 bit A/D, giving q_s a value of 4096.

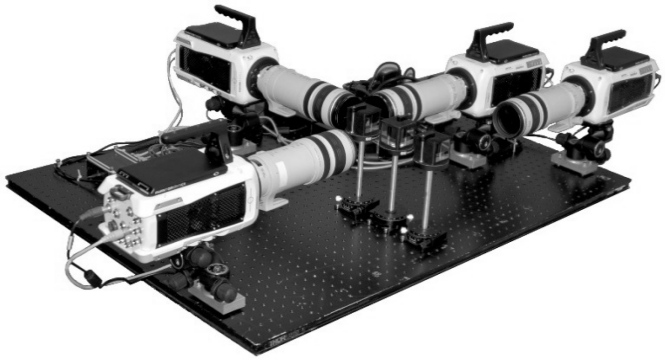


Figure 5-4. Image of setup used for high speed HDR video recording. Each camera is a Vision Research Phantom 711 camera coupled with Cannon 400 mm lenses. Covers removed for clarity.

The optimization routine was sequenced to find the optimal camera spacing for ranges spanning from the base dynamic range, $20\text{Log}_{10}(q_s)$, or 72 dB, every 3 dB, to the maximum, $J{\cdot}20\text{Log}_{10}(q_s)$, where J is the number of cameras in use, the results can be seen in Figure 5-5.

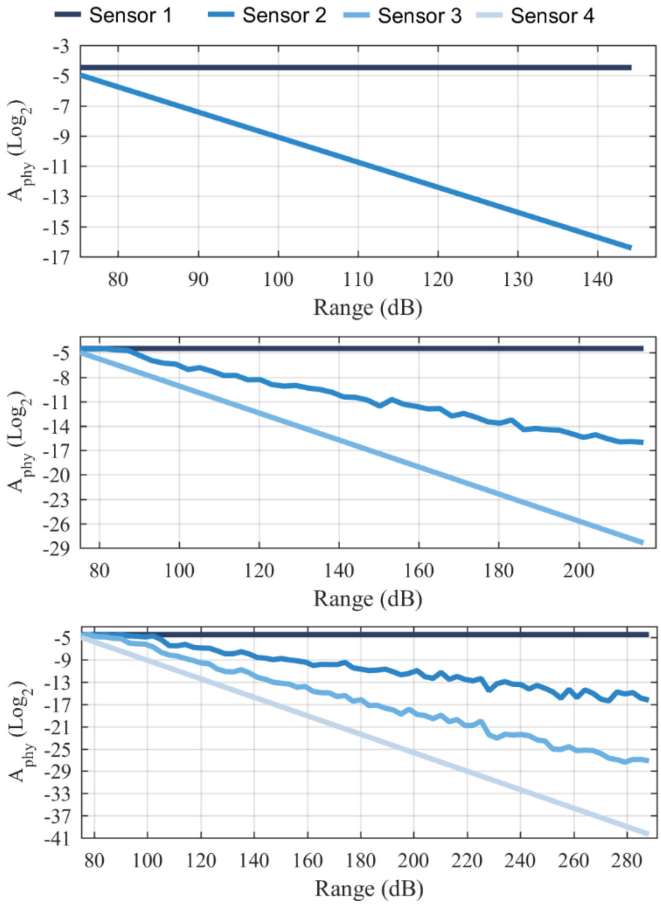


Figure 5-5. Optimal attenuation settings found for a two, three, and four camera system.

Due to the bit depth of the high speed cameras, at least three cameras will be needed to span the full 170 dB design envelope. Figure 5-5 indicates the camera attenuation spacing that minimizes the overall error on formation in all cases, A_{Opt}, can be defined by

$$A_{Opt} = \frac{1}{J-1} Log_2 \left[\frac{q_s}{10^{DR/20}} \right] \quad 5\text{-}33$$

Utilizing (5-33) and knowing DR and q_s, an error optimal system can be designed for any range scene with any set of camera sensors. This analysis was completed solely on the basis of what dynamic range the cameras could span. It was not computed based on the signal to noise ratio.

If the cameras continue to extend to ranges greater than the SNR allows, it is possible to get image degradation from the dark current noise source. The dark current will manifest itself as an additional bias in the lower SNR regions of the image. Figure 5-6 demonstrates what percentage of the camera's range is at or above a given SNR for a given dynamic range. Since the native dynamic range for a V711 camera is about 60 dB the highest quality range and dynamic range occurs at 120 dB with two cameras, 160 dB with three, and 180 with four.

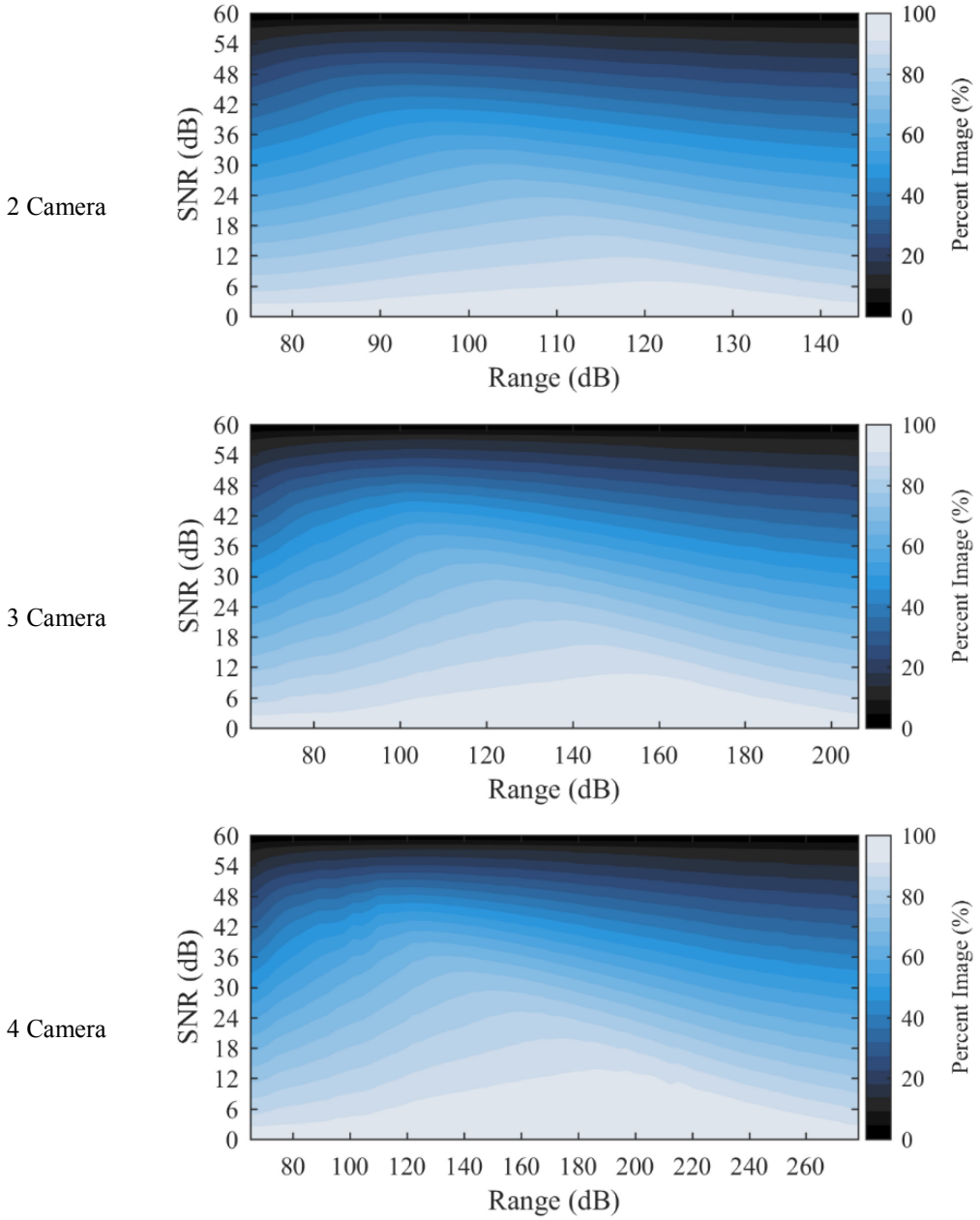


Figure 5-6. Percent of the dynamic range covered by a given SNR.

Design of a 170 dB camera system

An optimized camera system was constructed to view a Photogenic C4-19 flash bulb initiated by a Photogenic 2500 DR set to a 1000 Watt output. The Photogenic C4-19 flash bulb reaches peak brightness in 10 μs and spans a range greater than 160 dB at peak output when viewing the flash bulb directly. Three or more cameras must be used to view a scene with a

greater than 140 dB range due to the dynamic range limitation of a single camera. The camera system must be able to sample the event adequately to measure the rise time. A sample rate of 79,000 pps was utilized with a 256x256 spatial resolution. With this acquisition rate, 8 frames will be captured on the rise of the bulb initiation. With a 170 dB dynamic range, a min[A_{phy}] value of 5.83×10^{-7} or 21 stops attenuation will be needed to span the full range. Using (5-33), A_{opt} for a three camera system will be -8.1 stops attenuation per camera and -5.4 stops for the four camera system. The signal to noise ratio can be seen in Figure 5-7 for the three camera and four camera setup and indicates what percentage of the acquirable range will have a SNR greater than the value listed on the x axis.

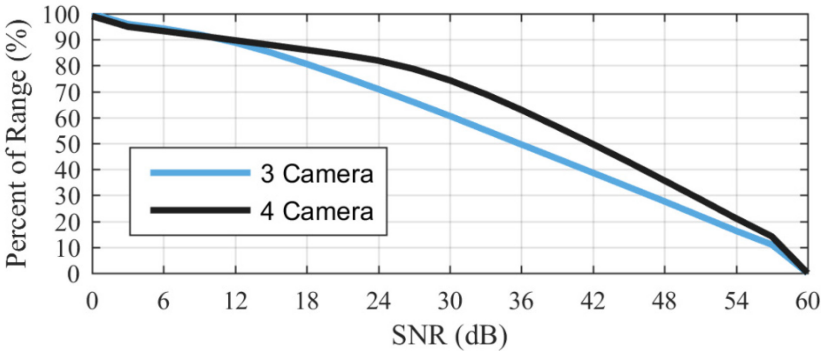


Figure 5-7. Comparison of the simulated SNR coverage for the acquisition system.

The four camera system will have a greater than 24 dB SNR for 80% of the acquirable range and the three camera system will have 70% of the acquirable range greater than 24 dB. Therefore the expectation is to have less noise in the four camera system when compared to the three camera system. Figure 5-8 shows the actual non-dimensional radiance compared to the reconstructed non-dimensional HDR radiant exitance. The four camera and three camera system have an error under 6 dB for the acquirable range, aside from the initial 10 dB. This is caused by the base noise floor of the cameras. The three camera system has a region of error around -50 dB and -100 dB created by the base noise floor of the higher attenuation cameras. In all cases this will cause an over estimation of the actual radiant exitance value. Therefore it is expected to see some regions in the final image that are unrealistically bright when the video traverses this radiant exitance range. The four camera system does not exhibit this same characteristic and therefore should produce an accurate video, but at the expense of added equipment.

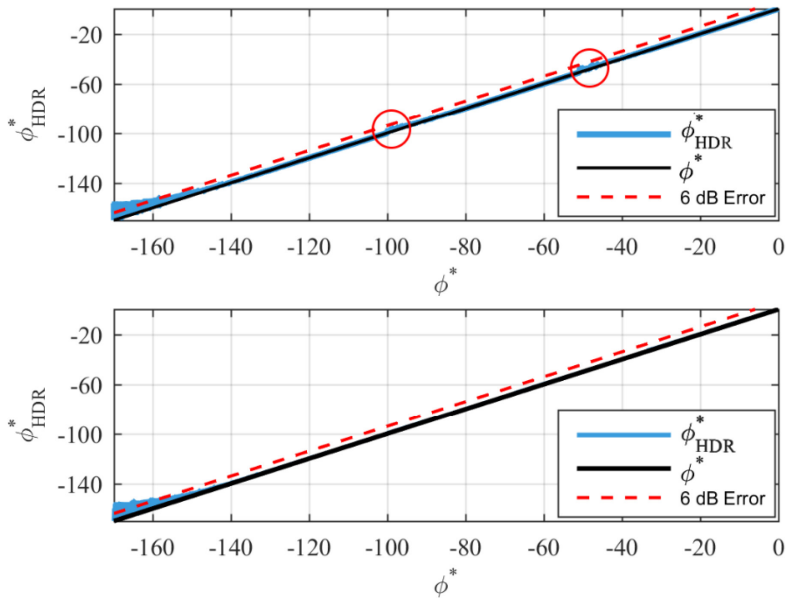


Figure 5-8. Top. Comparison of the HDR recreation and the expected value for the three camera system in units of dB. The circles help visualize the ranges the HDR image's noise reaches the 6 dB error line. Bottom. Comparison of the HDR recreation and the expected value for the four camera system in dB.

Results

The laboratory tested camera architecture was designed to have a 170 dB dynamic range. The cameras were synchronously triggered on the rising edge of a 5 V TTL signal generated by an Arduino Uno. The same trigger signal was used to trigger the flash bulb in the Photogenic 2500 DR. The cameras were time and frame synchronized to IRIG-B GPS timing providing a temporal accuracy of 10 ns. The maximum exposure time that could be used due to the acquisition rate of 79,000 pps was 12.2 µs. It was found the exposure time oversaturated the 170 dB range, therefore the exposure time was reduced to 3.2 µs to keep the event within the designed dynamic range. Additional lighting was provided by a halogen source focused by a Fresnel lens so the bulb could be seen at a time before initiation. The halogen light source amplitude was reduced until the bulb structure was resolvable and the video would span the full 170 dB dynamic range.

The data was recorded for 54 ms. The time amplitude histogram can be seen in Figure 5-9. The maximum amplitude was achieved in 8 µs for both three and four camera tests. Half brightness occurred after 89 µs for both tests. The amplitude reduced to 1% of the maximum

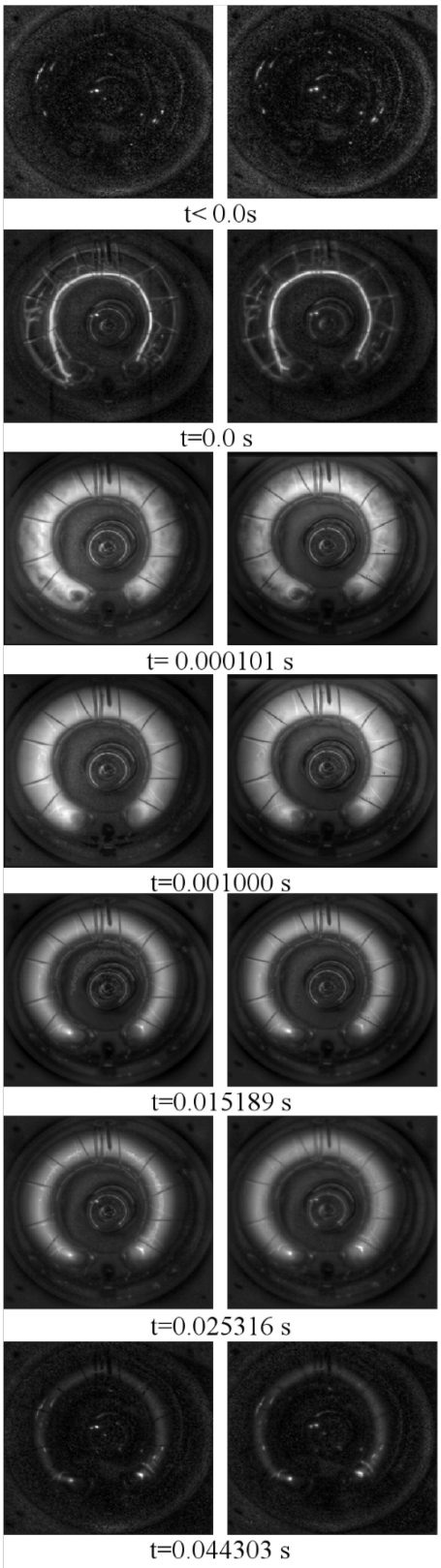


Figure 5-10. Left. 3 camera video frames. Right. 4 camera video frames.

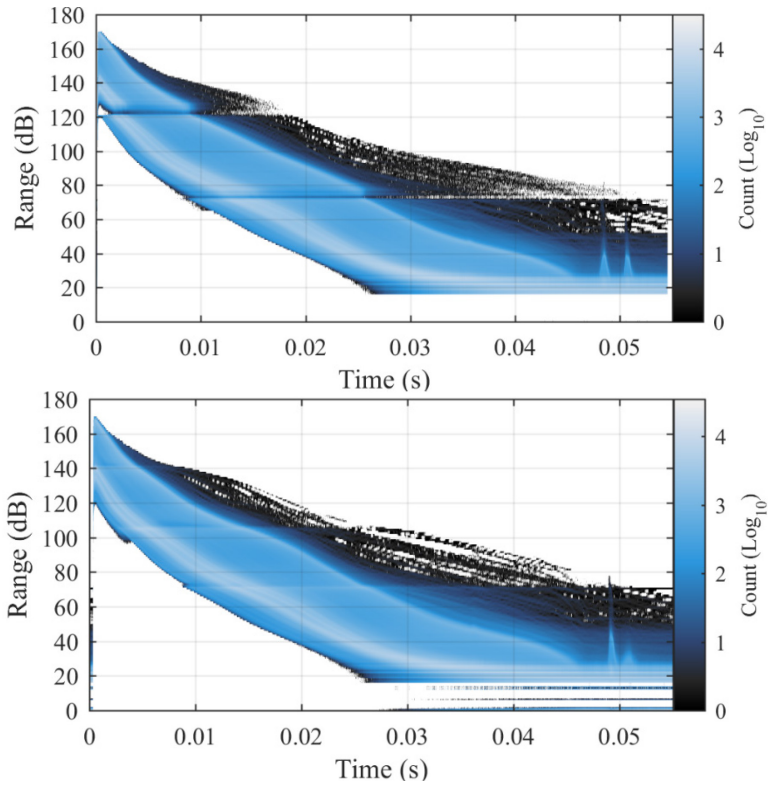


Figure 5-9. Comparison of the three camera (top) and four camera (bottom) system time amplitude and range histograms.

after 14.6 ms for the four camera system and 15.4 ms during the three camera test. The maximum amplitudes were estimated at 3.7×10^7 W/m^2 for the four camera test and 3.8×10^7 W/m^2 for the three camera test. With an exposure time of 3.2 μs and active pixel area of 22 nm^2 the max total energy recorded is estimated at 26 nJ. The estimated minimum resolvable radiant exitance was 0.119 W/m^2 for this camera configuration and exposure time.

Seven frames were tone mapped utilizing the Fattel [80] tone map operator for the purpose of display within an 8 bit intensity space, seen in Figure 5-10. There are several differences between the three and four camera system as predicted by Figure 5-8 and readily observed in Figure 5-9 and Figure 5-10. There is a discontinuous gap in the intensity information in the three camera system around

120 dB and 70 dB, corresponding to the errors predicted at -50 dB and -100 dB. The errors can be seen in the tone mapped images in the space surrounding the center illumination bulb. In the 5th row, t=0.015189 s, a bright region can be seen disproportionally in the three camera system, and not in the four camera system. In the 6th row another amplitude error can be seen in the flash bulb in the three camera system and not the four camera system. Other than the predicted image error the videos have similar visual attributes, only with slightly more noise in the three camera system compared to the four camera system.

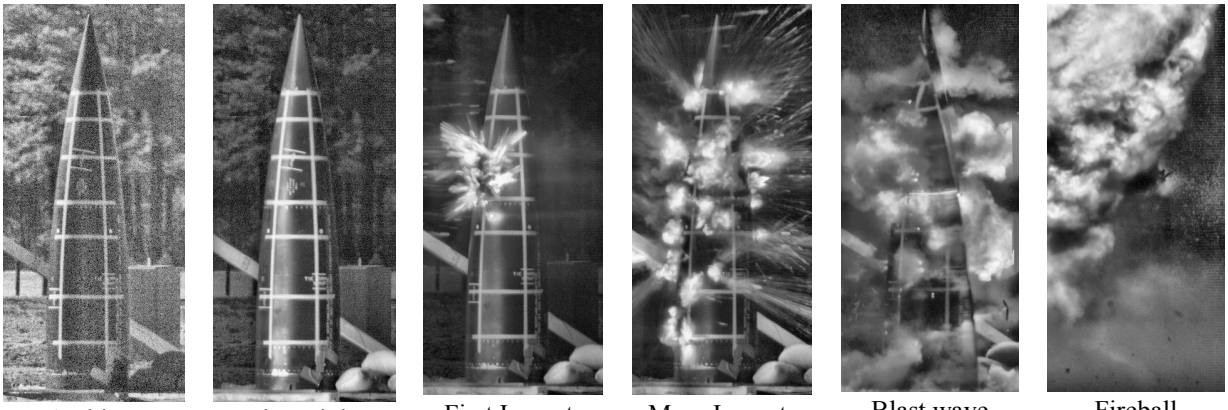


Figure 5-11. HDR Video frames at 34000 fps of fragments impacting a cylindrical test form.

As of the writing of this text the current system has been used to record eighty seven energetic test events. The dynamic range of each test and the cameras used have varied significantly ranging up to 27 stops or 162 dB throughout the video with four cameras in use. Currently the system has used with Vision Research Phantom 7.1, 7.3, 711, and 641 cameras. The minimum focal length used has been 135 mm and the maximum 400mm. An example of an energetic test can be seen in Figure 5-11. In this test fragments are explosively propelled toward

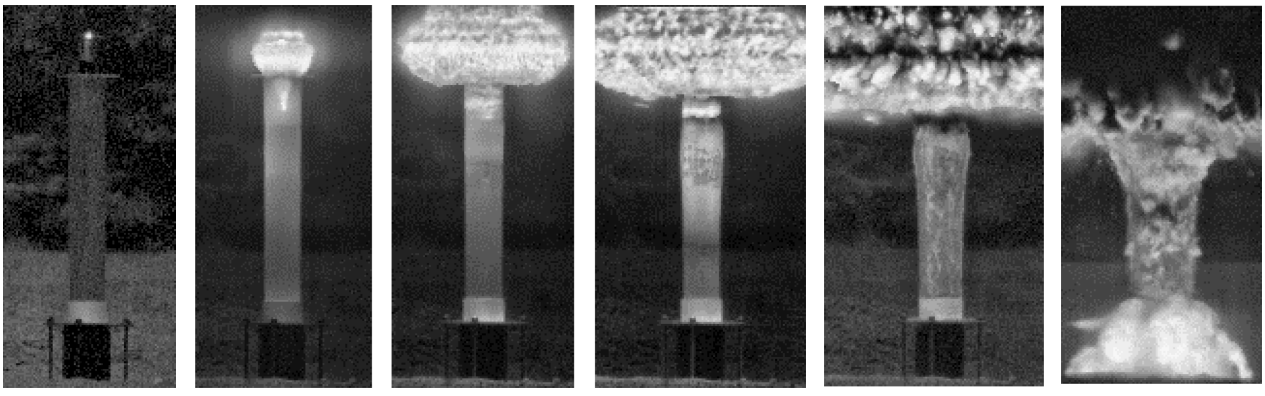


Figure 5-12. Frames from a high speed HDR video of a shape charge detonation, viewed through an acrylic tube, recorded at 140,000 fps.

an object and impact points noted. The data was recorded with four Vision Research Phantom 711 cameras running at 34000 fps.

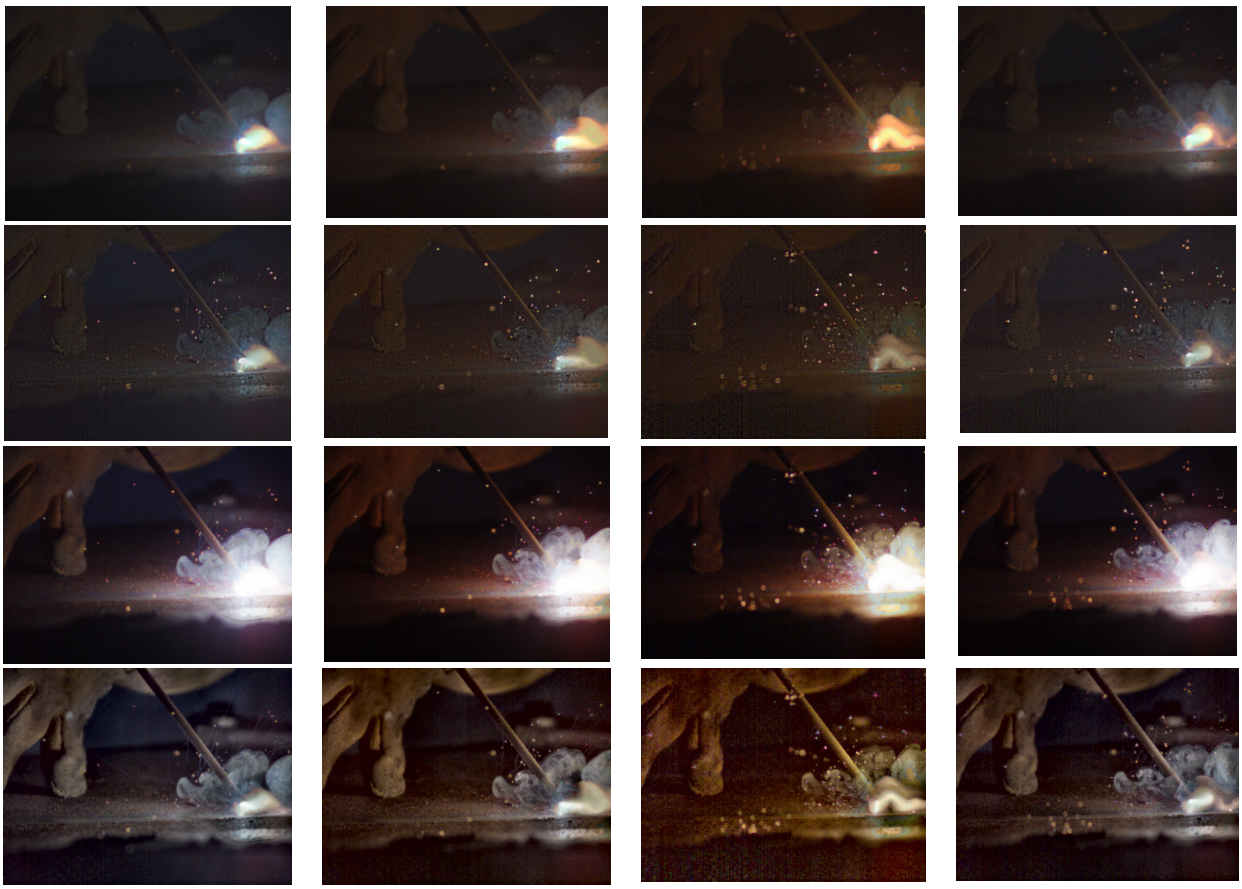


Figure 5-13. Arc weld video created with EMM method tone mapped via: Top Row. Griffiths Global. Top Middle. Griffiths Local. Bottom Middle. Reinhard. Bottom. Matlab's Histogram equalization method.

Figure 5-12 is another example of an energetic test filmed at 140,000 fps using four Vision Research Phantom 711 cameras. Frequently the range of the event is not known *a priori* so a best guess must be made from prior experience. If the range is known, then the camera system can be deigned to maximize the SNR coverage with the number of available cameras.

The dynamic range in this video was in excess of 110 dB. The initial frame is illuminated by ambient daylight seen in the first frame. The background is the ground and trees at the test site and the foreground is the test object.

Another example of video frames taken from a high speed video can be seen in Figure 5-13. There are four time frames from the recorded video presented in each row. Each column is the same time step and each row represents a different tone map method. The first row utilizes the mapping introduced in this document presented in Chapter 4 as its global variation. The

second row utilizes the tone mapping method introduced in Chapter 4 in the spatial variant form. The third row was mapped using Reinhard's local tone mapping method and the fourth row utilizes Matlab's built in method.

Conclusion

A high speed high dynamic range video system has been introduced utilizing a minimal number of cameras. A methodology to optimize the camera settings in a multi imager based video system has been derived. It was found the optimized camera spacing can be found with only knowledge of the number of quantization steps of each sensor and the dynamic range requirement. The maximum dynamic range a camera system with multiple sensors can accurately reproduce is limited to the number of quantization steps in the A/D and the number of cameras. An extended dynamic range can be created beyond the base SNR, however there will be bright discontinuous regions in the tone mapped video. Depending on the application, the reduction in assets may outweigh the need for added visual accuracy.

A camera system was introduced and a high speed HDR video was formed utilizing both three and four Vision research Phantom 711 cameras. The scene spanned a maximum of 169 dB dynamic range and was filmed at 79,000 pps demonstrating the concept of high speed HDR video with a reduced number of cameras. With the three camera system the added noise created from the read noise and dark current noise appeared in the recreated video was predicted and visualized in Figure 5-9.

Chapter 6. Concluding Remarks and Future work

In this document the history of high speed videography and photography has been reviewed. The chronological placement of this effort has been suggested as an evolutionary step in the development of high dynamic range high speed videography. In order to create a video with a minimal number of cameras at very high rates, and display them in a perceptually correct manner four primary steps were taken beyond the current state of the art.

First a simplified weighting function was introduced. The ideal weighting function was defined as the reciprocal of the variance of the error on the formation of the image. An engineering estimate of the variance associated with the image formation process was introduced based on a simplified camera model. An order of magnitude analysis was presented to define what parameters govern the variance throughout the majority of the formation process. A new weighting function was introduced that is only a function of the exposure time, physical attenuation and sensor gain. All of the needed parameters can be readily defined for any camera and lends to future real time implementation. This weighting function was compared to eleven other weighting functions that prior authors have proposed and the effectiveness to reduce the error was demonstrated relative to the other methods. In all cases the proposed method was within a 2.3 dB margin of error relative to the optimal recursive methods.

The second contribution presented in this work is a stochastic image formation methodology. A formation methodology was introduced based on a decision matrix on how to form the final image. The decision matrix weighed the probability of intersection with the uncertainty of formation out to a user specified confidence interval to decide which camera set would be idea for final image formation. This was done in an effort to detect non-symmetrical errors such as veiling glare or misalignment in the presence of extremely limited samples. These errors often present color artifacts or harsh discontinuities within the final image that do not originate from the scene being viewed. The method was tested with similar lenses and different lenses. It was shown how, in the presence of severe bias error, a more accurate image could be resolved when compared to the standard formation method using the proposed error minimization method. With the test scene presented, a greater than 500% reduction in the bias error was measured when the newly developed error minimization method was used to form the image when compared to the standard error formation method. Significant reductions in veiling glare and intraocular lens reflections can be observed when using the newly proposed use of

dissimilar lenses. The varying lens with multiple imagers forces nonlinear bias errors and allow the error minimization method to detect the glare and flaring pattern for removal.

The third contribution was the introduction of a new tone mapping methodology. The medical literature was reviewed and a function was derived for the light adapted semi-saturation parameter. The semi-saturation parameter controls the retinal response to an incremental change in the luminance relative to the background luminance. The semi saturation parameter is guided by both the photo bleaching phenomena and the cellular adaptation mechanism reviewed by Valeton. The new function for the semi saturation parameter was combined with the Naka-Rushton equation to calculate the normalized retinal response. The normalized retinal response was then used to replace the perceptual brightness and remapped into the standard RGB color space for final display. This method was demonstrated in both a spatially variant and spatially invariant methodology. The introduced tone map operators were shown and compared to the seven other methods for various standard scenes commonly used in the high dynamic range literature.

The fourth and final contribution is the design, demonstration, and optimization of a high dynamic range video system. The governing equations that model the image formation method from Chapter 2 were non-dimensionalized. The non-dimensional parameters were then optimized utilizing an analysis of experiments optimization. This optimization method was chosen since the governing equations are not continuous and a numerical method was needed. The ideal camera settings were found and formulated into a simple equality. The optimized settings were used to demonstrate the construction of a 170 dB dynamic range video system. The video was captured at 79,000 fps to visualize the excitement of a high output strobe bulb being energized whih spanned a measured 169 dB range. This was shown with both a three and four camera system, demonstrating the advantages and disadvantages of creating a high speed high dynamic range video with a limited number of cameras. To the best of the author's knowledge, this video is the highest speed digital high dynamic range video known to be produced to date. Subsequent fielded studies were then demonstrated with varying camera configurations, one running up to 140,000 fps.

In the proceeding photo series, seen in Figure 6-1, a final demonstration is given that utilizes every aspect presented in this document and concludes the flash bulb demonstrations observed at the end of each chapter. A high speed video was captured utilizing a 160 dB dynamic range using four Vision Research 641 cameras, with only 126 dB range being used.

The error minimization method was used to reform the captured video frames into the estimated radiant exitance values. The global tone map operator, introduced in Chapter 4, was used to reform the data into an 8 bit color image. The data was rescaled such that it would be perceived as if the light was room brightness and remapped with a gamma of 2.22 for final display. This video spans a total range of 126 dB, about 2,000,000:1. The time amplitude histogram for the full video can be seen in Figure 6-2. The corresponding formation methods that were used to form each pixel can be seen in the ribbon plot in Figure 6-3 with the corresponding camera sets listed in the Appendix. This figure demonstrates how the error minimization process changed for each frame as the scene evolved, demonstrating how the error minimization method works to mitigate bias errors.

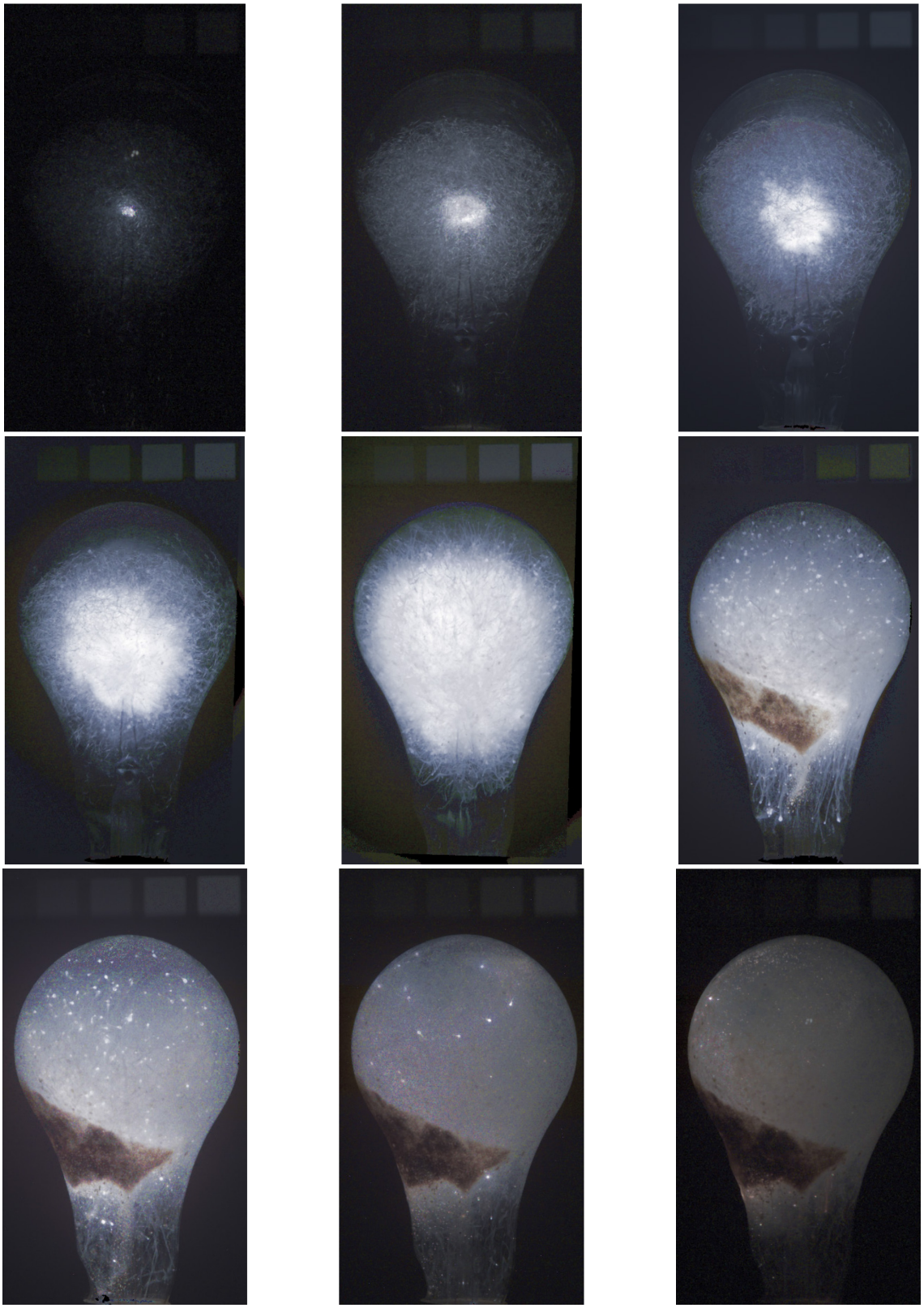


Figure 6-1. Flash bulb high speed video frames using the EMM formation, ideal camera spacing, and the new global tone map operator filmed at 5300 fps using four Phantom 641 cameras.

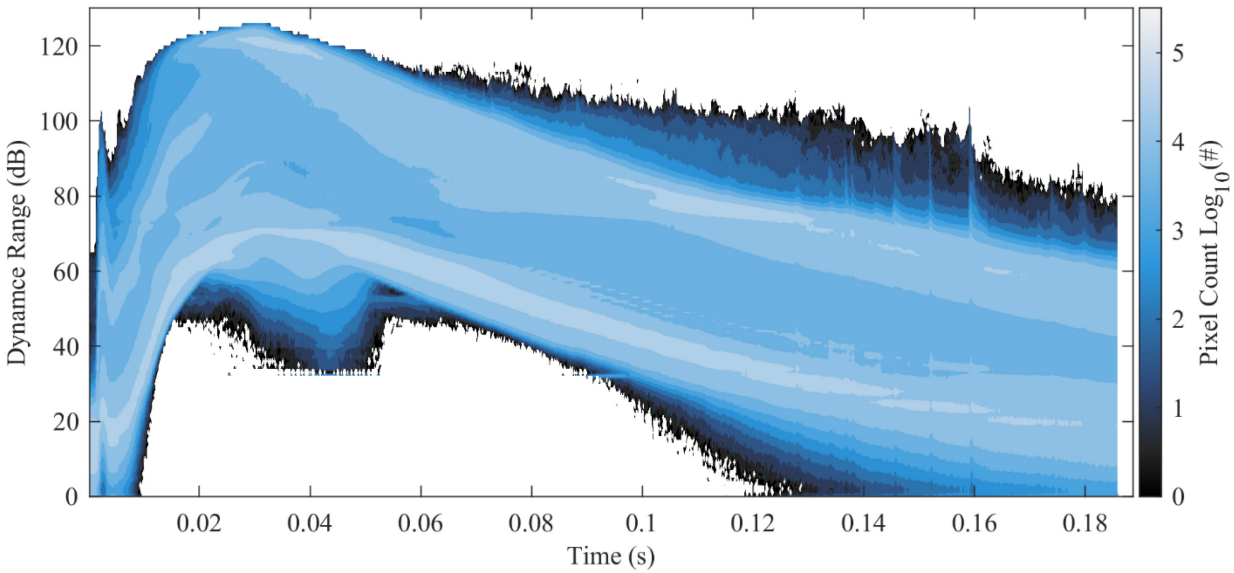


Figure 6-2. Time amplitude histogram of the flash bulb video demonstrating the dynamic range of the captured data.

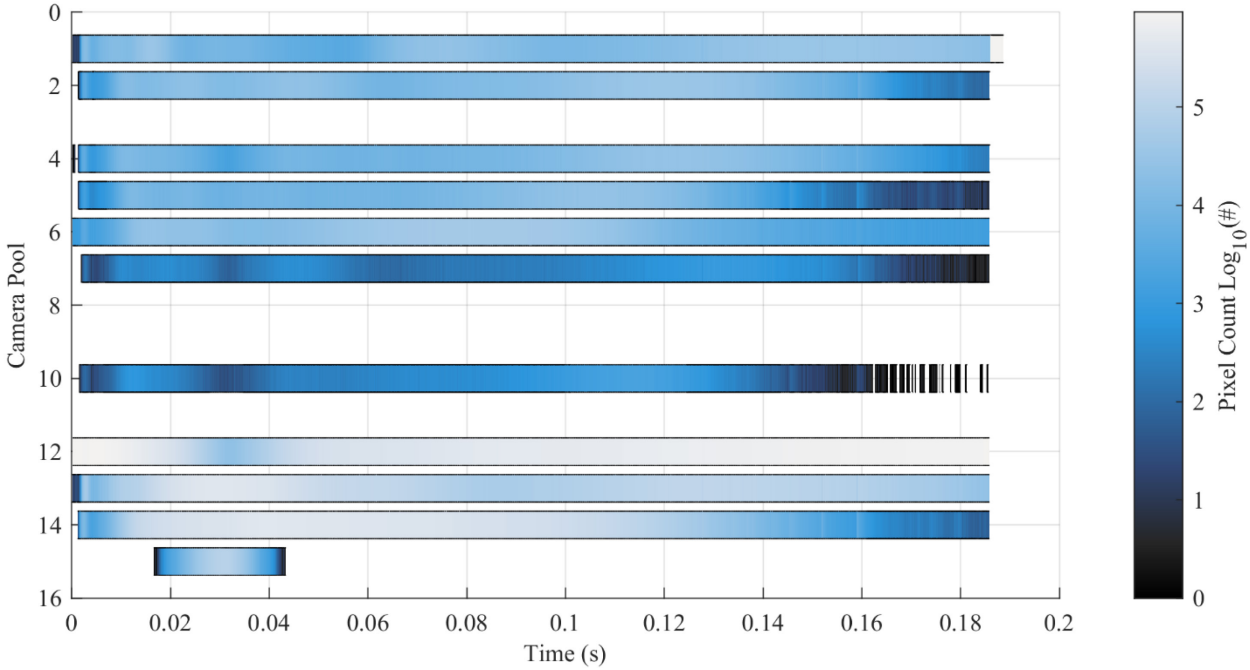


Figure 6-3. Histogram of the formation method used to form the image. The method changes as the scene changes for flare and misalignment mitigation (see Appendix for pool number to set translation).

Future Work

While this study was focused on the formation of the image and the display of a video in a perceptually correct way, the process that has been described takes time to compute. Currently it is a series operation performed for every pixel for every color channel. Operating on a single luminance channel was explored but the degradation in the final image was deemed too significant of a tradeoff for the speed that was recovered. A significant amount of the operations can be massively parallelized as they are not dependent on their neighboring pixels. Such an effort could reduce the processing time considerably, which is ultimately needed for widespread adaptation and implementation.

A potential method to improve the estimate would to look at how a pixel changes in time and the spatial neighbors. Potentially this information could be used to collect more samples in estimating the true mean.

The error minimization method introduced the concept of forcing nonlinear bias errors by using different lenses in the image acquisition chain. However the optimal characteristics of each lens to mitigate the bias error has not be investigated. It has been observed that various combinations of lenses reduce the error on formation more than other combinations, suggesting there could be optimal arrangement for example a lens system with alternating even and odd aperture blades, or varying number of blades and lensing elements.

Currently the tone map operators operate on a per frame basis. As previously discussed in the introduction, there is a temporal response to the human vision's system. The temporal response is not modeled in the current implementation, however it could be. This would entail the temporal adaptation mechanism and could smooth out some flicker in the final video formation. As it stands there is an assumption that there is an instantaneous adaptation, which can lead to abrupt changes in the mapping of the intensity values, which can look abnormal.

There are many extensions of this work into other scientific data analysis fields. One such extension is the near wall digital particle image velocimetry measurements. One significant problem in this field is near wall reflections that oversaturate the imager of the high speed camera. The author hypothesis this method could be used to reduce this problem. Extensions have been demonstrated in the Appendix to the X-Ray and IR fields as well as demonstrated by an HDR reconstructed image of a cordless drill and the HDR reconstruction of IR images viewing a running truck.

References

[1] Dolby Vision, "www.dolby.com," 2016. [Online]. Available: www.dolby.com/us/en/technologies/dolby-vision/dolby-vision-white-paper.pdf. [Accessed 1 11 2016].

[2] B. E. Bayer, "Color imaging array". USA Patent US 3971065 A, 20 Jul 1976.

[3] G. E. Healey, "Radiometric CCD Camera Calibration and Noise Estimation," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 16, no. 3, pp. 267-276, March 1994.

[4] H. Tian, "Noise Analysis in CMOS Image Sensors," Stanford Univeristy, 2001.

[5] P. J. Winzer, "Shot-noise formula for time-varying photon rates: a general derivation," *Journal of the Optical Society of America,* vol. 14, no. 10, pp. 2424-2429, 1997.

[6] F. E. Toole, Sound Reproduction, Oxford: Elsevier Ltd., 2009.

[7] M. Pharr and G. Humphreys, Physically Based Rendering: From Theory to Implementation, Burlington, MA: Morgan Kaufmann, 2010.

[8] G. Wyszecki and W. S. Stiles, Color Science: Concepts and Methods, Quantitative Data and Formulae, New York: John Wiley & Sons Inc., 1982.

[9] SMPTE, "Society of Motion Picture & Television Engineers," 2016. [Online]. Available: https://www.smpte.org/. [Accessed 2 11 2016].

[10] C. D'Arcy, "Sur la Durée de la Sensation de La vue," *Mémoires de l'Académie des Sciences de Paris,* p. 439–451, 1765.

[11] J. Plateau, Dissertation sur quelques propriétés des impressions produites par la lumière sur l'organe de la vue, Liège: Dessain, 1829.

[12] W. H. F. Talbot, "Experiments in Light," *Philosophical Magazine and Journal of Science,* vol. 5, pp. 321-334, 1834.

[13] W. Staunton, "The Prophet Louis Ducos Du Hauron and His Marvelous Moving Picture Machine," *Cinema Journal,* vol. 6, pp. 46-51, 1967.

[14] W. H. F. Talbot, "On the production of instantaneous photographic images," *Phil Mag.,* pp. 73-77, 1852.

[15] T. Skaife, Instantaneous Photography, H.S. Richardson, Greenwich, 1860.

[16] E. Mach and L. Mach, "Weitere ballistisch - photo-graphische Versuche.," *Mathematisch-Naturwisenschaft.,* pp. 1310-1332, 1889.

[17] A. M. Worthington, A Study of Splashes, London: Longmans, Green and Co., 1908.

[18] J. Cranz and B. Glatzel, "Photographic recording of ballistic and other rapid phenomena witht he aid of the quenched spark," *Phys. Gesell.,* vol. 14, pp. 525-535, 1912.

[19] S. F. Ray, High Speed Photography and Photonics, Oxford: Focal Press, 1997.

[20] W. Boyle and G. Smith, "Buried channel charge coupled devices". USA Patent US3792322 A, 19 Apr 1973.

[21] M. F. Tompsett, "Charge transfer imaging devices". USA Patent US 4085456 A, 30 Aug 1972.

[22] Photron, "Photron," 2016. [Online]. Available: http://photron.com/about-photron/company/. [Accessed 2 Nov

2016].

[23] A. Aciu, P. Pop and R. Corlan, "High-frame rate image acquisition and motion analysis system". USA Patent 5,625,412, 13 Jul 1995.

[24] Photron, "FASTCAM SA5," 2016. [Online]. Available: http://photron.com/high-speed/cameras/fastcam-sa5/. [Accessed 2 Nov 2016].

[25] Vision Research, "v711," 2016. [Online]. Available: https://www.phantomhighspeed.com/Products/Ultrahigh-Speed-Cameras/v2512. [Accessed 2 Nov 216].

[26] Vision Research, "EDR - Extreme Dynamic Range," 2016. [Online]. Available: https://www.phantomhighspeed.com/Portals/0/Files/Phantom%20Features/Notes_WEB-EDR.pdf. [Accessed 2 Nov 2016].

[27] A. B. Watson and J. I. Yellot, "A unified formula for light-adapted pupil size," *Journal of Vision,* vol. 12, no. 10, pp. 1-16, 2012.

[28] E. Reinhard, G. Ward, S. Pattanaik, P. Debevec, W. Heidrich and K. Myszkowski, in *High dynamic range imaging*, Elsevier, 2010.

[29] C. W. Wyckoff, "An experimental extended exposure responce film," in *SPIE Newsletter*, 1962.

[30] S. Mann and R. W. Picard, "On being 'undgital' with digital cameras: Extending dynamic range by combining differently exposed pictures," M.I.T. Media Lab Perceptual Computing Section, Boston, Massachusetts, 1995.

[31] A. O. Akyuz and A. Genctav, "A reality check for radiometric camera responce recovery algorithms," *Computers and Graphics,* pp. 935-943, July 2013.

[32] P. Debevec and J. Malik, "Recovering High Dynamic Range Radiance Maps from Photographs," *SIGGRAPH,* 1997.

[33] T. Mitsunaga and S. K. Nayar, "Radiometric self calibration," *Computer Vision and Pattern Recognition,* vol. 1, 1999.

[34] M. A. Robertson, S. Borman and R. L. Stevenson, "Estimation-theoretic approach to dynamic range enhancement using multiple exposures," *Journal of Electronic Imaging,* vol. 12, no. 2, pp. 219-228, 2003.

[35] M. D. Grossberg and S. K. Nayar, "Modeling the space of camera responce functions," *Pattern Analysis and Machine Intelligence,* vol. 26, no. 10, pp. 1272-1282, 2004.

[36] E. Dedrick and D. Lau, "A Kalman-Filtering Approach to High Dynamic Range Imaging for Measurment Applications," *IEEE Transactions on Image Processing,* vol. 21, no. 2, pp. 527-535, 2012.

[37] M. A. Robertson, S. Borman and R. L. Stevenson, "Dynamic range improvement through multiple exposures," *IEEE,* pp. 159-163, 1999.

[38] Y. Tsin, V. Ramesh and T. Kanade, "Statistical Calibration of CCD Imaging Process," in *Computer Vision*, 2001.

[39] E. Reinhard, G. Ward, S. Pattanaik and P. Debevec, High Dynamic Range Imaging: Acqusition, Display and Image-Based Lighting, San Franciso: Morgan Kaufmann Publishers, 2005.

[40] K. Kirk and H. J. Anderson, "Noise characterization of weighting schemes for combination of multiple exposures," in *Proceedings of British Machine Vision Conference*, 2006.

[41] A. O. Akyuz and E. Reinhard, "Noise reduction in high dynamic range imaging," *Journal of Visual Communication and Image Representation,* vol. 18, pp. 366-376, 2007.

[42] M. Granados, B. Ajdin, M. Wand and C. Theobalt, "Optimal HDR Reconstruction with Linear Digital

Cameras," *Computer Vision and Pattern Recognition ,* pp. 13-18, June 2010.

[43] S. W. Hasinoff, F. Durand and W. T. Freeman, "Noise-Optimal Capture for High Dynamic Range Photography," pp. 553-560, 2010.

[44] G. Ward, "Fast, Robust image registration for compositing high dynamic range photographs from hand-held exposures," *Journal of Graphics Tools,* vol. 8, no. 2, pp. 17-30, 2012.

[45] R. D. Fiete, Modeling the Imaging Chain of Digital Cameras, Bellingham: SPIE, 2010.

[46] ON Semiconductor, "AND9189/D," January 2016. [Online]. Available: www.onsemi.com. [Accessed January 2016].

[47] M. D. Tocci, C. Kiser, N. Tocci and P. Sen, "A versatile HDR video production system," in *ACM Transactions on Graphics*, 2011.

[48] M. Aggarwal and N. Ahuja, "Split aperture imaging for high dynamic range," *International Journal of Computer Vision,* vol. 58, no. 1, pp. 7-17, 2004.

[49] N. Bromberg and J. Bickford, "Veiling glare in the imaging chain," *Application of Optical Instrumentation in Medicine ,* vol. 454, 1984.

[50] J. Kronander, S. Gustavson, G. Bonnet, A. YnnermN and J. Unger, "A Unified Framework for Multi-Sensor HDR Video Reconstruction," *Signal Processing: Image Comminications,* 2013.

[51] E. Selmanovic, K. Debattista and T. Bashford-Rogers, "Enabling steroscopic high dynamic range video," *Signal Processing: Image Communication,* vol. 29, pp. 216-228, 2014.

[52] S. K. Nayar and T. Mitsunaga, "High Dynamic Range Imaging: Spatially Varying Pixel Exposures," *Computer Vision and Pattern Recognition,* vol. 1, pp. 472-479, 2000.

[53] B. S. Eastwood and E. C. Childs, "Image Alignment for multiple camera High Dynamci range Microscopy," in *IEEE Workshop on Applications of Computer Vision*, 2012.

[54] J. E. Harvey, "Parametric analysis of the effect of scattered light upon the modulation transfer function," *Optical Engineering,* vol. 52, no. 7, 2013.

[55] . K. Jacobs, C. Loscos and G. Ward, "Automatic high-dynamic range image generation for dynamic scenes," *IEEE Computer Graphics and Applications,* vol. 28, no. 2, pp. 84-93, 2008.

[56] T. Grosch, "Fast and robust high dynamic range image generation with camera and object movement," *Vision, Modeling and Visualization,* pp. 277-284, 2006.

[57] D. Sidibe, W. Puech and O. Strauss, "Ghost detection and removal in high dynamic range images," in *Signal Processing Conference*, 17th European, 2009.

[58] M. A. Fsichler and R. C. Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," *Communications of the ACM,* vol. 24, no. 6, pp. 381-395, 1981.

[59] Y. S. Heo, K. M. Lee and S. U. Lee, "Ghost-free high dynamic range imaging.," in *Asian Conference on Computer Vision*, Springer Berlin Heidelberg, 2010.

[60] A. Srikantha and D. Sidibé, "Ghost detection and removal for high dynamic range images: Recent advances," *Signal Processing: Image Communication,* vol. 27, pp. 650-662, 2012.

[61] J. J. McCann and A. Rizzi, "Veiling glare: the dynamic range limit of HDR images," in *Proc. SPIE 6492, Human Vision and Electronic Imaging XII*, San Jose, CA, 2007.

[62] S. Naimuddin, B. Hasegawa and C. A. Mistretta, "Scatter-glare correction using a convolution algorithm with

variable weighting," *Medical physics,* vol. 14, no. 3, pp. 330-334, 1987.

[63] A. J. Seiber, O. Nalcioglu and W. Roeck, "Removal of image intensifier veiling glare by mathematical deconvolution techniques," *Medical physics,* vol. 12, no. 3, pp. 281-288, 1985.

[64] M. Hullin, E. Eisemann, H. Seidel and S. Lee, "Physically-based real-time lens flare rendering," *ACM Transactions on Graphics,* vol. 30, no. 4, 2011.

[65] E.-V. Talvala, A. Adams, M. Horowitz and M. Levoy, "Veiling Glare in High Dynamic Range Imaging," *ACM Transactions on Graphics,* vol. 26, no. 3, 2007.

[66] M. Bullock, "The Secrect Garden of Math," 2013. [Online]. Available: www.mickybullock.com/blog/2013/10/statistical-outliers-impossible-in-small-samples/. [Accessed 10 9 2015].

[67] H. H. Ku, "Notes on the use of propagation of error formulas," *Journal of Research of the National Bureau of Standards,* vol. 70, no. 4, 1966.

[68] F. S. Woods, Advanced Calculus, Boston, USA: Ginn and Company, 1934.

[69] D. J. Griffiths and A. Wicks, "Simplified weighting function for high dynamic range video frame formation," *Journal of Applied Optics,* vol. 55, no. 31, pp. C9-C17, 2016.

[70] S. N. Pattanaik, J. E. Tumblin, H. Yee and D. P. Greenberg, "Time-Dependent Visual Adaptation for Realistic Real-Time Image Display," in *Proceedings of SIGGRAPH 2000*, New Orleans, 2000.

[71] D. C. Hood and M. A. Finkelstein, "Psychophysical Tests of Models of the Response Function," *Vision Research,* vol. 19, pp. 401-406, 1979.

[72] E. Reinhard and K. Devlin, "Dynamic Range Reduction inspired by Photoreceptor Physiology," in *IEEE transaction on visualization and computer graphics*.

[73] R. Mantiuk, K. Myszkowski and H. P. Seidel, "A perceptual framework for contrast processing of high dynamic range images," *ACM Transaction on Applied Perception,* pp. 286-308, 2006.

[74] J. Tumblin and H. Rushmeier, "Tone Reproduction for Realistic Images'," *Computer Graphics and Applications,* pp. 42-48, 1993.

[75] D. Lischinski, Z. Farbman, M. Uyttendaele and R. Szeliski, "Interactive Local Adjustment of Tonal Values," *ACM Trans. Graph.,* vol. 25, pp. 646-653, 2006.

[76] F. Durand and J. Dorsey, "Fast bilateral filtering for the display of high-dynamic-range images," *AXM transactions on graphics,* vol. 21, no. 3, pp. 257-266, 2002.

[77] H. Ahn, B. Keum, D. Kim and H. S. Lee, "Adaptive Local Tone Mapping Based on Retinex for High Dynamic Range Images," in *IEEE International Conference on COnsumer Electronics*, 2013.

[78] N. Moroney, M. D. Fairchild, R. Hunt and C. Li, "The CIECAM02 Color Appearance Model," in *Color and Imaging Conference*, 2002.

[79] M. H. Kim and J. Kautz, "Consistent Tone Reproduction," 2008.

[80] R. Fattal, D. Lischinski and M. Weerman, "Gradient domain high dynamci range compression," *ACM Transactions on Graphics,* vol. 21, no. 3, pp. 249-256, 2002.

[81] T. O. Aydin, N. Stefanoski, S. Croci, M. Gross and A. Smolie, "Temporally Coherent Local Tone Mapping of HDR Video," *Transactions on Graphics,* 2004.

[82] F. Rieke and D. A. Baylor, "Single-photon detection by rod cells of the retina.," *Reviews of Modern Physics,* vol. 70, no. 3, pp. 1027-1037, 1998.

[83] L. Holladay, "The fundamentals of glare and visibility," *Journal of the Optical Society of America,* vol. 12, no. 4, pp. 271-319, 1926.

[84] B. H. Crawford, "The dependence of pupil size upon external light stimukus under static and variable conditions," in *Proceedings of the Royal Society of London*, 1936.

[85] P. Moon and D. E. Spencer, "On the Stiles-Crawford effect," *Journal of the OPtical Society of America,* vol. 34, no. 6, pp. 319-329, 1944.

[86] S. G. De Groot and J. W. Gebhard, "Pupil size as determined by adapting luminace," *Journal of the Optical Society of America,* vol. 42, no. 7, pp. 492-495, 1952.

[87] A. P. Stanley and K. Davies, "The effect of field of view size on steady-state pupil diameter," *Ophthalmic and Physiological Optics,* vol. 15, no. 6, pp. 601-603, 1995.

[88] D. C. Hood and M. Finkelstein, "Sensitivity to light," in *Basic Sensory Processes*, 1986.

[89] D. Marr, Vision: A computational Investigation into the Human Representation and Processing of Visual Information, Cambridge: MIT, 2010.

[90] G. Osterberg, "Topography of the Layer of Rods and Cones in the Human Retina," *The Journal of the American Medical Association,* vol. 108, no. 3, 1937.

[91] K.-W. Yau, "Phototransduction mechanism in retinal rods and cones," *Invest Opthalmol Vis Sci,* vol. 35, pp. 9-32, 1994.

[92] D. G. Birch, D. C. Hood, K. G. Locke, D. R. Hoffman and R. T. Tzekov, "Quantitative Electroretinogram Measures of Phototransduction in Cone and Rod Photoreceptors," *Archive of Opthalmology,* vol. 120, pp. 1045-1051, 2002.

[93] W. H. Rushton,, "Increment threshold and dark adaptation," *JOSA,* vol. 53, no. 1, pp. 104-109, 1963.

[94] S. K. Shevell, "Saturation in Human Cones," *Vision Research,* vol. 17, pp. 427-434, 1977.

[95] L. S. Evans, S. N. Peachey and A. L. Marchese, "Compairison of three methods of estimating the parameters of the Naka-Rushton equation," *Documenta Opthalmologica,* vol. 84, pp. 19-30, 1993.

[96] R. W. Nygaard and E. T. Frumkes, "Calibration of the Retinal Illimination Provided by Maxwellian views," *Vision Research,* vol. 22, p. 433, 1982.

[97] J. M. Valeton and D. van Norren, "Light adaptation of Primeate Cones: An Analysis Based on Extracellular Data," *Vision research,* vol. 23, no. 12, pp. 1539-1547, 1983.

[98] R. M. Boynton and D. N. Whitten, "Visual adaptation in monkey cones: recordings of the late receptor potentials," *Science,* pp. 1423-1426, 1970.

[99] H. S. Fairman, M. H. Brill and H. Hemmendinger, "How the CIE 1931 color-matching functions were derived from Wright-Guild data," *Color Research and Applications,* vol. 22, no. 1, pp. 11-23, 1997.

[100] R. Hall, Illumination and Color in Computer Genereted Imagery, New York: Springer-Verlag, 1989.

[101] "MPI," Max Plank Institut informatik, [Online]. Available: http://resources.mpi-inf.mpg.de/hdr/gallery.html. [Accessed 8 11 2016].

[102] G. Ward, "Anyhere," [Online]. Available: http://www.anyhere.com/gward/hdrenc/pages/originals.html. [Accessed 8 11 2016].

[103] P. Debevec, "Recovering High Dynamic Range Radiance Maps from Photographs," Paul Debevec, [Online]. Available: http://www.pauldebevec.com/Research/HDR/. [Accessed 8 11 2016].

[104] S. Silk and J. Lang, "High dynamic range image deghosting by fast approximate background modelling," *Computers and Grpahics,* vol. 36, no. 8, pp. 1060-1071, 2012.

[105] K. B. Sing and W. M. Uyttendaele, "High Dynamic Range Video," in *ACM Transactions on Graphics*, 2003.

[106] T. Portz, L. Zhang and H. Jiang, "Random coded sampling for high speed HDR video," in *International Conference on Computational Photography*, Cambridge, MA, 2013.

[107] J. Culler, "NASA," [Online]. Available: http://www.nasa.gov/topics/shuttle_station/features/sts-134_launch_photo-video.html. [Accessed 2013].

[108] M. Gupta, D. Iso and S. K. Nayar, "Fibonacci Eposure Bracketing for High Dynamic Range Imaging," in *IEEE International Conference on Computer Vision*, 2013.

A. Appendix

In this section some material will be presented that was omitted from the main portion of the text. It is presented here as supporting material.

Imager Bloom Model

One form of error that was not discussed at length in the main body of text is imager blooming. It originates from the charge leakage from an over saturated potential well. The potential well will continue to generate charge in the presence of a photon flux, however, will no longer be able to store it, and so the charge flows over into the adjacent neighboring wells. There are several different methods of mitigation techniques to minimize this effect. The first is a lateral drain. The lateral drain allows excess charge to dissipate and not affect the adjacent pixels. This is commonly seen as either horizontal or vertical bands in an image where there are saturated pixels.

Another form of mitigation is a vertical drain. The vertical drain is located behind the sensor with a bias potential. Any overflow from the potential well gets trapped into this drain and prevents overflow into adjacent cells. Both methods can be effectively used to prevent blooming, however when these methods are used it reduces the fill factor and the sensitivity of the imager. In order to model this form of mitigation system, the method of mitigation and its effectiveness needs to be measured. This can be done by looking at the point spread of an over saturated pixel on the imager that utilizes a pinhole lens. Utilizing a pinhole lens effectively gives an infinite depth of field and mitigates any flares that can arise from lens imperfections. This can be accomplished by constructing a pinhole lens for the camera.

An example lens would a 50mm standoff and a 0.3mm diameter hole. This gives an equivalent f number of about 168 and will require a several second exposure with a low ISO settings. The incident light should be bright enough to over saturate the sensor, or the exposures must be taken long enough to over saturate the sensor. An HDR image can be created using multiple exposures and a calibration scene luminance can be found. The objective is to model how the imager handles both the spatial distribution and magnitude of the overflow charge from the wells. Some cameras will have a radial pattern and some will have a vertical or horizontal banding depending on their anti-blooming characteristics.

Once the pattern is identified the mitigation method can be identified. This can be simulated using a cellular automata simulation. Overall structure can be developed for any given pixel and its surrounding pixels. A Von Neumann or Moore window system can be used if there is no mitigation technique used for blooming. The rule system for this is simple and basically states a condition that a given cell cannot have more than a maximum number of electrons and if it does the remainder of the generated electrons will spill equally into the adjacent cells. This simulation can be carried out for pixels that are over saturated. Utilizing a Von Neumann window in a conservative system, the charge will exhibit a star like pattern that reaches out every 90° rotated 45°. If a Moore window is used, then a similar star pattern will emerge with the bands located at 0°, 90°, 180°, and 240°. If lateral or vertical dissipation is used then a new rule set can be implied that is not conservative. This rule set states that some percentage of the energy going in the defined spatial direction of the drain will be lost. What this leads to is a streak in the direction perpendicular to the direction of the drain. The final form of anti-bloom is the vertical drain. This is a drain where there is no spatial dependency since the excess charge flows to a ground plan that is located behind the well. This system will exhibit a similar pattern to the no dissipation case, however will require proportionally more light to exhibit the same pattern extents.

These simulations develop a pattern of blooming from oversaturation of a single pixel. In an image there may be several to thousands of pixels that are oversaturated and the response of the imager will be the summation of the effects from each pixel. To predict the bloom on an image the quanta of light must be known, therefore the HDR image must be formed to get an estimate of the amount of electrons generated by a given pixel. This information can then be used to find which pixels will be oversaturated by running the cellular automata simulation, once calibrated for the sensor, with an estimate of the amount of electrons generated by a given pixel. This information can then be used to find which pixels will be oversaturated by running the cellular automata simulation with found parameters from the measured blooming from the imagers. This information can be subtracted from the final HDR recreation algorithm. This process can be repeated until there is very little deviation from the previous prediction and the current prediction of the luminance. An example of the structure of a simulated imager bloom can be seen in Figure A-1.

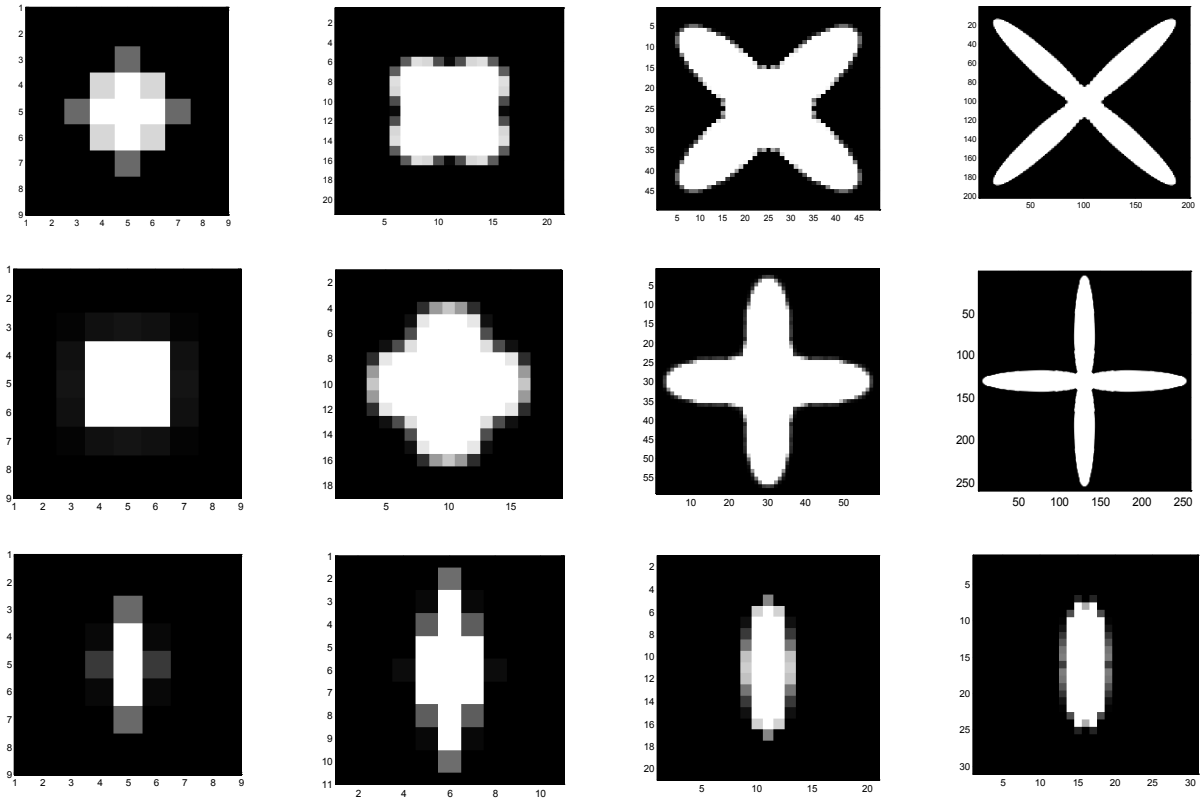


Figure A-1. Row 1 is a Von Neumann simulation with 10x, 100x, 1000x, and 10000x well capacity. Row 2 is a Moore simulation. Row 3 is a horizontal drain simulation

Characterizing the camera's noise characteristics

Throughout the document, various camera parameters are needed to perform the various tasks prescribed in this paper. The following is a description of how to find the camera parameters.

Dark Current

Dark current is caused by thermal agitation of the silicone. It generates free electrons and can be stored at the collection sites. It is proportional to the integration time and temperature. From the previous derivation we see

$$\varepsilon = V_{ga}(N_{dc} + N_s + N_r) \qquad \text{A-1}$$

where ε is the total error. If a cover is placed over the cap of the camera the shot noise goes to zero, therefore

$$V_{actual} = V_{ga}(N_{dc} + N_r) \qquad \text{A-2}$$

knowing the read noise is defined as

$$N_r = \overline{N_{dc}} - N_{dc} \quad \text{A-3}$$

substituting that into arrives at

$$V_{actual} = V_{ga}\overline{N_{dc}} \quad \text{A-4}$$

A series of dark frames averaged will allow for the acquisition of the dark current and dark current noise assuming unity gain. The dark current is the offset from zero seen in Figure A-2. The distribution about the mean is the dark current noise.

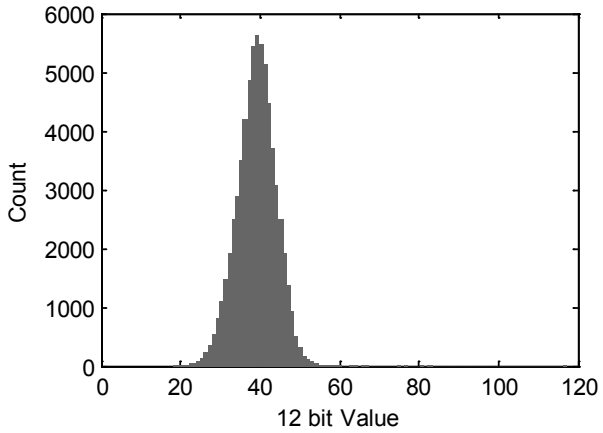


Figure A-2. Measured dark current from a Prosiclica 1900c camera.

Read Noise

The read noise is the noise induced from the internal circuitry from frame to frame. It is typically a zero mean Gaussian distribution. It can be found from imaging dark frames, finding the mean of the dark frames, then finding the difference between the mean and a subsequent capture. Using (A-3) an example of the measured read noise can be seen in Figure A-3. The example read noise was measured from a Prosilica 1900C machine vision camera.

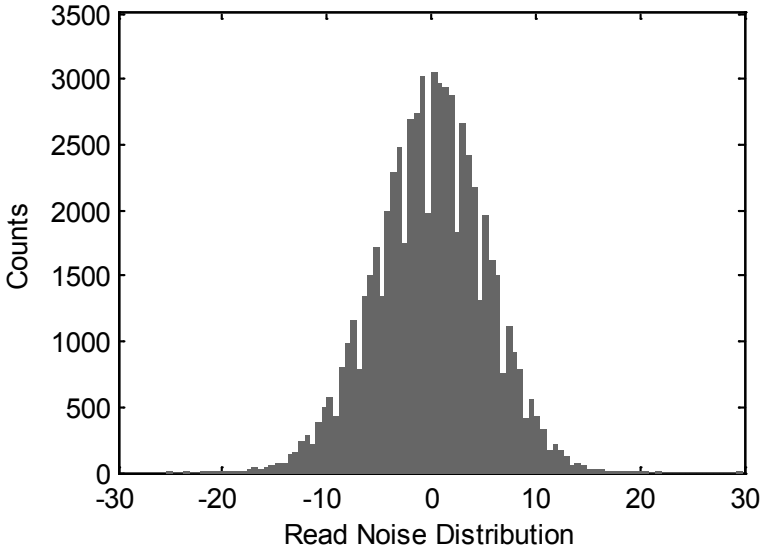


Figure A-3. Read noise from a Prosilica 1900c camera.

Camera Gain

The camera gain can be determined by fitting a curve to a plot of the read noise variance as a function of software level gain. The gain can be found if the gain setting is modified in the hardware and a bias frame is measured. The bias frame is subtracted from acquired dark frame and the variance found. The variance is plotted as the software gain is modified. The slope of the corresponding line would be proportional to the unity gain of the camera, an example of this for the Prosilica 1900c camera can be seen in Figure A-4.

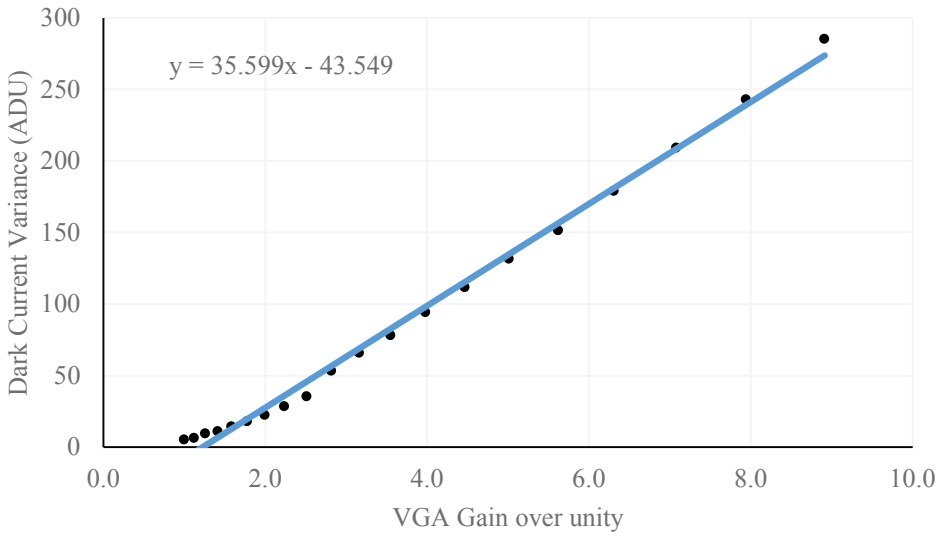


Figure A-4. Determination of hardware unity gain in a Prosiclica 1900c camera

Photon transfer curve

The photon transfer curve can be used to find the sensitivity of the camera if the quantum efficiency is not known. The photon transfer curve is found by looking a diffuse flat field of light and taking several images of the field. The variance of the measurement is then taken along with the mean intensity. The primary source of noise introduced in this method will be shot noise. The shot noise will be proportional to the number of photons arriving at the sensor sight at the given radiance level. The process is repeated for many different radiance levels, and a plot is formed. A linear function is fit to the plot where the slope corresponds to the number of photons per analog to digital unit. A lower number means the camera is more sensitive and a higher number means the camera is less sensitive. An example of a photon transfer curve for the same Prosilica camera can be seen in Figure A-5; there will be one slope per color channel.

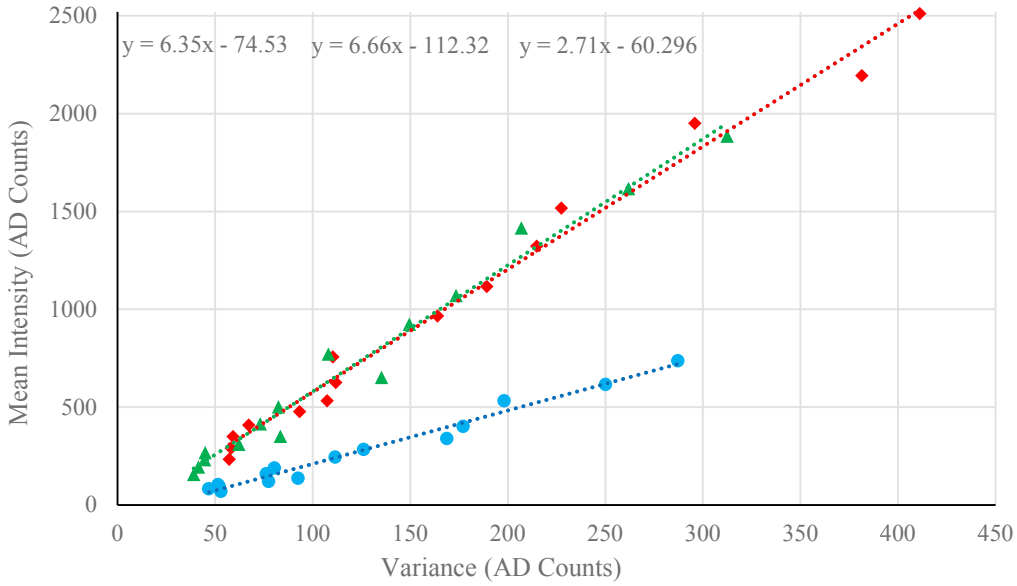


Figure A-5. Photon transfer curve for a Prosilica 1900c camera.

Intersection of two Gaussian curves

The intersection of the two probability density functions at q standard deviations can be found. This procedure can be done to find the intersections of the PDF instead of using a uniform distribution. This is done by equating the two curves

$$\frac{1}{\sigma_l\sqrt{2\pi}}exp\left[-\frac{(\Phi_\cap-\Phi_l-q\sigma_l)^2}{2\sigma_l^2}\right]=\frac{1}{\sigma_m\sqrt{2\pi}}exp\left[-\frac{(\Phi_\cap-\Phi_m-q\sigma_m)^2}{2\sigma_m^2}\right]$$

And solving for quadratic for points of $\Phi_\cap$

$$\Phi_\cap = \frac{2\sigma_l\left(\frac{\Phi_m\sigma_l}{\sigma_m^2}-\frac{q\sigma_l}{\sigma_m}-q-\frac{\Phi_l}{\sigma_l}\right)\pm\sqrt{\left(2\sigma_l\left(\frac{\Phi_m\sigma_l}{\sigma_m^2}-\frac{q\sigma_l}{\sigma_m}-q-\frac{\Phi_l}{\sigma_l}\right)\right)^2-4\left(1-\frac{\sigma_l^2}{\sigma_m^2}\right)\left(2\sigma_l^2\left[\frac{\Phi_m}{\sigma_m}\left[q-\frac{\Phi_m}{2\sigma_m}\right]\frac{\Phi_l q}{\sigma_l}-Ln\left(\frac{\sigma_l}{\sigma_m}\right)\right]+\Phi_l\right)}}{2\left(1-\frac{\sigma_l^2}{\sigma_m^2}\right)}$$

Other Applications for HDR

The image formation methodology introduced in this document is not solely for the use of videos. It can also be used for other forms of instrumentation. To demonstrate this two test cases are provided below. The first test case was produced from scans of X-Ray Film. Four X-

Rays were captured on silver halide film using different levels of energy. The advantage to this system is it allows for the viewing of all of the varying absorption materials on a single screen, seen in Figure A-6. The left two columns are the inverted scanned films that the HDR X-Ray image was taken from. On the right hand side image all of the electrical circuitry is visible, as well as the copper windings of the motor. Upon close inspection, the balls in the ball bearings are even visible with the single view.

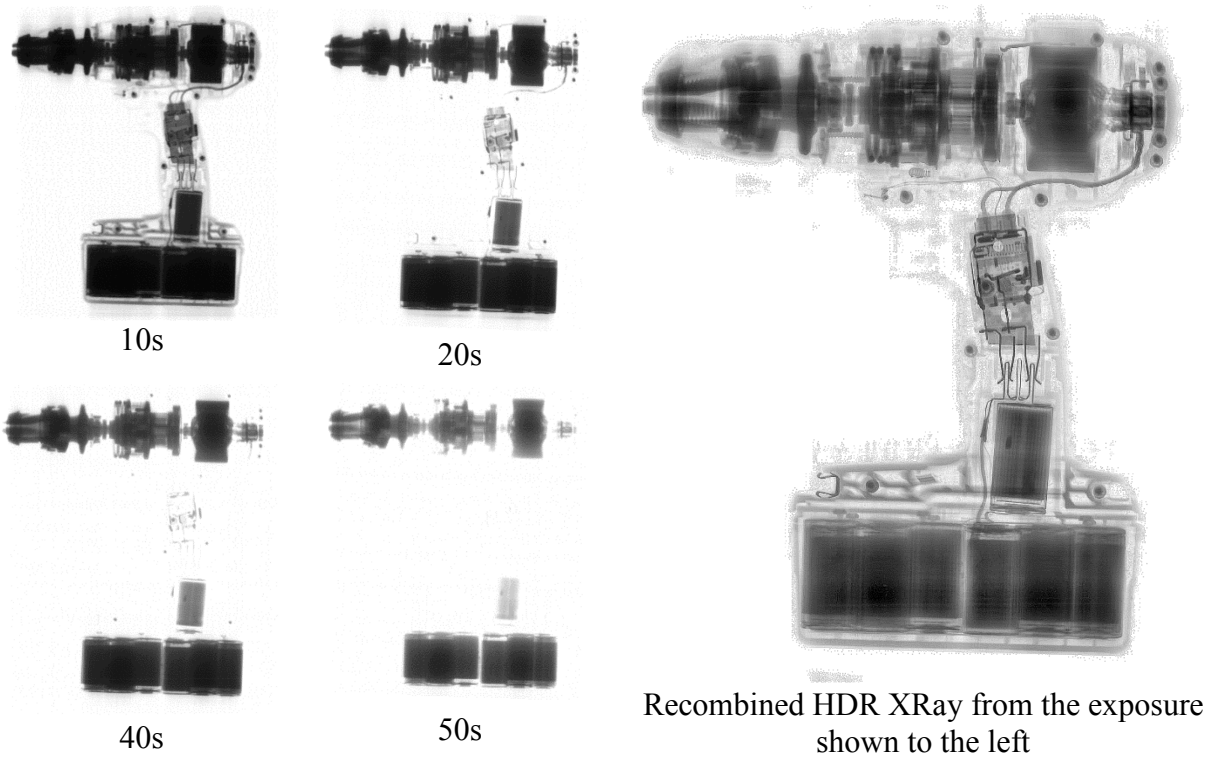


Figure A-6. HDR XRay of a cordless drill.

Another example application stems for the infrared range. Typically thermography cameras employ high bit depth sensors and must find a way to remap the data to the display space. Also oversaturation of the image can occur depending on what the temperature and emissivity of the object that is being viewed. An example of a tone mapped HDR IR image utilizing a FLIR RS 6700 RS camera can be seen in Figure A-7. In this figure the engine compartment of a running pickup truck was viewed. The temperature gradient in the radiator can clearly be scene. Also from the tone mapped data the presence of dirt is exemplified as cool spots on the oil cooler and transmission cooler that are not readily seen in the individual exposures.

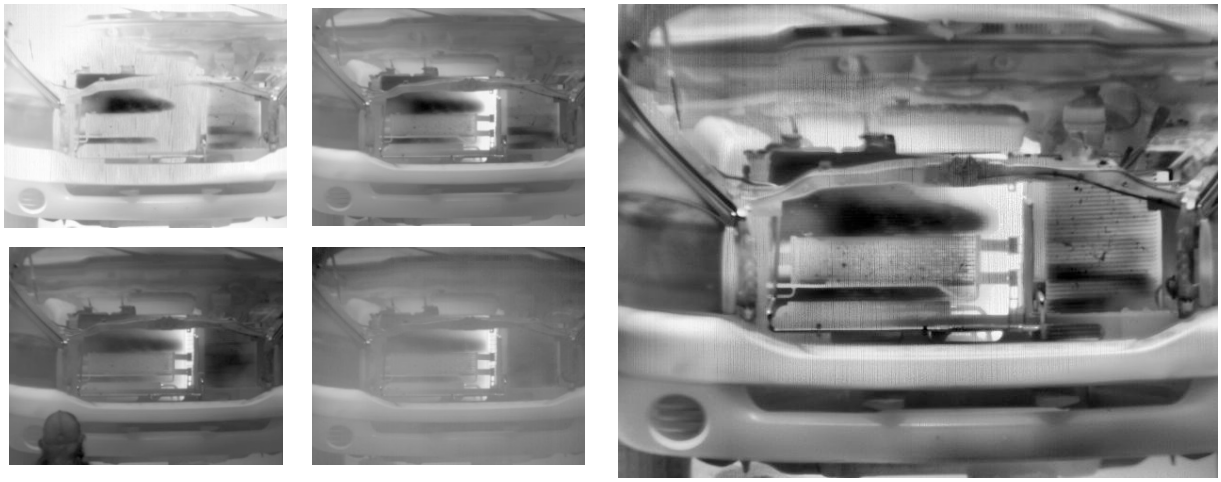


Figure A-7. HDR IR Image generated from a FLIR RS 6700 camera viewing the engine bay of a pickup truck.

Tone mapping settings for HDR Images in Chapter 4

For the sake or reproducibility the following tables indicate the tone mapping settings used for the formation of the images in Chapter 4.

Griffiths Global	Calibration	Correction	Window Size	Saturation	Gamma
Atrium Morning	10	0	11	0.6	3.2
Forest Path	10	0	11	0.6	2.222
Memorial	50	0	11	0.28	2.6
Office	10	0	11	0.4	3
Church	10	0	11	0.6	2.222
Park	10	0	11	0.45	2.222

Griffiths Variant	Calibration	Correction	Saturation	Gamma	FOV	Age
Atrium Morning	10	0	0.6	3.2	2	32
Forest Path	300	0	0.6	2.222	2	32
Memorial	3000	0	0.26	2.6	2	32
Office	500	0	0.5	3	2	32
Church	500	0	0.6	2.222	2	32
Park	4000	0	0.45	2.222	2	32

Fattal	fBeta	Gamma
Atrium Morning	0.89	1.2
Forest Path	0.95	1.2
Memorial	0.9	1.1
Office	0.88	1
Church	0.92	1.1
Park	0.95	1.1

Tumblin Rushmeier	Display Luminance	Max Display Luminance	Maximum display contrast	Calibration	Gamma
Atrium Morning	20	120	100	0.237	1.1
Forest Path	10	500	10000	1	1.6
Memorial	10	500	10000	1	1.7
Office	1	500	10000	1	1.7
Church	10	500	10000	1	1.6
Park	10	500	10000	1	1.7

Pattanaik	Rod Adaptation	Cone Adaptation	Gamma	Gain
Atrimum Morning	80	80	1.1	300
Forest Path	80	80	1	400
Memorial	80	80	1	2000
Office	8000	40000	1	$1*10^5$
Church	80	80	1	200
Park	80	80	1	1000

Reinhard Devlin	Contrast	Intensity	Light Adaptation	Chromic Adaptation	Gamma
Atrium Morning	0.1	-2	1	0	1
Forest Path	0.1	-2	1	0	1.03
Memorial	0.6	-3	1	1	1.2
Office	0.8	-5	1	1	1.08
Church	0.1	-2	1	0	1.05
Park	0.5	-2	1	1	1

Lischinski	Alpha	White	Gamma
Atrium Morning	0.142	47.7	1
Forest Path	0.143	3.6	1
Memorial	0.115	305	1
Office	0.15	44.7	1.01
Church	0.1052	51.37	1.05
Park	0.199	8.73	1.08

Matlab	Tiles	Range	Saturation	Gamma
Atrium Morning	6x6	.001-1	1.8	0.5
Forest Path	6x6	.001-1	1.8	0.8
Memorial	4x4	.001-1	1.2	0.5
Office	4x4	.1-1	3	1
Church	6x6	.001-1	2.5	0.5
Park	4x4	.3-1	3	1

Durand	Contrast	Gamma
Atrium Morning	4.5	1
Forest Path	4.3	1
Memorial	4	0.8
Office	4	1
Church	3.5	1
Park	4.2	1

Potential Camera combinations

Listed in the Tables below are the potential camera combinations for a 2, 3 and 4 imager selection. When looking at a figure such as Figure 6-3, the formation method number corresponds to the camera pool in the table indicated below.

Two camera Pool	
1	$\phi_1 \cap \phi_2$
2	ϕ_1
3	ϕ_2

Three Camera Pool	
1	$\phi_1 \cap \phi_2$
2	$\phi_1 \cap \phi_3$
3	$\phi_2 \cap \phi_3$
4	$\phi_1 \cap \phi_2 \cap \phi_3$
5	ϕ_1
6	ϕ_2
7	ϕ_3

Four Camera Pool	
1	$\phi_1 \cap \phi_2$
2	$\phi_1 \cap \phi_3$
3	$\phi_1 \cap \phi_4$
4	$\phi_2 \cap \phi_3$
5	$\phi_2 \cap \phi_4$
6	$\phi_3 \cap \phi_4$
7	$\phi_1 \cap \phi_2 \cap \phi_3$
8	$\phi_1 \cap \phi_2 \cap \phi_4$
9	$\phi_1 \cap \phi_3 \cap \phi_4$
10	$\phi_2 \cap \phi_3 \cap \phi_4$
11	$\phi_1 \cap \phi_2 \cap \phi_3 \cap \phi_4$
12	ϕ_1
13	ϕ_2
14	ϕ_3
15	ϕ_4