Rochester Institute of Technology

RIT Scholar Works

Theses

5-2020

Bayesian Methods for Radiometric Calibration in Motion Picture Encoding Workflows

Ricardo R. Figueroa
rrfppr@rit.edu

Follow this and additional works at: <https://scholarworks.rit.edu/theses>

Recommended Citation

Figueroa, Ricardo R., "Bayesian Methods for Radiometric Calibration in Motion Picture Encoding Workflows" (2020). Thesis. Rochester Institute of Technology. Accessed from

This Dissertation is brought to you for free and open access by RIT Scholar Works. It has been accepted for inclusion in Theses by an authorized administrator of RIT Scholar Works. For more information, please contact ritscholarworks@rit.edu.

Bayesian Methods for Radiometric Calibration in Motion Picture Encoding Workflows

by

Ricardo R. Figueroa

A dissertation submitted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy
in Computing and Information Sciences

B. Thomas Golisano College of Computing and
Information Sciences

Rochester Institute of Technology
Rochester, New York

May 2020

Bayesian Methods for Radiometric Calibration in Motion Picture Encoding Workflows

by

Ricardo R. Figueroa

Committee Approval:

We, the undersigned committee members, certify that we have advised and/or supervised the candidate on the work described in this dissertation. We further certify that we have reviewed the dissertation manuscript and approve it in partial fulfillment of the requirements of the degree of Doctor of Philosophy in Computing and Information Sciences.

Dr. Pengcheng Shi, Dissertation Advisor	Date
---	------

Dr. Nathan Cahill, Dissertation Committee Member	Date
--	------

Dr. Rui Li, Dissertation Committee Member	Date
---	------

Dr. Linwei Wang, Dissertation Committee Member	Date
--	------

Dr. Shanchieh Jay Yang, Dissertation Committee Member	Date
---	------

Dr. Carl Salvaggio, Dissertation Defense Chair	Date
--	------

Certified By:

Dr. Pengcheng Shi, Ph.D. Program Director	Date
---	------

© 2020 Ricardo R. Figueroa

All rights reserved.

Bayesian Methods for Radiometric Calibration in Motion Picture Encoding Workflows

By Ricardo R. Figueroa

May 2020

Abstract

A method for estimating the Camera Response Function (CRF) of an electronic motion picture camera is presented in this work. The accurate estimation of the CRF allows for proper encoding of camera exposures into motion picture post-production workflows, like the Academy Color Encoding Specification (ACES), this being a necessary step to correctly combine images from different capture sources into one cohesive final production and minimize non-creative manual adjustments.

Although there are well known standard CRFs implemented in typical video camera workflows, motion picture workflows and newer High Dynamic Range (HDR) imaging workflows have introduced new standard CRFs as well as custom and proprietary CRFs that need to be known for proper post-production encoding of the camera footage. Current methods to estimate this function rely on the use of measurement charts, using multiple static images taken under different exposures or lighting conditions, or assume a simplistic model of the function's shape. All these methods become problematic and tough to fit into motion picture production and post-production workflows where the use of test charts and varying camera or scene setups becomes impractical and

where a method based solely on camera footage, comprised of a single image or a series of images, would be advantageous.

This work presents a methodology initially based on the work of Lin, Gu, Yamazaki and Shum that takes into account edge color mixtures in an image or image sequence, that are affected by the non-linearity introduced by a CRF. In addition, a novel feature based on image noise is introduced to overcome some of the limitations of edge color mixtures. These features provide information that is included in the likelihood probability distribution in a Bayesian framework to estimate the CRF as the expected value of a posterior probability distribution, which is itself approximated by a Markov Chain Monte Carlo (MCMC) sampling algorithm. This allows for a more complete description of the CRF over methods like Maximum Likelihood (ML) and Maximum A Posteriori (MAP). The CRF function is modeled by Principal Component Analysis (PCA) of the Database of Response Functions (DoRF) compiled by Grossberg and Nayar, and the prior probability distribution is modeled by a Gaussian Mixture Model (GMM) of the PCA coefficients for the responses in the DoRF. CRF estimation results are presented for an ARRI electronic motion picture camera, showing the improved estimation accuracy and practicality of this method over previous methods for motion picture post-production workflows.

Acknowledgements

Thank you to Dr. Pengcheng Shi for all his support, guidance and patience during these past years. Your knowledge, experience and wit helped me balance all aspects of my life to be able to complete this work. In addition, thank you to my dissertation committee for all their guidance and feedback, especially to Dr. Rui Li, for all the technical advice and the insightful conversations that led to finalizing the work presented in this dissertation. Last but not least, thank you to Dr. Jinwei Gu for providing the original inspiration and ideas that helped start up this work.

Dedicated to my family, Anabelle Lee, Ricardo José and Paola Beatriz.

I love you all very much.

Contents

1	Introduction	1
1.1	The Problem	1
1.2	Methodology Overview	6
1.3	Contributions	10
1.4	Dissertation Organization	13
2	Motion Picture Imaging Background	14
2.1	Introduction	14
2.2	Imaging Fundamentals Review	15
2.2.1	Scene Radiance to Digital Image Code Values	15
2.2.2	White Balancing	19
2.2.3	Gamma Correction	19
2.3	Motion Picture Workflows	23
2.3.1	Traditional Film Workflows	23
2.3.2	Modern Workflows	25
2.3.3	CINEON	27
2.3.4	ACES	29
2.3.5	Image Interchange Challenges	31
3	Radiometric Calibration Background	33
3.1	Introduction	33
3.2	Camera Response Function Estimation	34
3.2.1	Camera Response Function Model	37
3.3	General Probability Review	39

3.4	Radiometric Calibration Through Edge Color Distributions . .	43
3.4.1	Edge Color Distributions	43
3.4.2	Edge Color Boundary Detection	44
3.4.3	Edge Colors Analysis	44
3.4.4	Transformation of Edge Color Non-linearities to Linear Distributions	46
3.4.5	Probabilistic Framework	49
3.4.6	Prior Probability Distribution	51
3.4.7	Gaussian Mixture Model	51
3.4.8	Likelihood Probability Distribution	52
3.4.9	Bayesian MAP Solution	52
3.4.10	Ensuring Monotonicity	53
3.4.11	Camera Response Function Estimation Ambiguity . .	53
3.5	Summary	55
4	Novel Radiometric Calibration	56
4.1	Introduction	56
4.2	Radiometric Calibration with Multiple Frames	57
4.3	Noise Image Features for Radiometric Calibration	58
4.3.1	Noise in Electronic Images	58
4.3.2	Linearity of Noise Features	62
4.3.3	Pearson Correlation Coefficient and Pearson's Distance	67
4.3.4	Noise Patch Detection	69
4.3.5	Bayesian Framework with Posterior Probability Distri- bution Estimation	70

4.3.6	Prior Probability Distribution	73
4.3.7	Likelihood Probability Distribution	73
4.3.8	Posterior Probability Distribution	73
4.3.9	Markov Chain Monte Carlo Sampling	75
4.3.10	Metropolis-Hastings Algorithm	77
4.4	Summary	80
5	ACES Input Device Transform	82
5.1	Introduction	82
5.2	IDT for Image Capture	84
5.2.1	ACES IDT Cost Function	84
5.3	Summary	90
6	Alternative Approach for CRF Estimation	91
6.1	Introduction	91
6.2	Neural Networks	92
6.2.1	Bayesian Methods in Neural Networks	93
6.2.2	Model Selection	95
6.3	Summary	97
7	CRF Estimation Experimental Results	98
7.1	Introduction	98
7.2	Camera, Images and Setup	100
7.3	Multiple Frames MAP CRF Estimation	101
7.4	CRF Estimation Using Edge Color Distributions	104
7.5	CRF Estimation Using Noise Image Features	107

7.6	IDT Computation Color Error Analysis	108
7.7	Methodology Limitations Summary	112
7.8	Hardware Platform Details	115
8	Summary	116
8.1	Conclusions	116
8.2	Future Work	121
A	APPENDIX	126
B	APPENDIX	130
C	APPENDIX	132

List of Figures

2.1	Vignetting Effect	16
2.2	Image Formation Diagram	18
2.3	CRT Transfer Function	20
2.4	Rec. 709 and sRGB CRFs and CIE L*	22
2.5	Traditional Film Workflow	23
2.6	Film System Jones Diagram	25
2.7	Film Digitization Workflows	27
2.8	CINEON Exposure Prediction Log Error	29
2.9	ACES Workflow	30
2.10	ACES IDT Implementation Steps	31
3.1	Probability Diagram	39
3.2	Canny Edge Detection Map	45
3.3	Measured Color Non-linear Distribution at Image Edges (Adapted from Lin et al.)	46
3.4	Non-linear to Linear Distribution Transformation (Adapted from Lin et al.)	48
3.5	Database of Response Functions	50
4.1	Total Noise vs. Raw Code Value	66
4.2	Effect of CRF on Noise vs. Raw Code Value Relationship . . .	66
4.3	Sample Pearson Correlation Coefficient Scatter Diagram Ex- amples	68
4.4	Edge Map and Noise Patches	71
4.5	Noise Features Plot for Non-Linear Image	72

4.6	Noise Features Plot for Linearized Image	72
4.7	MCMC Trace Plots for Mean and Standard Deviation for 30,000 ESS	77
5.1	MacBeth Color Chart	83
6.1	Neural Network Node	93
6.2	Log-Sigmoid, Tan-Sigmoid and Linear Activation Functions .	94
7.1	Arri D-21 Gamma Encoding Response Function	100
7.2	Arri D-21 LogC Encoding Response Function	101
7.3	Sample ARRI D21 Frame	102
7.4	Camera Response Function Estimation Plot for 10 and 20 Frames	103
7.5	MCMC CRF Estimation with Edge Color Distributions	105
7.6	MCMC Approximation of Target Function for Coefficient 1 through 5	106
7.7	MCMC CRF Estimation with Noise Image Features	108
7.8	MCMC Approximation of Target Function - 3 Coefficients . .	109
A.1	ARRI ISO Dependent CRFs	127
A.2	Sony CRFs	128
A.3	Canon CRFs	129
B.1	CRF Post-Production Workflow Diagram	131
C.1	ACES Workflow Diagram	133

List of Tables

2.1	End-to-End Power Functions	21
7.1	RMS Error	103
7.2	IDT deltaE Calculation Error	110

1 Introduction

1.1 The Problem

Digital processes generate large amounts of data. These processes exist in a large number of applications and in most cases, the generated data exists by itself without any information or details about the process that created it. It is of interest to use computers to analyze the data, learn information about the process, and be able to predict future data from an estimation of the process that generated that data. This is also known as the field of machine learning.

One of such processes is the capture of images by a digital system. For more than a 100 years, audiences have enjoyed the movie going experience. Whether it is a comedy, mystery, drama or action movie, audiences enjoy being immersed in a story that will bring all sort of emotions while experiencing the images and sounds associated with that story. Moviemakers have always focused on delivering the highest quality images possible and work tirelessly behind the scenes, during production, post-production and exhibition, to make sure that audiences fully experience the story and don't get distracted by any of the many technical details behind the movie making process. Since the beginnings of this industry, and even more today, movies have been created by combining images from different capture mediums. From high resolution 65mm silver halide film, to small form factor consumer digital cameras, all have been used in the movie making process depending on the camera characteristics

needed for the situation. Some of these characteristics include ease of use, form factor, image quality, and cost to name a few. While in many cases selecting a different camera based on the needs of a specific scene ensures capturing the best possible images for that situation, all footage captured has to be integrated to make a movie that looks consistent and uniform, this being one of the most important post-production processes that audiences are rarely exposed to. Different cameras have different ways of encoding the light captured from a scene. This encoding converts the light falling onto the imaging sensor into digital pixel code values that are then manipulated in post-production.

One of the most important steps in this encoding is the implementation of a Camera Response Function (CRF) which allows the camera to have a more efficient encoding and to encode the light from the scene in a way that mimics the human visual system. Another important step following the implementation of the CRF is a color encoding transform that ensures subsequent appropriate display color reproduction. Although the main objective of the CRF is the same for all cameras, different cameras implement different types of functions due to different existing standards, advancements in camera technology, proprietary advantages, and image sensor characteristics to name a few reasons. This means that the CRF is one of the reasons why images will look different when brought into the post-production workflow and why time has to be spent in adjustments so they all look uniform for a given scene or movie. These adjustments, also known as technical grading, are a time consuming endeavor that could be optimized by implementing methods to estimate the camera specific characteristics that created the images, like the

CRF, without the need for specific camera analysis procedures that are not necessarily viable or practical in post-production environments. In summary, the post-production workflow can be optimized by learning about the camera encoding process from digital image data. This will result in a reduction of post-production time spent in technical grading, allowing for the time to be spent in creative grading operations.

Combining images from different sources into the same workflow is again, a very common practice in the motion picture industry and development of the Academy Color Encoding System (ACES) [1] began in 2004 as an effort by the Science and Technology Council at the Academy of Motion Picture Arts and Sciences (AMPAS) to improve these workflows. The goal of ACES was to establish standards for the interchange of digitally mastered motion picture images. One example of earlier attempts to accomplish this was the CINEON platform established by Eastman Kodak in the early 90s, which was widely adopted and used in the industry, especially for bringing film originated images into digital workflows. Even though CINEON provided a way of combining film originated images with digital originated images, many technical details were left up to the end-user along with the adoption of simple approximations of film characteristics. This led to different interpretations and implementations resulting in increased post-production inefficiency, increased costs, constant manual adjustments and reduced image quality. Typical television or motion picture productions involve different image manipulation workflows that are very challenging to integrate. Many of these productions can use various digital cameras and also include film in their production. Add to this the fact that during post-production, images from all these sources will arrive

at a facility in a number of different formats, color encodings and in many cases without helpful metadata. A critical part of ACES is the Input Device Transform (IDT) which aims at characterizing every possible input so it can be integrated into the ACES workflow. The CRF is a key camera characteristic that needs to be known for the correct calculation and/or implementation of the IDT since it relates camera code values to linear scene radiometry, but factors like camera model, camera settings and shooting conditions can play a role in the variability of a CRF. It would then be advantageous to estimate each input CRF only using captured images which are already part of the post-production workflow or by techniques with very limited interruption to post-production.

This work focused on developing a method to estimate a motion picture CRF from an image or a sequence of images to ensure an accurate motion picture color encoding transformation. As stated above, professional motion picture production requires images from different sources to be seamlessly combined. This in turn requires many camera characteristics like the CRF, sensor spectral sensitivity, color encoding transform, image processing chain and sensor characteristics among others, to be understood so this task can be accomplished accurately and efficiently. Although these camera characteristics are easily measurable in a laboratory environment, they are hard to measure during production or post-production where time is short and other processes have priority. Furthermore, modern cameras offer the user a wide range of options to modify some of these characteristics making the interchange of images even more challenging during post-production.

Understanding the practical motivation for this research, it's evident

that to solve that problem, an computational inference problem must be solved to estimate the CRF implemented by an electronic camera during capture without having any information about the capture process, with the exception of the images themselves. Here is where a probabilistic framework brings benefits to finding a solution that depends exclusively on the data in question, the images. This framework is summarized in the next section and further expanded in this dissertation.

1.2 Methodology Overview

The computational problem in this research involved the statistical inference of a CRF based on captured image data. The data is originated in an electronic camera system, providing a large spatial and temporal sampling of the scene being captured. A Bayesian probabilistic framework presents a viable and preferred solution to this problem due to the fact that it falls under the category of evidential probabilities. This means that the problem can be formulated in terms of prior beliefs about the process that created the data, and the calculation of new beliefs as a result of additional evidence, like new image data, being taken into account and producing a posterior probability distribution.

By understanding the motion picture imaging process, important facts about the history of the industry, standard and non-standard methods to capture images, and post-production methods to combine images, it is possible to justify the decision to attack this problem with a Bayesian probabilistic approach. Historical and standard capture methods present a solid foundation for prior beliefs about the mathematical description of the CRF. But this can only represent a probabilistic belief, as new capture methods and non-standard methods exist and bring uncertainty to the process. Representing this uncertainty as a prior probability, allows for a statistical inference starting point, that can then be modified based on relevant evidence, modeled by a likelihood probability distribution over the captured image data, to then produce a posterior probability distribution as the result.

With the solution methodology proposed as a Bayesian probabilistic framework, the prior probability distribution and the likelihood probability distribution were modeled. The prior distribution was based on previous work

by Grossberg and Nayar [2], a principal component analysis of a database of existing measured response functions, and a subsequent Gaussian mixture model associated with the coefficients obtained after the principal component decomposition for all CRFs in the database. The likelihood was modeled as an exponential distribution with a specific image feature based metric being used as the main parameter in the likelihood probability distribution. Two different image features were considered in this research. The first one is based on the work of Lin et al. [3] and takes into account edge color distributions that are affected by the non-linearity imparted by a CRF on image data. The second image feature is a novel contribution and is based on image noise signatures and the relationship between the combination of read and photon noise, and the measured image code value. This relationship is also affected by the non-linearity imparted by a CRF and is used to establish an additional distance metric that is then incorporated in the likelihood probability distribution. Both of these image features were incorporated in the Bayesian framework. First, an extension of Lin's work included additional edge color distributions from multiple image frames, increasing the amount of edges considered in the approach and minimizing the potential of noise influencing one specific image. This approach focused on a Maximum A Posteriori (MAP) solution to the Bayesian posterior probability estimation problem. The second method involved noise image features as the distance metric for the likelihood function, and the estimation of a posterior probability distribution utilizing a Markov chain Monte Carlo (MCMC) sampling approach.

By expanding Lin's approach to a Bayesian framework that takes advantage of the information contained in the posterior probability distribution

and not only in a single point estimate like in Maximum Likelihood (ML) or MAP, the posterior probability distribution can be considered for estimating an expected value for the CRF. Probabilistic solutions like ML and MAP are viable solutions to the CRF estimation problem, but don't take full advantage of all the process variables' information contained in a posterior probability distribution. By obtaining an estimate of the posterior probability distribution for the parameters in question, the calculation of an expected value for the posterior distribution improves results especially when dealing with non-Gaussian shaped distributions like a bi-modal distribution, where the mode of the distribution is not an accurate representation of the probability distribution. Computing the posterior probability distribution though, comes with the challenge of computational complexity, which is mitigated by the implementation of MCMC sampling.

In summary, the methodology chosen brings a solution that only depends on image pixel code values and takes advantage of inherent image features that require no additional measurements or scene setups. The methodology is not only optimized to produce a minimal mathematical error calculation but also minimize a perceptual color error metric, to ensure industry practicality. The estimation of an expected value of the posterior probability distribution brings benefits when dealing with complex parameter probability distributions and improves the error metrics and estimation results over methods based on a single point, like the mode of the distribution. Lastly, the methodology brings a novel use of image noise as an image feature signaling radiometric linearity of the image data. This is especially advantageous for the problem presented in this dissertation, but also applicable to other imaging

applications where radiometrically referenced image code values can be useful (e.g. remote sensing).

1.3 Contributions

The Bayesian probabilistic framework summarized in the previous section, presents various advantages over existing radiometric calibration methods. These advantages are related to improvements to the state of the art probabilistic methods used in radiometric calibration, new relevant image feature identification to include in the Bayesian framework and practical benefits that the methodology brings into post-production workflows. These improvements are summarized next.

1. First, improvement of the method utilizing edge color distributions as image features, by expanding the number of total color edges identified. This expansion was carried out by using not only a single image frame, but a sequence of frames and restructuring the existing algorithm to incorporate image sequences instead of single frames. The increase in the number of edge color distributions provided additional data for the methodology originally presented by Lin et al. The additional data compensated for errors in the estimation of the CRF due to noise or excessive image processing influencing the edge color distributions considered.
2. Second, the extension of Lin's methodology to compute an estimate of the posterior probability distribution function. Lin's method implements a MAP solution to the Bayesian framework to solve the statistical inference problem presented in the previous section. MAP provides a single point solution and does not take into account all the information included in the posterior probability distribution. This research implements a

sampling algorithm to estimate the posterior probability distribution and subsequently compute an expected value for the CRF. Estimation improvements are shown by estimating the posterior distribution and computing an expected value versus a MAP solution.

3. Third, identification of a novel image feature to be incorporated into likelihood function of Bayesian framework. Although image noise is thought of as an undesirable influence in edge color distributions, analysis of the relationship between the total noise in non-textured parts of an image with different exposures, versus the corresponding measured image code value, showed that a CRF has an impact in the linearity of that relationship and an appropriate metric could be developed to utilize this information in the probabilistic framework. In addition to identifying this image feature, a metric based on the Pearson correlation coefficient was identified and incorporated into the likelihood probability distribution. The linearity of image data is not only important to the problem presented, but also in other industries like remote sensing. This novel image feature has the potential to be very useful in any applications where radiometrically referenced image data is of use for understanding of light/matter interaction properties of objects that have been imaged. Remote sensing and photogrammetry are examples of applications that can benefit from this novel feature and approach.
4. Fourth, the development and implementation of this solution methodology was done emphasizing the need to only depend on capture image data obtained during normal production in a motion picture workflow.

This is an important improvement over state of the art methods that require specific images for analysis or constrained or limited capture conditions and lighting. By only depending on normal workflow images from typical productions, the solution in this research doesn't interfere with well-established workflows in the motion picture industry, and presents the opportunity to incorporate the algorithm into post-production encoding workflows seamlessly.

5. Last, the solution presented in this research can improve the efficiency of post-production workflows focused on the combination and interchange of images from different capture sources. The ability to estimate a CRF accurately, ensures the proper linearization of the images for subsequent encoding in ACES, or any other workflow or application that requires radiometrically linear image data. This results in color error reduction during the implementation of an IDT or during the computation of an IDT for the ACES workflow, reducing unnecessary time spent in manual adjustments during technical grading, and leaving more time for creative grading of images. This inclusion of a perceptual metric in the optimization of the approach presented, serves as an example where a perceptual metric can complement a mathematical error metric during optimization to ensure a practical result for a specific application. Industries like photography and printing can also benefit from this type of approach.

1.4 Dissertation Organization

This dissertation is organized as follows. Chapters 2 and 3 cover the motion picture imaging background along with background associated with radiometric calibration methods and the details of the methodology that served as a starting point for this research. Chapter 4 covers the novel radiometric methodology that expands on chapter 3. Chapter 5 describes the details of the ACES IDT calculation and implementation, as well as the influence of the CRF in the IDT performance. Chapter 6 covers a possible alternative approach, its benefits and why the current methodology was chosen at this time. Chapter 7 describes the experiments carried out and the associated results along with limitations to the methodology. And lastly, chapter 8 covers a discussion of the results and contributions, and presents possible future work.

To keep compatibility with the terminology in the literature, the terms Camera Response Function (CRF), Radiometric Response Function (RRF) and Opto-Electronic Transfer Function (OETF), a specific term for digital capture, should all be interpreted as referring to the same function that relates scene radiometry to measured code values. The term gamma also refers to the non-linear reproduction in cameras but also includes aspects from perception and video in addition to physics and it is further explained in the next section.

2 Motion Picture Imaging Background

2.1 Introduction

The following sections introduce image formation fundamentals along with how the manipulation of images, in the television and motion picture industry, requires the understanding of the CRF for proper interchange of images. Traditional and current workflows in the motion picture industry are described, along with old and modern methods that have been developed to overcome the many challenges that exist in the interchange and combination of images from different sources during post-production. The description of the image formation process establishes the linear relationship between radiometric energy from a scene and the measured image code value and also establishes the reasoning for the implementation of a CRF after the capture of radiometric energy from a scene to properly represent brightness changes according to human perception and to ensure efficient use of the available encoding bit depth in digital systems.

2.2 Imaging Fundamentals Review

How is a digital image formed and how are digital code values related to scene brightness? This section reviews how images are formed in a camera and how the values obtained are related to the light coming from the scene. This process is divided into two parts, how the scene radiance is related to the image irradiance falling on an imaging sensor and how the image irradiance is related to pixel values.

2.2.1 Scene Radiance to Digital Image Code Values

The amount of light coming from a scene is defined by the radiance L , which is the radiant power per unit of foreshortened area per unit solid angle, with units of watts per square meter per steradian (W/m^2sr). After the radiance passes through a lensed system, the energy falling onto an image sensor is known as image irradiance E , with units of watts per square meter (W/m^2). The irradiance falling onto the image sensor is proportional the scene radiance but will vary from the center towards the outside of the sensor due to various factors. The irradiance will decrease as $\cos^4$ of the angle θ that an incident ray makes with the optical axis. Equation 2.1, presents this relationship between scene radiance and image sensor irradiance, sometimes also known as the camera equation. In the equation $f\#$ is the f-number, which is the ratio of the lens focal length to the diameter of the aperture.

$$E = \frac{L\pi \cos^4 \theta}{4(f\#)^2} \quad (2.1)$$

Figure 2.1: Vignetting Effect
Image showing the $\cos^4$ light fall-off associated with vignetting.

Another source of irradiance fall-off is the phenomenon known as vignetting. This refers to the darkening of the edges of an image due to blocking of the incident rays at oblique angles, by the effective aperture size. Figure 2.1 shows an example of this effect on an uncorrected uniform exposure.

Although many professional cameras employ advanced algorithms for automatic correction of fall-off effects, close attention has to be paid to images captured and confirm automatic corrections to avoid CRF estimation errors due to any of the previously mentioned factors.

As defined in equation 2.1, the amount of light collected at an image sensor is the irradiance E . Irradiance is then transformed to image code values through the CRF. Equation 2.2 describes this relationship where E is irradiance and M is the image intensity (i.e. measured pixel code values).

$$M = f(E) \tag{2.2}$$

Increasing the irradiance onto an image sensor will increase, or keep constant, the image intensity or code values, making this function monotonic and allowing it to be invertible. There are various reasons for implementing

a CRF, including extending the capture dynamic range of the image sensor, improving the efficiency of code value distributions throughout the dynamic range, processing the linear radiometric captured signal for display and/or to mimic the non-linear human visual system response to brightness changes as is explained in section 2.2.3.

In most cases the dynamic range of a scene exceeds the capture dynamic range of a camera and the exposure h plays a role in the final image intensity that will be recorded by controlling the amount of light exposed onto the image sensor. Exposure is defined as the product of irradiance and time at the image sensor as shown in equation 2.3. Aperture and shutter speed are the two principal exposure controls available to the camera operator and due to reciprocity, different combinations of these controls will produce the same sensor exposure. Decisions regarding depth of field and/or motion blur will dictate which combination is the most appropriate for a specific situation.

$$h = Et \tag{2.3}$$

Another camera setting that impacts the final image intensity is the International Standards Organization (ISO) speed. ISO speed is a numerical value inversely proportional to the exposure needed to produce a specified output by the sensor and is used by camera manufacturers to rate the light capture efficiency of their imaging systems. Two popular measurement techniques that exist are saturation based ISO and noise based ISO [4]. The saturation based ISO refers to the maximum possible exposure before clipping occurs while allowing for highlights up to 141% (specular highlights). The noise based ISO

Figure 2.2: Image Formation Diagram

General diagram describing light sensing and encoding into measured pixel values.

refers to the exposure that will produce a given signal to noise ratio image, typically 10:1 and 40:1 for what the standard calls an *acceptable* image and an *excellent* image respectively. The effects of different ISO settings in the estimation of the CRF are studied in this research and methods to make the estimation independent of ISO are presented.

Figure 2.2 shows the overall process from radiance to image intensity pixel code values.

2.2.2 White Balancing

Another setting that affects the image intensity is white balance. Digital cameras implement white balancing by adjusting a gain in each color channel until a neutral scene object outputs equal pixel code values or intensities. These gains in turn affect the intensity of the overall image. Although camera manufacturers implement white balancing as simple channel gain corrections, the IDT computation depends on an error metric that measures perceptual color differences and a more elaborate treatment of white balancing is used incorporating human vision chromatic adaptation models. This is further explained in section 3.3.2.

2.2.3 Gamma Correction

In the computer, television and motion picture industries, the gamma correction term is a numerical parameter that represents the non-linearity of luminance reproduction by a display device. It traditionally has been somewhat mysterious and hard to understand because it not only involves the physics of the display when reproducing an image, but also the perception of those images under different surround conditions. From a strictly physics point of view, gamma correction in video and computer graphics systems comes from the process of correcting for the inherent non-linearity of old cathode ray tube (CRT) display devices, where the luminance produced on the screen was a non-linear function of the input voltage as shown in figure 2.3 with normalized axes.

Figure 2.3: CRT Transfer Function

Typical cathode ray tube transfer function relating encoded input voltage signals to luminance from the screen

In addition to this, the perceived luminance sensation of human vision known as lightness, CIE L^* (CIE is French acronym for International Commission on Illumination), is roughly a 0.3-0.4 power function of luminance, where luminance is defined as radiance weighted by the spectral sensitivity function that is characteristic of vision, also known as the luminous efficiency function [5]. As it is mentioned later in section 2.3.1 when discussing how silver halide records light exposure, lightness can also be approximated by a logarithm function based on Weber's Law for Just-Noticeable-Differences and is still an active area of study [6].

Regardless of the non-linear model used for representing the lightness sensitivity in human vision, this is a necessary adjustment when encoding images to avoid imaging artifacts and make effective use of available encoding bits. This imaging artifact is called contouring or banding, and it occurs when the luminance ratio between adjacent code values is larger than 1%, known also as the Weber contrast, and is easiest to detect in areas of smoothly varying shades [5]. By combining these two concepts, an interesting coincidence is revealed where the physics related non-linearity of a CRT display is very similar

IMAGING SYSTEM	Encoding Exponent	Decoding Exponent	End-to-End Exponent
CINEMA (dark surround)	0.6	2.5	1.5
TELEVISION (dim surround)	0.5	2.5	1.25
OFFICE (light surround)	0.45	2.5	1.125

Table 2.1: End-to-End Power Functions

End-to-end contrast adjustments for different viewing environments.

to the inverse of the lightness sensitivity of human vision. This justifies the use of gamma correction regardless of the physics involved in the display. Even more today, when CRT displays have been replaced by newer technologies with linear characteristics, gamma correction or non-linear encoding is still necessary for perceptual reasons.

One additional concept to introduce is that of rendering intent. Rendering intent refers to the alteration of a reproduced tone scale to take into account perceptual effects when viewing images reproduced with substantially lower luminance than that of the original scene, reproduced with a limited contrast ratio, or viewed in a dim lit surround. This correction is achieved by subjecting the luminance to an end-to-end power function between approximately 1.1 and 1.6 and it will depend on the ratio of the scene luminance to the reproduced luminance, the display physics and the viewing environment. The data in table 2.1, acquired from Poynton [5], presents the approximate exponents for a power function to achieve a perceptually correct encoding depending on the surround condition. This means that the product of the encoding exponent and the decoding exponent set the end-to-end power function that imposes the rendering intent.

In summary, non-linear encoding of captured scene radiance is necessary for various reasons, including display physics, visual perception, repro-

Figure 2.4: Rec. 709 and sRGB CRFs and CIE L*
Standard CRFs for encoding video to be delivered to video displays and
computer displays compared with the human luminance perception
(lightness) function.

duction brightness and surround conditions. In many cases depending on the specific application or industry, the non-linear encoding will be different. As an example, different functions have been standardized for video systems and are in use today. These include the ITU Rec. 709 international standard specifying the parameters for HDTV, SMPTE 240M specifying the first studio HDTV standard, Rec. 1361 extending the ITU Rec. 709 coding to accommodate wider color gamuts and sRGB specifying the parameters for computer based displays to mention a few. Figure 2.4 shows a comparison of the ITU Rec. 709 and the sRGB CRFs along with the CIE L*.

In addition to these standardized CRFs, many other non-standard CRFs exist in camera systems that have been developed to extend the capabilities of traditional video systems adding complexity when combining images from different sources.

Figure 2.5: Traditional Film Workflow
Analog motion picture image workflow for silver halide media (film).

2.3 Motion Picture Workflows

In today's motion picture industry, images are created in a number of different ways including silver halide film, digital cameras and computer generated imagery (CGI). Furthermore, the industry is largely entangled with the computer and video/broadcasting industries where different independent standards have been created for the capture and display of images. This section presents the background from traditional to modern motion picture workflows and some of the systems that have been implemented to attempt the seamless interchange of images from different sources.

2.3.1 Traditional Film Workflows

Since around 1900, the motion picture industry has been entertaining audiences all around the world. For the next 100+ years, technical developments along with extensive human psychophysical tests were combined to produce the images that we've come to enjoy on the big screen. Figure 2.3.1 shows the basic steps of this traditional workflow where silver halide negative film is

optically printed onto a silver halide print film for exhibition.

These developments focused in perfecting a system where silver halide film was the basis for capturing, duplicating, storing and displaying images and many of the image quality standards from this system are still the aim for modern motion picture systems. Figure 2.6 shows what is known as a Jones diagram, or a four quadrant tonescale diagram. The two graphs on the top share the horizontal axes with the two bottom graphs, and the two graphs on the left share the vertical axes with the two graphs on the right. This diagram shows how the response function of a capture negative film is designed to work perfectly with the response function of a negative print film so that images on the screen are perceptually correct and produce the right contrast on the movie screen. This system is designed to produce the correct rendering intent. This is analogous to gamma correction in video cameras to prepare the images to be displayed on video monitors. But in this film scenario, gamma takes a slightly different meaning from the gamma in video. In film systems, gamma is the slope of the straight-line portion of the film's response function, also known as the characteristic curve, for a specific film stock and it relates how changes in log exposure produce changes in density formed on the film. This means that the axes are logarithmic. This is also an interesting coincidence. Silver halide emulsions react to light in a logarithmic manner in the same way as one of the models for human lightness sensitivity [6] and plotting characteristic curves in log-log axes makes intuitive sense.

Even though these workflows produced good visual results and had been perfected over time, some problems still persisted. These included slow and expensive post-production, limited editing capabilities and photochemical

Figure 2.6: Film System Jones Diagram

Graphical diagram of the tone transfer from scene to projected print film for a film workflow.

processing variability as examples. Trying to eliminate some of these problems, the industry started to adopt digital workflows to benefit from immediate feedback and ease of editing. One of the main challenges would be how to find a standard way of digitizing silver halide images to guarantee correct interchange with digitally originated media that would also be part of motion picture workflows.

2.3.2 Modern Workflows

Modern motion picture workflows lie at the intersection of traditional silver halide workflows, video/broadcast workflows and CGI workflows. Video workflows have evolved over the last 75+ years and are easily understood and well implemented on properly calibrated professional equipment. This makes it easy for video industry professionals to take advantage of tone and color ma-

nipulation tools that follow the general rules and standards established. So why not adopt the same standards for digital workflows in the motion picture industry? Actually, many of these have been adopted and implemented when the media being manipulated originated with video systems. The problems arise when film originated images are digitized and brought into a digital workflow. Video standards work well but are restricted in color gamut and dynamic range, making them not practical for future workflows where new technologies bring advancements like wide color gamut and high dynamic range to capture and display systems [7, 8].

Two related but distinct general workflows have evolved regarding digitization of film. One focused on digitizing film images for video/broadcast distribution and the other focused on digitizing film images for motion picture distribution, also known as the digital intermediate (DI) process. Both workflows require knowledge of the film response function when converting film density to code values and in both cases approximations of the characteristic curve have been implemented in the industry. These approximations lead to errors that in some cases can be resolved through color correction but in other cases cause problems in the interchange of images [9, 10]. Figure 2.7 shows the general steps in these workflows.

One last part of modern workflows is CGI. These have become ubiquitous in the industry not only to add special effects and characters but also to add image content that might have been hard or unavailable to capture initially. The interesting part is that CGI content is always created in reference to an artist reference display and more importantly, it employs linear light encoding unlike video or film, making it straightforward to composite

Figure 2.7: Film Digitization Workflows

Hybrid workflow showing the digitization of film for post-production and optional recording back to film for analog projection.

with linearized photographic images, but typically problematic with images processed through a CRF.

2.3.3 CINEON

The CINEON Digital Film System was introduced by Kodak in the early 90's and was quickly adopted by the industry for the digitization of film and subsequent interchange of film and digital images [11]. The system was built with the goal of encoding film printing density. Printing density is described as the log exposure onto the negative print film during the printing stage. Equation 2.4 describes how to formally calculate printing density, where PD_i is the corresponding color channel printing density, T is the transmittance of the negative film being exposed onto the print film, SS_i is the spectral sensitivity of the negative print film for that color channel, S is the spectral power distribution of the printer light source and λ is the wavelength.

$$PD_i = -\log \frac{\int_{\lambda} T(\lambda) SS_i(\lambda) S(\lambda) d\lambda}{\int_{\lambda} SS_i(\lambda) S(\lambda) d\lambda} \quad (2.4)$$

Evidently, printing density is a logarithmic quantity and the CINEON system encodes it digitally with 10 bit quantization where each code value represents 0.002 PD, based on an average 2.046 film density range. The CINEON system assumes an average characteristic curve from negative films from the mid 1990s. This average characteristic curve is described as a linear function in log-log space with a slope of 0.6 without any toe or shoulder roll off as it is shown in figure 2.6. This approximation can cause errors when predicting exposure from digital code values (linearization) versus using the actual characteristic curve for the film for the prediction. Calculating scene exposure is the main step in the encoding of images for the interchange of different sources, and the reason for understanding the CRF or film transfer function. This error is shown in figure 2.8 plotted in log units for better visual representation and it shows how the error increases as the code values deviate from a reference scene white code value and approach the lower and higher code values associated with the toe and shoulder of the film transfer function or characteristic curve. This is an example of the errors introduced by incorrect approximations of a transfer function.

Equation 2.5 shows the approximation implemented in the CINEON system where IN is the input code value representing printing density, $refwhite$ is the reference scene 100% white code value and OUT is the predicted exposure.

$$OUT = 10^{(IN - refwhite)(\frac{0.002}{0.6})} \quad (2.5)$$

Another issue with the CINEON system has been the variability across

Figure 2.8: CINEON Exposure Prediction Log Error

This graph shows the CINEON model exposure prediction error due to wrong assumptions about the film's response function.

post-production facilities in the implementation of the CINEON equations and necessary calibration of film scanners to produce accurate printing density code values [9]. These reasons, along with new technologies with extended color gamuts and dynamic range, prompted the industry to look for other solutions and eventually develop the ACES platform with the hopes that it could mitigate some of the inconsistencies in the CINEON system and also support future developments.

2.3.4 ACES

The ACES platform aims to enable the unambiguous interchange of images from film and digital sources. Figure 2.9 shows a high level diagram of ACES from origination to display including every possible input and output.

ACES is designed to be a colorimetric appearance-based color space and a high precision encoding for mastering, interchange and archiving to be used as a standard for all motion picture images. This is key, since adopting this encoding aims at having a seamless interchange of material from dif-

Figure 2.9: ACES Workflow

Overall ACES diagram including film and electronic inputs. The reference input device is also shown as producing ACES values directly.

ferent sources. To this end, ACES defined a reference input capture device (RICD). This RICD is a theoretical camera that would directly generate ACES values for all images captured without any additional processing and serves as the aim model for any other motion picture camera entering an ACES workflow. As can be seen in figure 2.9, part of the ACES workflow includes the input device transform (IDT). The IDT takes care of transforming the output of a real motion picture camera into the output of the theoretical RICD, making it interchangeable with any other material also encoded in ACES. An essential part of the computation of this IDT is knowing the CRF along with the white balance condition, spectral sensitivity and ISO settings. It can also be approximated by taking a known set of exposures with the camera and comparing them to the equivalent exposures that would be captured with the RICD. Figure 2.10 shows the general steps in calculating the IDT for a digital

Figure 2.10: ACES IDT Implementation Steps

Steps to implement an IDT to images from a real electronic camera. Steps include correcting for the CRF (LUT), white balancing the signals, clipping the signals to allowable values, color space and chromatic adaptation implemented through the 3x3 IDT matrix, and a final scaling of values for ACES.

motion picture camera and are further explained in section 3.3.

The CRF has to be known so that camera code values can be converted to radiometrically proportional linear values indicative of light hitting the sensor for further conversion to ACES values through a linear matrix transformation operation. In most cases the CRF is not represented accurately because access to scientific measuring techniques is simply not available in a production or post-production environment or because is being approximated by a simple approach, like in the CINEON system.

2.3.5 Image Interchange Challenges

The motion picture industry has constantly been faced with the challenge of how to best accomplish the interchange of images from different sources. As presented in the previous sections, knowing the CRF is critical to understanding the relationship between the camera's encoded code values and the scene

linear radiometry. Standardization of the CRF would be an obvious solution for every camera manufacturer to follow, unfortunately there are different standards that specify slightly different CRFs and existing image processing systems tend to trivially approximate them, causing errors in subsequent calculations. Furthermore, for industry differentiation purposes, newer camera systems that expand on video imaging standards, have turned to proprietary non-standardized CRFs in their image processing, leading to further challenges in post-production.

3 Radiometric Calibration Background

3.1 Introduction

Radiometric calibration refers to the computation of the CRF, or inverse CRF, for a specific capture system with the purpose of understanding the mathematical relationship, typically non-linear, between linear scene radiometry and encoded digital image code values. The standard practical approach is to photograph a variable reflectance or transmittance chart with known measured radiance/luminance values and relate these radiance/luminance measurements to the obtained pixel code values [12]. This approach is very practical in a research or academic environment but it's not practical in a motion picture production or post-production environment.

This section presents current state of the art methods of estimating the CRF based on different image statistics without having access to direct CRF measurements as well as a description of the CRF mathematical model used in many of the estimation methods including this research. In addition, relevant probability fundamentals are reviewed. Lastly, a state of the art methodology, independent of many image capture constraints as is done in many existing methods, is described and presented as starting point for the novel methodology presented further on in this dissertation.

3.2 Camera Response Function Estimation

Aside from measuring the CRF through direct photography of a variable reflectance or transmittance chart, some of the most popular methods for estimating the camera CRF involve taking multiple registered images of a static scene while varying the camera exposure and iteratively solving a system of equations for both the CRF and the exposure ratio unknown variables as presented in [13] and [14]. Work has also been done by Grossberg and Nayar [2] to relax this dependency on multiple spatially registered images by using histograms of images at different exposures. Manders, C. et al. [15] estimate the response by capturing registered images of a static scene illuminated by different combinations of light sources and locking the exposure instead of varying it. Similar work was presented by Kim et al. [16] where the estimation is done based on a video sequence with varying exposure. All of these methods require significant effort to obtain a series of registered, static scene or exposure dependent images.

Other approaches, like Farid [17], assume that the response function is in the form of a gamma curve, or power function, to estimate it from one single image by exploiting the fact that the non-linearity introduced by the gamma correction also introduces higher-order correlations in the frequency domain that can be detected with tools from polyspectral analysis. However, many cameras in use today in the motion picture industry have CRFs that differ significantly from a gamma curve, especially higher end motion picture cameras where their design deviates from the typical video imaging standard. Lin et al. [3] estimate the response by looking at the edge color distributions in a single image. Their results depend heavily on the kinds of edges selected

and noise and compression artifacts have a an impact on the accuracy of the estimation. Nonetheless, this work serves as motivation for this research with the aim to expand its methodology to analyzing edge color distributions in a sequence of motion images, identify additional image features that affected by the CRF in addition to color edge distributions and expand their Bayesian probabilistic approach to include the estimation of the posterior probability distribution function.

More recent work has also focused on using multiple images, especially photo collections. Diaz and Sturm [18] use a set of images from an internet collection to recover the camera’s geometric calibration and a 3D scene model, which in turn are used as input to determine the camera radiometric response function. Obviously, the requirement to obtain the scene 3D model would be an undesired workflow step for motion picture post-production workflows focused on image integration and color correction. Shafique et al. [19] introduce a method that uses differently illuminated images and estimate the response function by assuming that the properties of materials in a scene should remain constant and use cross-ratios of image values in the different color channels to compute the response function. They also model the response function as a gamma curve only. This again excludes the modeling of cameras that have CRFs that differ from a gamma curve and their algorithm was only verified by synthetic experiments. In [20], Kuthirummal et al. found priors for statistics in large photo collections and use them to estimate the response function of generic camera models assuming that all instances of a camera model have the same properties and that many images are available for that specific camera. They empirically found that the joint histogram of irradiance

values for a specific color channel, looks very similar across different camera models, based on the symmetric Kullback-Leibler divergence between corresponding histograms. They consequently use this similarity of the histograms computed for any camera model as a non-parametric prior on these statistics. A major disadvantage of their method is that it doesn't allow modeling of cameras with interchangeable lenses. This is also an impractical approach for the motion picture industry where lens changes are necessary. In [21], Badki et al. combine intensity mapping functions without the need for pixel to pixel correspondence with a rank minimization approach to perform the radiometric calibration. The drawback of their approach is also the need of multiple exposures from the scene. The approach by H. Lin et al. in [22] presents a novel method for compensating for gamut mapping operations before estimating the CRF, but requires an extensive calibration procedure including over 10,000 images from over 30 cameras. Lastly, in [23] Lee et al. present a method based on the invariant low-rank structure of sensor irradiance values from a static scene recorded with a series of different exposure times.

The existing methods described above, with the exception of Lin et al., suffer from general practical limitations in their application to the motion picture industry. They are restrictive in the position or movement of the capture device, they are restrictive in the exposure required for the camera sensor, they assume a simplistic mathematical formulation of the CRF that does not characterize all possible CRFs or they require co-capture of images and some additional scene information. This research again aims to expand on the work of Lin et al. to avoid these limitations and provide a solution based on unconstrained captured images that are typical of the motion picture industry.

In addition, many of the methods described use the CRF model proposed by Grossberg and Nayar [24] based on principal component analysis where a monotonic CRF can be well modeled with a linear combination of a few basis functions. This model will also be used initially in this research and is described in the next section.

3.2.1 Camera Response Function Model

As shown in equation 2.2, the CRF f relates captured scene radiance L , or proportional image irradiance E , to its measured intensity M in the image, represented by the pixel code value. Typically, radiometric calibration methods solve for the inverse response function $g = f^{-1}$, which is invertible since the sensor output increases monotonically with respect to E . Previous works relying on multiple exposures compute the inverse response function based on the relationship:

$$g(m_A) = rg(m_B) \quad (3.1)$$

where m_A and m_B represent image intensities in images A and B respectively for corresponding points and r denotes the exposure ratio between the two images. Requirements like precise registration of images or known exposure ratios have again been worked around with the use of histogram equalization and the use of iterative methods, referenced in the previous section, to solve for r and g .

One important obstacle in computing the camera radiometric response function from equation 3.1 is the exponential ambiguity, also known as u-ambiguity. From equation 3.1 it can be seen that if g and r are solutions for a specific image set, then g^u and r^u could also be solutions. To deal with

this ambiguity, these prior methods require a good initial estimate of the exposure ratio, if unknown, and assumptions on the structure of the radiometric function model as presented by Grossberg and Nayar [2]. Their results showed that the relationship between image intensity $M = f(E)$ and image irradiance E can be expressed as a linear combination of an average CRF h_0 and principal components h_n . They called this model the Empirical Model of Responses (EMoR) and based it off of a Database of Response Functions (DoRF) containing 201 inverse response functions from a variety of digital cameras and films up to the year 2003. The model is described by equation 3.2:

$$f(E) = h_0(E) + \sum_{n=1}^N w_n h_n(E) \quad (3.2)$$

Although other mathematical functions, like logarithmic or power functions, can be used to model a CRF, Grossberg and Nayar's model offers the most flexibility to approximate the CRF. This is important because camera manufacturers can deviate from pure logarithmic or power functions when implementing a CRF in their image processing workflow.

Random variables X and Y

Figure 3.1: Probability Diagram

Diagram to visualize fundamental probability concepts and derive the product and sum rules with regards to two random variables.

3.3 General Probability Review

Probability theory plays a key role in the field of machine learning. It provides a way to quantify and manipulate uncertainty by following a set of rules. These rules of probability are best explained by deriving them from the following example adapted from [25]. Let's consider two random variables X and Y and the diagram in figure 3.1.

Random variable X can take any of the values x_i where $i = 1, 2, \dots, M$, and Y can take the values y_j where $j = 1, 2, \dots, L$. Now let's consider N sampling trials of X and Y , and let n_{ij} be the number of trials in which $X = x_i$ and $Y = y_j$. In addition, the number of trials for which $X = x_i$ irrespective of Y , is c_i , and the number of trials for which $Y = y_j$ irrespective of X , is r_j . The probability that X will take the value x_i and Y will take the

value y_j is called the joint probability of X and Y , $p(X = x_i, Y = y_j)$. This is represented by the number of points in cell i, j as a fraction of the total number of points, N , considering the limit $N \rightarrow \infty$. Equation 3.3 shows this.

$$p(X = x_i, Y = y_j) = \frac{n_{ij}}{N} \quad (3.3)$$

In the same way, $p(X = x_i)$ represents the probability that X is equal to x_i irrespective of the value of Y . It is given by the fraction of points that fall into column i in figure 3.1 and described by equation 3.4.

$$p(X = x_i) = \frac{c_i}{N} \quad (3.4)$$

By noting that $c_i = \sum_j n_{ij}$, equation 3.4 can be written in the following form.

$$p(X = x_i) = \sum_{j=1}^L p(X = x_i, Y = y_j) \quad (3.5)$$

Equation 3.5 is known as the sum rule of probability and $p(X = x_i)$ is referred to as the marginal probability. Considering the fraction of instances for which $Y = y_j$ when $X = x_i$, referred to as the probability of $Y = y_j$ given that $X = x_i$, is called the conditional probability and it is obtained by finding the fraction of points in column i that fall in cell ij . Equation 3.6 shows this.

$$p(Y = y_j \mid X = x_i) = \frac{n_{ij}}{c_i} \quad (3.6)$$

Combining equations 3.3, 3.4 and 3.6, we can derive the following relationship, known as the product rule of probability.

$$p(X = x_i, Y = y_j) = \frac{n_{ij}}{N} = \frac{n_{ij}}{c_i} \cdot \frac{c_i}{N} = p(Y = y_j | X = x_i) p(X = x_i) \quad (3.7)$$

By the symmetry property for probabilities, $p(X, Y) = p(Y, X)$, and taking the product rule into account, the following relationship can be derived for conditional probabilities. This relationship is known as Bayes' theorem, and is presented next in a simplified and generic notation.

$$p(Y | X) = \frac{p(X | Y) \cdot p(Y)}{p(X)} \quad (3.8)$$

Using the sum rule, the denominator in equation 3.8 can be expressed in the following form, representing the normalization constant to ensure that the probability on the left hand side of the equation over all the values of Y equals one.

$$p(X) = \sum_y p(X | Y) p(Y) \quad (3.9)$$

One known statistic to review for this research is the Pearson correlation coefficient, which indicates the extent to which two random variables are linearly related. The Pearson correlation coefficient is defined to have a value between -1 and 1, where -1 represents a total negative correlation, 0 represents no correlation, and 1 represents a total positive correlation, and it is calculated by computing the covariance of two random variables, divided by the product of their corresponding standard deviations. Equation 3.10 defines the Pearson correlation coefficient for a population of the random variables

X and Y , where the covariance of X and Y is expressed in terms of means, μ_X and μ_Y , expectation and standard deviations, σ_X and σ_Y .

$$\rho_{X,Y} = \frac{\text{cov}(X,Y)}{\sigma_X \cdot \sigma_Y} = \frac{\text{E}[(X - \mu_X) \cdot (Y - \mu_Y)]}{\sigma_X \cdot \sigma_Y} \quad (3.10)$$

The derivation of the Bayes' theorem brings a different interpretation of probabilities compared to the more traditional view of probabilities as frequencies of random, repeatable events, which is known as the frequentist approach. In the Bayesian approach, probabilities provide a way to quantify uncertainty about an event. The main significance of the Bayes' theorem is that it allows the conversion of a prior probability into a posterior probability by taking into account observed data from an uncertain event. The prior probability represents the assumptions made about an event without any observed data, $p(Y)$. Once data about the event is obtained, it can be taken into account and is expressed as a conditional probability known as the likelihood function, $p(X | Y)$. The likelihood function expresses how probable is the data X , for different values of parameters Y . This means that Bayes' theorem can be expressed in words as *posterior* $\propto$ *likelihood* $\times$ *prior*. This Bayesian framework is again the foundation for the different probability estimation methodologies presented in this research.

3.4 Radiometric Calibration Through Edge Color Distributions

The following sections present the radiometric calibration methodology followed by Lin et al. [3]. Lin's method takes advantage of non-linearities introduced by a CRF at color edges in one single image. This is advantageous since it very easily fits into a motion picture post-production workflow and avoids dependency on multiple images, different exposures, multiple images or any additional calibration data. This dependency on a single image, made this methodology the preferred starting point for this research.

3.4.1 Edge Color Distributions

Radiometric calibration from a single image can't be accomplished by the use of equation 3.1 because of the dependency on different exposures. To avoid this dependency, the relationship between the CRF and measured edge color distributions can be studied. Single sensor (CCD or CMOS) cameras, like most used in the motion picture industry, require a color filter array (CFA) to capture color images. Each array element, or pixel x , images a solid angle of the scene and the whole array of pixels is $S(x)$. For each x , the image irradiance E depends on the sensitivity q_k of the pixel and filter pair, where k is the color of the filter, and the incoming scene radiance L incident on the image plane points p included in $S(x)$:

$$E(x, k) = \int_{\lambda_k} \int_{S(x)} L(p, \lambda) q_k(\lambda) dp d\lambda \quad (3.11)$$

Here λ represents the range of wavelengths that are transmitted by the

color filter k and it is assumed that all three colors red, green and blue are measured accurately because of effective demosaicing in the camera. Demosaicing refers to the mathematical process of estimating missing primary color values at each pixel position due to each pixel only capturing one primary color with a CFA. It is important to also recognize that image noise or non-optimal demosaicing can impact any calculations where these edge color distributions are taken into account.

3.4.2 Edge Color Boundary Detection

The detection of edges in the image is done by applying a Canny edge detection algorithm as described in [26]. The algorithm utilizes an intensity gradient detection method with a threshold parameter to control the edge detection sensitivity after applying a Gaussian filter to reduce the influence of noise in the detection. For edge color distributions detection, the threshold is set empirically to a high value to allow only true edges to be detected and minimize the amount of noise or non-edge textured features from being mistaken and detected as edges. Figure 3.2 shows an example edge map.

3.4.3 Edge Colors Analysis

To ensure proper edge color distributions, a color analysis is performed on the candidate edges and the mean and variance values for each of the two candidate color regions in each patch are computed. If the variance for each color region falls below a certain threshold, the areas are each considered uniform. In addition, the two areas' color mean values must lie at a specified Euclidian RGB distance from each other to avoid noise influencing the distance computation. Last, the edge colors must also lie within the color range delimited by

Figure 3.2: Canny Edge Detection Map
High threshold Canny edge detection to ensure only true edges considered in
edge color distributions

Figure 3.3: Measured Color Non-linear Distribution at Image Edges (Adapted from Lin et al.)

Diagram showing the CRF impact on edge color distribution triplets. The linearity exhibited in 3D space, turns to a non-linear relationship when a CRF is applied to the images.

the two edge regions. If all these conditions are met, the color triplet is added to the dataset used for the CRF estimation.

3.4.4 Transformation of Edge Color Non-linearities to Linear Distributions

To analyze edge color distributions two regions of distinct but uniform color are considered, like in figure 3.3.

Since image sensors have limited spatial resolution, a pixel x imaging a scene area where the two regions come together on an edge, will be receiving scene radiance $L_1(\lambda)$ and $L_2(\lambda)$ respectively from region 1 and region 2, and the overall radiance incident on it can be expressed as:

$$\int_{S(x)} L(p, \lambda) dp = \int_{S_1(x)} L_1(\lambda) dp + \int_{S_2(x)} L_2(\lambda) dp$$

$$= \alpha L_1(\lambda) + (1 - \alpha) L_2(\lambda) \quad (3.12)$$

where $\alpha = \int_{S_1(x)} dp$ and $S(x)$ is of unit area. Looking back at equation 2.2 and substituting equations 3.11 and 3.12 into it, the measured color at pixel x will be:

$$m(x, k) = f[\alpha E_1(x, k) + (1 - \alpha) E_2(x, k)] \quad (3.13)$$

For a linear relationship between the image irradiance E and the measured color $f(E)$ to exist, the following property has to hold true:

$$f[\alpha E_1 + (1 - \alpha) E_2] = \alpha f(E_1) + (1 - \alpha) f(E_2) \quad (3.14)$$

This means that the measured colors of pixels that lie on an edge, would fall on a straight line in the RGB color space between both regions. But, since f is typically non-linear, the measured edge colors no longer fall in a straight line, and this feature can provide information about the CRF. This non-linearity of the measured edge color distributions is used to estimate the CRF, based on the assumption that the inverse CRF should transform the measured values back into values linearly related to image irradiance. For an edge region with measured colors M_1 and M_2 , the inverse CRF, g , should map the measured color M_p of each point p in the region to a line defined by $g(M_1)$ and $g(M_2)$. Figure 3.4 shows this graphically. A function g satisfies this property if the distance from $g(M_p)$ to line $g(M_1)g(M_2)$ is zero as described by equation 3.15 where $\times$ is the cross product operation between the two

Figure 3.4: Non-linear to Linear Distribution Transformation (Adapted from Lin et al.)

The correct inverse CRF will restore the linear edge color distributions relationship in 3D space.

vectors.

$$\frac{|[g(M_1) - g(M_2)] \times [g(M_p) - g(M_2)]|}{|g(M_1) - g(M_2)|} = 0 \quad (3.15)$$

All obtained edge color triplets are combined into a set $\Omega_c = \{M_1, M_2, M_p\}$, and the total distance is then defined as:

$$D(g; \Omega_c) = \sum_{\Omega_c} \frac{|[g(M_1) - g(M_2)] \times [g(M_p) - g(M_2)]|}{|g(M_1) - g(M_2)|} \quad (3.16)$$

The best inverse response g will be the one that gives the smallest total distance among all edge color distributions selected. The relationship between the estimated irradiance given by g and the absolute irradiance hitting the camera sensor will have unknown scaling factors for each color channel. To deal with these differences, the domain and codomain of g are normalized so that $g(0) = 0$ and $g(1) = 1$ as it is done in [24]. The model used for g is based

on the EMoR from real-world camera response information compiled into the DoRF from Grossberg and Nayar’s work [24]. The DoRF is shown in figure 3.5. Using the PCA based CRF model, introduced in section 3.2.1, the inverse response function g is modeled by:

$$g = g_0 + Hc \quad (3.17)$$

where g_0 is the mean inverse response, H is a matrix with columns representing the eigenvectors and c is a coefficient vector that represents an inverse response function. Having defined the CRF model and the edge color distribution distance features, Lin’s method implements the Bayesian approach described next to estimate the CRF.

3.4.5 Probabilistic Framework

Probability based estimation methods have been used for many years and in many applications with two popular solution methods to Bayes’ theorem, Maximum Likelihood (ML) and Maximum A Posteriori (MAP). In the case of ML, the result is process most likely to have generated the observed data, whereas in MAP the result is the process that is most likely, given the observed data [27]. An important difference is that MAP estimation applies Bayes’ theorem, shown again in equation 3.18 with relevant variables to this research, so that the estimate can take into account prior parameter knowledge in the form of a prior probability distribution, $p(g)$, while evaluating the likelihood of the data, $p(\Omega_c|g)$, generated by those parameters. MAP can be considered an improvement over ML, but both, ML and MAP, give only single point best estimates and not a distribution of the parameters in question. On the other

Figure 3.5: Database of Response Functions
Set of CRFs collected by Grossberg and Nayar. This database included known CRFs up to 2003.

hand, a benefit is that these methods allow the computation of this single best estimate in a fast and efficient manner.

$$p(g|\Omega_c) = \frac{p(\Omega_c|g)p(g)}{p(\Omega_c)} = \frac{p(\Omega_c|g)p(g)}{\int_g p(\Omega_c|g)p(g)dg} \quad (3.18)$$

3.4.6 Prior Probability Distribution

Following this Bayesian approach, a prior probability distribution model can be created based on the EMoR PCA coefficients compiled by Grossberg and Nayar [24]. This prior probability distribution, $p(g)$, describes prior knowledge about the EMoR mathematical description and dimensional space of the CRFs and is modeled as a Gaussian Mixture Model (GMM), with means and covariances μ_i, Σ_i for each mixture component i and mixture proportion α_i . In this implementation the prior was precomputed using the Expectation Maximization (EM) algorithm, briefly described next.

$$p(g) = \sum_{i=1}^K \alpha_i N(g; \mu_i, \Sigma_i) \quad (3.19)$$

3.4.7 Gaussian Mixture Model

A Gaussian Mixture Model can serve as a clustering algorithm where each cluster is assumed to have their own independent Gaussian distribution. In this research, the GMM is applied to the set of coefficients obtained in the EMoR. The coefficient clusters can provide prior probability information about the DoRF. To obtain these clusters, the EM algorithm is used. The process starts with an initialization step where random points are selected to be the means of each cluster and the covariance of each cluster is selected to be the covariance

of the whole data set. After this initialization step, two additional steps are performed iteratively until convergence. The first step is the Expectation step, where the probability of each data point belonging to a each cluster is calculated. The next step is the Maximization step where the means and covariances of each cluster are recalculated based on the probabilities obtained in the previous Expectation step.

3.4.8 Likelihood Probability Distribution

From equation 3.16, the estimated inverse CRF should yield a low total distance for the correct inverse CRF. The likelihood probability distribution, $p(\Omega_c | g)$, can then be modeled by incorporating this distance calculation into an exponential distribution, with λ set empirically and Z being a normalization constant:

$$p(\Omega_c | g) = \frac{1}{Z} \exp(-\lambda D(g; \Omega_c)) \quad (3.20)$$

3.4.9 Bayesian MAP Solution

After modeling the prior and the likelihood functions, the problem can be formulated in a Bayesian framework following equation 3.18, with an MAP solution approach as described in equations 3.21 and 3.22. The optimal response function for data set Ω_c , is then:

$$g = \arg \max p(g | \Omega_c) = \arg \max p(\Omega_c | g) p(g) \quad (3.21)$$

Or taking the log of equation 3.21, g can also be written as:

$$g = \arg \min \lambda D(g; \Omega_c) - \log p(g) \quad (3.22)$$

This again represents a single point best estimate of the inverse CRF, g .

3.4.10 Ensuring Monotonicity

As presented in section 3.2.1, $g = f^{-1}$, and it is invertible due to the monotonicity of the output of an electronic sensor with respect to the irradiance falling on it. To ensure that the estimate of g satisfies this, monotonicity is imposed by following the method presented by Grossberg and Nayar [24]. For the function to be monotonic, its derivative must be positive. This can be expressed by defining a discrete derivative matrix V , and based on the EMoR, this monotonicity constraint becomes a system of inequalities presented in matrix form in equation 3.23.

$$V g_{mon} \geq 0 \quad (3.23)$$

Following equation 3.17, g_{mon} is of the form $g_{mon} = g_0 + Hc_m$, where the coefficient vector c_m is then determined with the constraint $VHc_m \geq -Vg_0$. This becomes an optimization problem to minimize the squared error as follows:

$$c_m = \operatorname{argmin}_c \| Hc - g - g_0 \|^2 \quad (3.24)$$

3.4.11 Camera Response Function Estimation Ambiguity

Previous radiometric calibration methods have dealt with CRF estimation ambiguity, especially the exponential ambiguity presented in section 3.2.1, by estimating exposure ratios and making assumptions about the form of the

CRF. A similar ambiguity can arise in this color edge distributions approach when the color triplets distance function in equation 3.16 is minimized by more than one CRF. This can happen when the distributions lie along the $R=G=B$ axis (neutral colors edge). Since it is quite unlikely for all the edges in a color image to be a purely neutral color, it is also unlikely for this ambiguity to exist as long as black and white images are not processed through this methodology. In addition, knowledge about the shape of the CRF along with the monotonicity constraint, allow this method to avoid a potential ambiguity.

3.5 Summary

This research focuses on developing a Bayesian probabilistic model to estimate the CRF building off of a state of the art single image estimation method based on edge color distributions of a single image. Past state of the art research has focused on ML and MAP solutions to the Bayesian framework as described in this section. The benefits of this approach present a great starting point for our methodology with the implementation of a MAP solution to the statistical inference problem, but at the same time, open the door to taking advantage of current computing resources and additional image data, to develop a new Bayesian methodology and approximate the posterior probability distribution instead of only calculating an MAP single point solution.

4 Novel Radiometric Calibration

4.1 Introduction

As discussed before, the methodology presented in the following sections is inspired by the work of Lin et al. [3], and aims at expanding their approach to a Bayesian framework that takes advantage of the information contained in the posterior probability distribution and not only in a single point estimate like in ML or MAP. Probabilistic solutions like ML and MAP are viable solutions to the CRF estimation problem, but don't take full advantage of all the process variables' information contained in a posterior probability distribution.

Computing the posterior probability distribution though comes with the challenge of computational complexity due to the potential high dimensionality of the problem. This is addressed in this work by implementing a sampling algorithm to obtain an estimate of the CRF. In addition, this section also includes the description of a novel image feature based on noise measurements, that is incorporated in the likelihood function of the probabilistic framework to estimate the CRF.

4.2 Radiometric Calibration with Multiple Frames

The first expansion to Lin’s work focused on implementing the MAP CRF solution methodology presented in section 3.4, but using multiple frames to expand the number of edge color distributions considered in the distance calculation in equation 3.16. The main objective was to minimize impact of image noise relative to number of edge color distributions and look to improve the CRF estimation result.

The implementation expanded the amount of edge color distribution information included in set Ω_c from color edges in one frame, to color edges in 10 and 20 frames respectively. The likelihood and prior probability distribution functions were formulated as in section 3.4 as well as the MAP solution approach following equation 3.22. Details about the experimentation results are included in the results section of this document and published in [28].

4.3 Noise Image Features for Radiometric Calibration

The next step in this research included the expansion of the Bayesian framework to include the estimation of the posterior probability distribution function of the CRF through a Markov Chain Monte Carlo sampling process. In addition an update to the DoRF with current motion picture camera CRFs and additional gamma function models was also performed, expanding the database from 201 entries to 277. The CRF PCA model described in section 3.2.1 is based on data acquired from imaging systems up to the year 2003. With many different camera systems having been developed since then, an update to the data used in this research was important. Following this DoRF update, thought was given to how noise not only affected edge color distributions, but also to the relationship of noise measurements and average intensity measured code values. A novel idea to use image noise as an image signature containing information about the CRF was consequently investigated and implemented as part of the methodology. The following sections describe these noise image features, their relationship to the non-linearity introduced by a CRF and the sampling approach to estimate the posterior probability distribution function.

4.3.1 Noise in Electronic Images

Following the previously described Bayesian framework, an innovative CRF estimation approach was investigated to again avoid the need for different exposures to estimate a CRF as required by equation 3.1, and avoid the potential influence of noise in edge color distributions. Instead of relying on the edge color distributions non-linearity introduced by a CRF, we can tackle the problem by looking at the non-linearity introduced in the relationship between

several image area intensity measured code values and the total noise in those image areas.

The total noise in the image data captured by an imaging sensor is a combination of several independent sources of noise. These include photon noise, read noise, which includes a fixed pattern noise component, thermal noise, pixel response non-uniformity noise and the quantization error as explained in [29]. Equation 4.1 shows how different independent noise sources, $N_1, N_2, N_3 \dots N_n$ are added in quadrature to contribute to the total noise N_t . In this research $n = 5$, as five sources of noise are considered.

$$N_t = \sqrt{N_1^2 + N_2^2 + N_3^2 + \dots + N_n^2} \quad (4.1)$$

Photon noise is the noise associated with the fluctuations around the average flux of photons arriving at an imaging sensor. These fluctuations are governed by Poisson statistics and they are equal to the square root of the average flux, or number of photons, counted by the imaging sensor. Equation 4.2 shows this relationship, where N_{Ph} represents the photon noise and P represents the photon count.

$$N_{Ph} = \sqrt{P} \quad (4.2)$$

Read noise, N_R , is associated with the fluctuations introduced by the electronic components present in the readout, gain and digitization circuitry of the sensor at each pixel. The digitized signal is ideally directly proportional to the photon count that arrived at each pixel. But in real camera systems, the read noise introduces a deviation in the raw measured code value. As an

example, the read noise can be isolated and measured by capturing a black frame. This implies taking a frame, or a few frames, with a lens cap on, to avoid light hitting the sensor, at the highest shutter speed, or shortest integration time, possible. Once this black frame is captured, the read noise can be calculated as the standard deviation of that frame, since all variation is solely due to the electronics and not to any arriving photons. A component of the read noise is the fixed pattern noise and it can be isolated by taking the average of J black frames, typically ten or more. This averaging has the effect of removing the variable component of the noise, reducing it by a factor of $\sqrt{J}$ relative to the fixed pattern component, and leave only a fixed pattern noise template, that can then be subtracted by the camera processing.

A third noise component is thermal noise. Thermal noise is due to the thermal agitation of electrons in a sensor pixel and they are indistinguishable from electrons generated by an exposure. The thermal noise tends to rise with exposure time, but this is not a concern for motion imaging applications as the exposure time is set consistent each time and it is restricted by the chosen frame rate of the motion imaging camera.

Pixel response non-uniformity (PRNU) is the next noise component considered, and it is related to the difference from pixel to pixel in the efficiency to capture and count arriving photons. PRNU grows proportionally with the exposure level and it is the most dominant noise component at high exposure levels if not accounted for in the camera processing. The relationship between PRNU and raw code values is given by equation 4.3, where N_{PRNU} represents the PRNU noise, v represents the variation in the response of each pixel and S_{AVG} represents the average raw code value.

$$N_{PRNU} = v \cdot S_{AVG} \quad (4.3)$$

The last noise component is the quantization error. This noise is due to the rounding off in the conversion from an analog voltage signal into a digitized raw code value. The contribution of this error to the noise is considered negligible in modern electronic motion imaging cameras and not considered further in this research.

In modern camera systems, some of these contributors to noise will be minimized by in-camera processing. As described above, the fixed component of the read noise can be eliminated by averaging a number of black frames. The PRNU can also be eliminated by a process called flat field correction since it is a fixed property of the imaging sensor that does not change from image to image. The process involves again taking J number of images of an evenly illuminated featureless surface. Since PRNU is most dominant at higher exposures, and the variable photon noise will also be an important contributor at those exposure levels, taking the average of J flat field images will reduce the variable photon noise by a factor of $\sqrt{J}$ relative to the PRNU. Now, PRNU is a variation on each pixel's response to light. The average flat field image represents a map of that variation and can then be used to divide the captured image raw pixel code values by the flat field image pixel code values, removing the variation in pixel to pixel response. This will also have the effect of dividing the raw pixel code values by the average illumination level of the flat field image L_{AVG} , so it will need to be restored in the correction. Equation 4.4 shows how both the fixed pattern correction and PRNU correction are implemented where I_C

is the corrected image, I_U is the uncorrected image, FP_T is the fixed pattern noise template and $PRNU_T$ is the flat field PRNU template. Applying the correction in this way also takes care of removing any camera specific raw code value offset applied to images, as the fixed pattern template would include the offset.

$$I_C = \frac{L_{AVG} \cdot (I_U - FP_T)}{(PRNU_T - FP_T)} \quad (4.4)$$

It is important to realize that although noise contributions from fixed pattern noise and PRNU can be minimized or eliminated in modern camera systems, contributions from the variable component of the read noise and from photon noise are going to be present in every image. Again thermal noise will be indistinguishable from photon noise but constant at a set exposure time and the quantization noise will be negligible. These facts allow the introduction of a novel image feature that can be used to estimate the CRF.

4.3.2 Linearity of Noise Features

As shown in the previous section, many sources of noise combine when capturing an electronic image, and some of these sources can be eliminated or discounted. The two most important sources of noise considered in typical exposures are read noise, N_R , and photon noise, N_{Ph} , and together they determine the image quality and signal to noise ratio at a given exposure. The two can be combined in quadrature to determine the total noise, N_T , as presented in equation 4.5.

$$N_T^2 = N_R^2 + N_{Ph}^2 \quad (4.5)$$

Photon noise is dictated by Poisson statistics as shown in equation 4.2, where the photon noise is the square root of the average photon count. Image data is accessible in the form of raw code values instead of photon counts and there exists a proportionality constant that relates this photon count and the raw code value. This is known as the gain, and it is represented with the variable ζ . This allows equation 4.2 to be rewritten in terms of raw code values (RCV). With ζ photons/RCV, the noise and the photon count are represented in raw code values as $N_{PhCV} = \frac{N_{Ph}}{\zeta}$ and $S_{CV} = \frac{P}{\zeta}$, and equation 4.2 can be rewritten in the following form:

$$N_{PhCV} = \sqrt{S_{CV}} \quad (4.6)$$

In addition, ζ is inversely proportional to the camera ISO setting. Equation 4.7 shows this relationship, where ISO represents the camera ISO setting and ISO_N is a constant that represents the native ISO of the camera, which is reported by the camera manufacturer following the ISO photographic speed standard ISO 12232 [4]. This quantity is also known as the Unity Gain.

$$\zeta = \frac{ISO_N}{ISO} \quad (4.7)$$

This results in the following relationship between the photon noise raw code value, the signal raw code value and the gain.

$$N_{PhCV}^2 = \left(\frac{N_{Ph}}{\zeta} \right)^2 = \frac{N_{Ph}^2}{\zeta^2} = \frac{P}{\zeta^2} = \frac{S_{CV} \cdot \zeta}{\zeta^2} = \frac{S_{CV}}{\zeta} \quad (4.8)$$

Now equation 4.5 can be expressed in the following form where the total noise, read noise and photon noise are obtained by measuring the noise and

signal in raw pixel code values. This allows the measurement of the noise at a variety of exposure levels, plotting the square of the noise versus the average raw code value and fitting it to a straight line, where the inverse of the slope is the gain controlled by the camera ISO setting, and the y-axis intercept is the read noise squared.

$$N_T^2 = N_R^2 + \frac{S_{CV}}{\zeta} \quad (4.9)$$

The process for analyzing the noise in a camera that provides access to raw code values can now be summarized in the following steps.

1. Obtain two images of the same scene and under the same exposure conditions and camera settings. If working on a controlled environment, a uniformly lit, slightly out of focus color checker chart provides a series of exposures with uniform patches where to measure the noise, but this is not absolutely necessary for this research. Taking the image slightly of focus ensures that any surface imperfections don't contaminate the measurement. The use of two successive images having the same exposure and subtracting them ensures the elimination of systematic effects like PRNU and subtle gradients that can also contaminate the measurement of the total noise. Obtaining two successive images for this purpose turns out to also be trivial when capturing motion images as long as there is not fast scene movements or fast camera movements. Additional edge-detection based pre-processing can be implemented to ensure the selection of uniform areas for the noise analysis and avoid the need for

specific noise measurement charts.

2. The two images are then added together and the raw code value average for each of the patches for each of the channels is measured. The resulting raw code values are then divided by two to obtain the average signal for each exposure in raw code values.
3. The two images are subtracted and the standard deviation is calculated in each of the color patches considered in step 2. The result is divided by $\sqrt{2}$ to obtain the combined read and photon noise for each of the patches. This division adjustment by $\sqrt{2}$ is necessary due to the fact that independent noise sources add in quadrature, so in combining the two images by subtracting them, the resulting standard deviation measurements will be $\sqrt{2}$ times the standard deviation of a single image.
4. Finally, a plot can be created of the square of the total noise obtained in step 3 vs. the average signal raw code values from step 2. This plot will be a straight line where the y-axis intercept is the read noise squared and the slope of the line is the inverse of the camera gain.

Figure 4.1 is an example of the straight line total noise plots for the red, green and blue channels following the steps listed above. The image analyzed was of a MacBeth color checker chart and its 24 patches following this procedure to show the noise squared vs. raw code value linearity described in this section.

Any non-linear function applied to the raw code values to produce the measured color as shown back in equation 3.14, will also disrupt the linear

Figure 4.1: Total Noise vs. Raw Code Value
Linear relationship between the square of the total noise versus the measured pixel code value for channels R, G, and B.

Figure 4.2: Effect of CRF on Noise vs. Raw Code Value Relationship
Non-linear relationship between the square of the total noise and the measured pixel code value for channels R, G, and B. This is introduced by the CRF.

relationship presented in equation 4.9 between the total noise squared and the raw code value. Figure 4.2 shows the effect of applying a non-linear CRF to the image and how the linear relationship in equation 4.9 doesn't hold any more. This non-linearity of the total noise squared vs. raw code values is then used to estimate the CRF, as processing the image through the inverse CRF will restore the linear relationship.

It's important to highlight that for a different camera gain ζ , or different ISO setting, the slope of the linear relationship presented in figure 4.1 will vary, but the linear relationship between total noise squared and raw code

value will still hold, making this a very convenient image feature for analyzing images captured at different ISO camera settings.

4.3.3 Pearson Correlation Coefficient and Pearson's Distance

Looking at equation 4.9, it is necessary to define a metric that indicates the level of linearity between the noise squared measurements and the corresponding raw code values. With this in mind, the Pearson correlation coefficient introduced in section 3.3 can be used to measure the linearity between the noise squared variable and the raw code value variable as described in [30]. The Pearson correlation coefficient for the noise squared vs. raw code value plots presented in figure 4.1 is 1, representing a total positive correlation. Figure 4.3 shows the sample Pearson correlation coefficient for various examples of a scatter diagram between random variables X and Y .

In the case of sampling a population, as it's the case when analyzing a select number of image patches, the Pearson correlation coefficient is known as the sample Pearson correlation coefficient and it is derived from the sample mean and the sample standard deviation, both computed from the total number of samples. In addition, a distance metric for the random variables X and Y can be calculated from the sample Pearson correlation coefficient. This distance metric is known as Pearson's distance and is shown in equation 4.10, where d represents the Pearson's distance. Pearson's distance will have minimum value of 0, for a total positive correlation and a maximum value of 2 for a total negative correlation. This distance can then be incorporated into an exponential distribution, in the same manner described in section 3.4.8, maxi-

Figure 4.3: Sample Pearson Correlation Coefficient Scatter Diagram Examples
 Pearson correlation coefficient value ranges for different data patterns.

mizing the likelihood probability distribution when a total positive correlation exists.

$$d = 1 - \rho_{X,Y} \quad (4.10)$$

4.3.4 Noise Patch Detection

The procedure presented in section 4.3.2 computes the total noise utilizing a pair of images of a color checker, or similar chart, to ensure smooth non-textured image surfaces appropriate for noise measurements. This is a practical process to follow in a laboratory environment but very impractical for typical production workflows. This research focuses on computing the total noise on any non-textured image areas where the measurement can be done reliably. The task of detecting these areas is carried out in two general steps. First, an edge detection algorithm is used on the image after a smoothing kernel is applied. This step ensures that true edges and textured image areas are detected. Second, the edge map created in the edge detection stage is used to exclude those areas from where the noise measurement can be done reliably.

The detection of edges in the image is done by applying a Canny edge detection algorithm as described in [26] and in a similar way as in section 3.4.2. The algorithm utilizes an intensity gradient detection method with a threshold parameter to control the edge detection sensitivity after applying a Gaussian filter to reduce the influence of noise in the detection of edges. For the noise patches detection approach, the threshold is set empirically to a low value to allow not only edges but also high noise and non-edge textured features to be detected. This edge map is then used to exclude those areas when performing

the noise patch detection and subsequent total noise calculations.

The patches selected to be analyzed for noise are compiled into a set $\Omega_N = \{\Phi_1, \Phi_2, \Phi_3 \dots \Phi_n\}$, where Φ represents an image patch and n is the total number of patches. Various patch regions sizes were tested and negligible CRF estimation result differences were found for patch sizes 25x25 or larger, although the algorithm performance was impacted by patch sizes above 50x50. The patch size was empirically selected to 30x30 pixels for most experiments. The total number of patches was also empirically selected to be between 500-1000 patches per image. Figure 4.4 shows an example edge map and subsequent patches selected for noise analysis. Figures 4.5 and 4.6 show the non-linear noise features plot for the red, green and blue channels, and the linearized noise features plot respectively, after selecting a range of patches.

4.3.5 Bayesian Framework with Posterior Probability Distribution Estimation

ML and MAP methods produce single value solution estimates following Bayes' theorem. Instead, it is preferred, if possible, to compute the posterior probability distribution and calculate an expected value from such distribution, taking advantage of all the information included in the whole posterior distribution that ML and MAP discard. The caveat is that this becomes computationally harder as the dimensionality of the parameter space grows. To solve this, estimation methods can be implemented, offering practical solutions to the posterior probability distribution presented in equation 4.11.

$$p(g|\Omega_N) = \frac{p(\Omega_N|g)p(g)}{p(\Omega_N)} = \frac{p(\Omega_N|g)p(g)}{\int_g p(\Omega_N|g)p(g)dg} \quad (4.11)$$

Figure 4.4: Edge Map and Noise Patches

Example edge detected image and accompanying noise patch detection avoiding textured image areas identified by the edge detection algorithm.

Figure 4.5: Noise Features Plot for Non-Linear Image
Non-linear relationship between square of the total noise and the measured pixel value for R, G, and B channels, for ARRI camera image.

Figure 4.6: Noise Features Plot for Linearized Image
Non-linear relationship correction through inverse CRF, between square of the total noise and the measured pixel value for R, G, and B channels, for ARRI camera image.

4.3.6 Prior Probability Distribution

Similarly to section 3.4.6, a prior probability distribution is modeled based on the EMoR PCA coefficients compiled by Grossberg and Nayar [24], with the addition of current systems CRFs. This prior probability distribution, $p(g)$, describes prior knowledge about the mathematical description and space of CRFs and is modeled as a Gaussian Mixture Model (GMM), with means and covariances μ_i, Σ_i for each mixture component i and mixture proportion α_i . In this implementation the prior was also precomputed using the Expectation Maximization (EM) algorithm.

$$p(g) = \sum_{i=1}^K \alpha_i N(g; \mu_i, \Sigma_i) \quad (4.12)$$

4.3.7 Likelihood Probability Distribution

From section 4.3.3, the estimated inverse CRF should yield a low Pearson's distance for the right inverse CRF. The likelihood probability distribution, $p(\Omega_N | g)$, can also then be modeled by incorporating this distance, d , into an exponential distribution with again λ set empirically and Z being a normalization constant.

$$p(\Omega_N | g) = \frac{1}{Z} \exp(-\lambda d(g; \Omega_N)) \quad (4.13)$$

4.3.8 Posterior Probability Distribution

Bayes' theorem, shown in equation 4.11, describes how to calculate the posterior probability distribution function. The problem in calculating the posterior is that integrals, like the one in the denominator of equation 4.11, can be quite difficult to compute. Analytical solutions of the integral for a high dimensional

posterior distribution can be impossible to obtain and in turn, the exact form of $p(g \mid \Omega_N)$ can't be determined. One way to solve this problem is with solution methods like ML or MAP. The drawback again, is that these methods compute a single estimate and throw away information included in the posterior probability distribution. To overcome this, we can implement a sampling approach where we can draw samples from a distribution that asymptotically approaches the posterior, $p(g \mid \Omega_N)$, without having to calculate the integral in equation 4.11. To take advantage of the information included in the posterior probability distribution, we can then calculate an expected value for the function in question.

To first estimate the posterior probability distribution of g , we implement a sampling approach. Once the posterior probability distribution is estimated, an expected value is easily computed to find the coefficients to estimate the inverse CRF g . This expectation is shown in equation 4.14 for the general continuous case. For the discrete case, the integral turns into a summation.

$$E[f(z)] = \int f(z) p(z) dz \quad (4.14)$$

In this case z is the random variable, and $p(z)$ is the probability distribution over possible values of z . The goal is then to calculate the expected value of $f(z) = g$, which is the CRF, over the whole posterior probability distribution of g , $p(g \mid \Omega_N)$.

$$E[g] = \int g \cdot p(g \mid \Omega_N) dg \quad (4.15)$$

4.3.9 Markov Chain Monte Carlo Sampling

The problem in calculating the expected value in equation 4.15 arises from the fact that computing the integral can be an intractable problem due to high dimensionality[31]. Another way to think about this calculation is to sample K points $z^{(1)}, z^{(2)}, z^{(3)} \dots z^{(K)}$ randomly from the posterior probability distribution function $p(g \mid \Omega_N)$. This sampling will result in an unnormalized estimate of the posterior probability distribution, from which we can then obtain an expected value in the following form.

$$E_{p(z)}[g(z)] = \lim_{K \rightarrow \infty} \frac{1}{K} \sum_{t=1}^K g(z^{(t)}) \quad (4.16)$$

By thinking of the problem in this way, we can implement a Markov Chain Monte Carlo (MCMC) sampling algorithm, like the Metropolis-Hastings algorithm, to approximate the posterior probability distribution in equation 4.11 for g , given a set of noise patches pixel data Ω_N . This approach is, of course, also viable with the color edge distributions set Ω_c and that solution approach. The dimensionality of the sampling space is determined by the number of coefficients in the EMoR computed from the DoRF, since it provides a concise descriptor for the inverse CRF. The work by Grossberg and Nayar [24] showed that 3-5 bases modeled all the CRFs in the DoRF well, with 3 bases, or principal components, capturing at least 99.6% of the information in the DoRF.

MCMC is an approach for generating samples $z^{(t)}$ in equation 4.16, while exploring a state space Z using what is known as a Markov chain. The mechanism is constructed in such a way that the chain spends the most time

in the important regions of the state space. By important regions we mean that the chain produces samples that mimic samples drawn from a target distribution $p(z)$, a multidimensional posterior probability distribution. This is important because samples cannot be drawn directly from the distribution in question, but it can be evaluated up to a normalizing constant. This stochastic process $z^{(t)}$ is called a Markov chain if its evolution only depends on the current state and a fixed transition matrix U , meaning that $U \triangleq U(z^{(t)} | z^{(t-1)})$ remains invariant for all t . This is also known as a homogeneous chain.

$$p(z^{(t)} | z^{(t-1)}, \dots, z^{(1)}) = U(z^{(t)} | z^{(t-1)}) \quad (4.17)$$

The chain will converge to the invariant distribution $p(z)$ for any starting point, if the stochastic transition matrix U satisfies two properties. The first one is irreducibility. This means that for any state in the Markov chain, there is a positive probability to transition to all other states. The second one is aperiodicity. This means that the chain should not get trapped in any cycles. With these two properties satisfied, convergence is guaranteed.

The values obtained with the Markov chain must be representative of the posterior distribution, the chain should be of enough length so that the estimates are accurate and stable and last, it should be efficient so that the estimates are calculated with the least number of steps possible to avoid exceeding computing power or introducing impracticality [32]. To this end, estimates of central tendency (like mode or median) along with the limits of the 95% High Density Interval should not vary much when the MCMC is run again with different initial starting values. Following visual inspection of trace plots

Figure 4.7: MCMC Trace Plots for Mean and Standard Deviation for 30,000 ESS

Effective sample size trace plot example for mean and standard deviation of the estimated posterior probability distribution.

along with other guidelines and a 10,000 effective sample size (ESS) heuristic presented by [32], an ESS of at least 30,000 was selected for most experiments to ensure accuracy and stability while maintaining computational efficiency. Figure 4.7 shows example trace plots for mean and standard deviation of the MCMC samples to approximate a target probability function of 3 coefficients for the CRF PCA model.

4.3.10 Metropolis-Hastings Algorithm

The Metropolis-Hastings algorithm is one of the most well-known MCMC methods. A Metropolis-Hastings step of invariant distribution $p(z)$ and proposal distribution $q(z^* | z)$ involves sampling a candidate value z^* , given a current value z , and according to $q(z^* | z)$. The proposal distribution is an easy to sample distribution that helps to sample the state space while evaluating the probabilities of the target distribution. The chain moves towards new candidate value z^* with the following acceptance probability, staying at z otherwise.

$$A(z, z^*) = \min\left\{1, \frac{p(z^*) \cdot q(z \mid z^*)}{p(z) \cdot q(z^* \mid z)}\right\} \quad (4.18)$$

By choosing a symmetric random walk proposal distribution, like a multivariate normal distribution, $q(z^* \mid z) = q(z \mid z^*)$, the acceptance probability simplifies to the following equation. This special case is known as the Metropolis algorithm.

$$A(z, z^*) = \min\left\{1, \frac{p(z^*)}{p(z)}\right\} \quad (4.19)$$

The pseudo-code for the Metropolis-Hastings algorithm is shown next.

- Initialize $z^{(0)}$
- For $t = 0 : K - 1$
 - Sample $u \sim U_{[0,1]}$
 - Sample $z \sim q(z^* \mid z^{(t)})$
 - If $u < A(z^{(t)}, z^*)$
 - * $z^{(t+1)} = z^*$
 - Else
 - * $z^{(t+1)} = z^{(t)}$

With the implementation of this sampling approach, the posterior probability distribution, $p(g \mid \Omega_N)$, can be efficiently estimated and it allows for the expected value of the inverse CRF, g , to be computed based off of all the information included in the posterior probability distribution. Once the CRF is estimated, it can be used to linearize a real camera's images so they can be incorporated into an ACES workflow, either for the calculation of an IDT or for the implementation of an IDT as presented in the next section.

4.4 Summary

This section described key extensions to Lin’s methodology and novel contributions. First, the expansion from a single image to a sequence of images, introduced additional image data to include more edge color distributions in the likelihood probability distribution and minimize the impact of noise in edges only from a single image. This lead to identifying a noise based metric that is affected by the non-linearity of the CRF. This noise metric is based on the relationship between the square of the total noise contained in non-textured image areas, dominated by photon noise and read noise, and the measured code value. This relationship is a linear relationship for images that are radiometrically linear, but deviates from linear for images that have been captured and processed through a CRF. A Pearson distance metric was then incorporated into the likelihood probability distribution and Bayesian framework.

Along with the noise image features, the estimation of the posterior probability distribution is necessary to find a solution to the inference problem by computing an expected value from the posterior probability distribution to obtain the CRF. This was achieved by implementing an MCMC sampling algorithm known as the Metropolis-Hastings algorithm. This in turn allowed for the computation of the CRF by taking into account the estimated posterior probability distribution instead of just a single point estimate like in MAP solutions.

By combining the new noise based image feature and the MCMC sampling of the posterior probability distribution, a more robust and accurate estimation of the CRF was possible and demonstrated in the experiments dis-

cussed in this dissertation.

5 ACES Input Device Transform

5.1 Introduction

An accurate estimate of the CRF will enable the correct implementation or calculation of an IDT, which is necessary to input content from any capture device into the ACES workflow. The calculation of an IDT is described next and follows the guidelines in the AMPAS Recommended Procedure for computing IDT's [33, 1].

The first objective of the IDT is to convert radiometrically linear signals from any real camera to radiometrically linear signals as they would be captured by the RICD. This is the main reason for calculating an accurate estimate of the real camera CRF since the error in the optimization of the IDT is directly impacted by the accuracy of the image linearization through the inverse CRF. Also, any correctly calculated IDT matrix applied to non radiometrically linear image data, will cause a similar color error impact in the IDT processing result. The second objective of the IDT is to incorporate a treatment of the native camera white balancing such that a neutral reflector would produce equal camera signals under the illuminant for which the IDT is being designed. The calculation of an IDT is typically done in a research environment by a camera manufacturer or research group, and the results are then included in various post-production ACES workflow software packages. The key is that these workflows expect radiometrically linearized images before being processed through an IDT matrix. In the next sections, the process for computing an IDT is described to explain how the CRF can impact the accuracy of the IDT calculation or the subsequent ACES workflow

Figure 5.1: MacBeth Color Chart
Example of industry known chart used in the IDT computation.

IDT implementation.

The ACES literature recommends two general approaches for an IDT calculation based on the user's access to the camera's spectral sensitivity curves. In the case where the spectral sensitivity of the camera is available, the camera exposures for a set of training patches are computed vs. when the spectral sensitivity curves are not available, in which case the camera exposures can be obtained from exposing the camera to a set of known patches. One example set of patches is the well known and previously mentioned MacBeth chart, shown in figure 5.1, for which the spectral reflectances of the color patches are also well known and published. Other sets of known reflectance patches also exist and can be used.

5.2 IDT for Image Capture

The first step in the calculation of the IDT involves capturing a series of images from a set of known patches. If the camera produces non-linear signals due to a non-linear CRF, the data will need to be linearized. The next step is to white balance the images based on the scene illuminant, for which the IDT is being created. The white balance normalization scalars k_i are calculated by equation 5.1 where $S_{aw}(\lambda)$ is the light source power spectral distribution and $SS_i(\lambda)$ is the camera spectral sensitivity for each color channel i . These steps are typically carried out through image processing inside a camera.

$$k_i = \frac{1}{\int_{\lambda} S_{aw}(\lambda) SS_i(\lambda) d\lambda} \quad (5.1)$$

5.2.1 ACES IDT Cost Function

Next comes the generation of the IDT matrix used to transform the camera spectral sensitivities to the RICD native spectral sensitivities and apply a chromatic adaptation as described in [33]. The conversion of the spectral sensitivities is necessary so the output of the camera approximates the linear exposures that would be captured by the RICD. Additionally, chromatic adaptation is necessary since ACES is scene-referred appearance based colorimetry with D60 (noon daylight with 6000K correlated color temperature) defined as the light source. Chromatic adaptation is the process by which the human visual system adjusts its response to the average chromaticity of the existing light source. Images taken under any other light source and properly white balanced, have to then be chromatically adapted to represent the perception of those same scene colors viewed under a D60 light source.

The IDT matrix coefficients are computed by optimizing a cost function built with an appropriate color difference metric, following ACES guidelines. Equation 5.2 shows this cost function.

$$\text{delta}E = \arg \min (W \text{delta}E[\text{Lab}(T_{ca}, T_{D60}), \text{Lab}(M_{ACES} \cdot B \cdot \text{Exp}_{wb}, T_{D60})]) \quad (5.2)$$

One popular color difference metric used is the deltaE color differences metric, for which various versions exist, each aiming to improve on perceptual color difference prediction performance [5]. The color difference calculation is done between standard 1931 CIE tristimulus values, XYZ, for a set of training patches and the tristimulus values obtained after conversion of digital camera exposures of the same training patches. Both sets of tristimulus values are converted to the CIE L*a*b* color space for the calculation due to CIE L*a*b* being closer to a perceptually uniform color space. Measuring the accuracy of color difference metrics in regards to actual perceptual differences is an area of active research where newer color spaces, like ICtCp, have shown recently to be more perceptually uniform [34]. An important related measure is the minimum detectable perceptual difference, known as a just noticeable difference (JND). A deltaE of 1-2 is considered to be the threshold of color difference detectability for human observers, with a tolerance of up to 3 in many applications [34], and represents a practical minimum error aim for the IDT. The tristimulus values can be calculated according to the color-matching properties of the CIE 1931 Standard Colorimetric Observer [35]. Equations 5.3 and 5.4, show this calculation.

$$\begin{aligned}
X &= k \int_{\lambda} R(\lambda) S_{aw}(\lambda) \bar{x}(\lambda) d\lambda \\
Y &= k \int_{\lambda} R(\lambda) S_{aw}(\lambda) \bar{y}(\lambda) d\lambda \\
Z &= k \int_{\lambda} R(\lambda) S_{aw}(\lambda) \bar{z}(\lambda) d\lambda
\end{aligned} \tag{5.3}$$

$$k = \frac{1}{\int_{\lambda} S_{aw}(\lambda) \bar{y}(\lambda) d\lambda} \tag{5.4}$$

In equation 5.3, X , Y and Z are the 1931 CIE tristimulus values, $R(\lambda)$ is the patch reflectance, $S_{aw}(\lambda)$ is the light source power spectral distribution, $\bar{x}, \bar{y}, \bar{z}$ are the 1931 CIE color matching functions and k is the normalization term, which by convention, is the same in the calculation of X , Y and Z .

The cost function presented in equation 5.2 also includes W , which represents a set of optional weighting factors applied to the training patch set when more importance is given to a particular set of colors because of a specific application, like for example emphasizing the accurate reproduction of skin tones. T_{ca} are the chromatically adapted tristimulus values for the training spectra and T_{D60} are the tristimulus values for the ACES white D60, [0.953, 1, 1.009]. The chromatic adaptation transform predicts corresponding colors under the given illuminant D60, where corresponding colors refers to two colors which match in appearance when the first is viewed under a primary illuminant and the second is viewed under a secondary illuminant. The chromatic adaptation method used is comprised of two steps, first a primary

color space rotation from 1931 CIE tristimulus values to CIE CAT pseudo-cone fundamentals which redefine the 1931 CIE X, Y, Z colorimetric primaries in more relevant psycho-anatomical terms, $R_{CAT02}, G_{CAT02}, B_{CAT02}$. Second a von Kries transform is applied as presented in [35] where the chromatically adapted values are obtained by applying a diagonal matrix with diagonal values equal to the ratios of the red, green and blue colorimetric primaries for the adapted white, w_2 (D60 for ACES), to the red, green and blue colorimetric primaries for the scene white, w_1 , both also in CIE CAT space. These transforms are described by equations 5.5 and 5.6.

$$\begin{bmatrix} R_{CAT02} \\ G_{CAT02} \\ B_{CAT02} \end{bmatrix} = \begin{bmatrix} 0.7328 & 0.4296 & -0.1624 \\ -0.7036 & 1.6975 & 0.0061 \\ 0.0030 & 0.0136 & 0.9834 \end{bmatrix} \cdot \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \quad (5.5)$$

$$\begin{bmatrix} R_{ca} \\ G_{ca} \\ B_{ca} \end{bmatrix} = \begin{bmatrix} \frac{R_{w2}}{R_{w1}} & 0 & 0 \\ 0 & \frac{G_{w2}}{G_{w1}} & 0 \\ 0 & 0 & \frac{B_{w2}}{B_{w1}} \end{bmatrix} \cdot \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad (5.6)$$

Equation 5.7 shows the combination of these two matrix transformations to obtain the chromatically adapted 1931 CIE tristimulus values.

$$\begin{bmatrix} X_{ca} \\ Y_{ca} \\ Z_{ca} \end{bmatrix} = M_{CAT02}^{-1} \cdot M_{vK} \cdot M_{CAT02} \cdot \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \quad (5.7)$$

M_{ACES} is another matrix defined by ACES to convert from ACES RGB

tristimulus values to 1931 CIE tristimulus values and is shown in equation 5.8.

$$M_{ACES} = \begin{bmatrix} 0.9525523959 & 0 & 0.000936786 \\ 0.3439664498 & 0.7281660966 & -0.0721325464 \\ 0 & 0 & 1.0088251844 \end{bmatrix} \quad (5.8)$$

B represents the IDT matrix containing the coefficients, $\beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6$ being computed by the optimization and Exp_{wb} are the linearized digital camera exposures of the training patches white balanced to the scene illuminant. The IDT matrix B is shown in equation 5.9, where it can be seen that the structure of the matrix enforces row-unity normalization. This is appropriate to retain neutral balance of the scene white neutrals which have been white balanced.

$$B = \begin{bmatrix} \beta_1 & \beta_2 & (1 - \beta_1 - \beta_2) \\ \beta_3 & \beta_4 & (1 - \beta_3 - \beta_4) \\ \beta_5 & \beta_6 & (1 - \beta_5 - \beta_6) \end{bmatrix} \quad (5.9)$$

Examination of equation 5.9 shows that the optimal coefficients for matrix B will be obtained when the Exp_{wb} values are proportional to linear scene radiance, or equivalently, sensor irradiance values. This again reinforces the need for properly linearizing the Exp_{wb} values through the CRF prior to the IDT calculation. This research found that to ensure the linearity of the optimization data, a seventh parameter, β_7 , could be included in equation 5.2 as a power function exponent CRF correction to linearize the training data in the form $Exp_{wb}^{\beta_7}$. This allows the optimization to find the optimum value for a simple exponential model of the CRF, along with the IDT matrix coefficients. This would be beneficial during the calculation of an IDT, due to the

controlled environment and access to training image data, but not during the implementation where only production images are available. For the linearization of image data before the implementation of an IDT, the methodology presented in this work is recommended.

5.3 Summary

Understanding the full calculation of the IDT, gives evidence to the fact that non-linear measured pixel values will produce a larger color error metric when optimizing the IDT matrix. This is not only evident when computing the IDT matrix, but it can also translate to larger errors when implementing the IDT in post-production workflows on image data that is believed to be linear, but that in reality is not because the CRF is not known or is erroneously assumed to be a wrong function. To improve the process of computing an IDT and what's more important, to improve the adoption and implementation of IDTs in the industry, a method that can accurately estimate the CRT is beneficial, as it translates to a reduction in the time necessary to manually correct for color errors in post-production workflows.

6 Alternative Approach for CRF Estimation

6.1 Introduction

A Bayesian inference model takes the most advantage of the image data available providing us with an entire distribution of answers and because of this reason the methods described so far in this document were selected as the primary solution methods. It's important to also recognize that neural network (NN) based machine learning or deep learning, could also be considered as a viable solution the problem.

Potential approaches could focus in modeling distance functions like those in equations 3.16 and 4.10 with a neural network architecture, and subsequently introduce this result into a Bayesian solution framework. In addition, both traditional and Bayesian approaches can be implemented for the supervised training of the network weights. Lastly, a Bayesian framework could also be implemented to perform model selection between different neural network architectures. To provide some background on these approaches, a brief introduction to neural networks is presented in this section.

6.2 Neural Networks

Neural networks differ from typical computer programs by being able to “learn” from a set of training data. Generally a computer program will give a pre-defined answer when presented with data, but a neural network will generalize when presented with data extending outside of the training data. Neural networks have been proven to work well in pattern recognition and non-linear curve fitting between a set of inputs and outputs [36] and this is a reason for considering them as a viable solution for this research. Some essential aspects about neural networks processing are the highly parallel execution of computation and robust performance against noise, both of which are attractive properties. More recently, studies have shown the ability of neural networks to learn by reinforcement learning without direct human knowledge involved [37], and more generally, over the last 5 years, the field of machine learning has seen rapid growth, with recent developments resulting in the field now call deep learning. What differentiates machine learning and deep learning is the ability of the neural network architecture to automatically extract features from the data that are meaningful to the task at hand, this having been a manual task done by humans in the traditional machine learning sense. Deep learning has been popularized by what are known as convolutional neural networks (CNN), which have proven to be very effective in object identification in images and many other applications. Next is a general description of the basic neural network architecture which sits at the heart of both machine learning and deep learning architectures, followed by a brief introduction on how Bayesian methods can be applied to neural network training and model selection.

Figure 6.1: Neural Network Node
Fundamental neural network node description.

6.2.1 Bayesian Methods in Neural Networks

In a neural network, the node composes the fundamental processing unit of the network. A node receives an input and calculates an output as shown in figure 6.1.

Each input is multiplied by a weight and summed before it is passed through an activation function, $f_a(\cdot)$. Equation 6.1 describes this step where q represents a node, x_n are inputs to that node, y_q is the output, w_{qn} are the weight values at each input connection and b is a bias term.

$$y_q = f_a\left(\sum_n w_{qn} x_n + b\right) \quad (6.1)$$

In Equation 6.1, the activation function $f_a(\cdot)$ is an important factor

Figure 6.2: Log-Sigmoid, Tan-Sigmoid and Linear Activation Functions
Classic neural network node activation functions.

in the non-linear behavior of neural networks. Three classic functions that are typically used in feedforward neural networks are the log-sigmoid function $f_a(x) = \frac{1}{1+e^{-x}}$, the tan-sigmoid function $f_a(x) = \tanh(x)$ and the linear function $f_a(x) = x$. Figure 6.2 shows each graph for the functions, each presented with a zero bias term.

In a feedforward neural network neurons are grouped in layers. The network can have one or many layers, depending on the application and computational objective. The first layer receives the inputs and the last layer is the output of the network. All middle layers are called hidden layers. Once a network topology is established, in a machine learning or deep learning architecture, the node weights can be determined by what is known as supervised training. Supervised training of a neural network consists of applying an iterative error minimization process while providing the network with labeled training data. A typical gradient descent based algorithm used for this purpose is the Levenberg-Marquardt algorithm [38]. Another option is to apply Bayesian methodology to the learning of the network weights during training [39]. Equation 6.2 expresses Bayes' theorem for finding a posterior distribu-

tion over the network weights w , where learning the weights means changing our belief from a prior distribution, $p(w)$, to the posterior, $p(w \mid \Omega)$, as a consequence of seeing the data Ω . Once the posterior of the weights has been calculated, we can also consider producing a distribution over the network outputs, where the expected value of the posterior distribution of the network outputs, $p(y \mid x, \Omega)$, can be obtained by marginalizing over the network weights, as it's shown in equation 6.3. Here x and y represent new input output pairs of the trained network. Typically, in real life applications, this integral also needs to be approximated by a sampling method like MCMC due to the high dimensionality of the network weights.

$$p(w \mid \Omega) = \frac{p(\Omega \mid w)p(w)}{\int p(\Omega \mid w)p(w)dw} \quad (6.2)$$

$$p(y \mid x, \Omega) = \int p(y \mid x, w)p(w \mid \Omega)dw \quad (6.3)$$

6.2.2 Model Selection

Considering different neural network architectures, like networks with different number of hidden layers, inputs, nodes or outputs, a Bayesian framework can also be implemented for model selection without the need for a separate cross-validation data set. Applying Bayes' theorem, the posterior distribution can be computed over the different network architectures, or models H_i , allowing the model with the largest posterior to be selected. Equation 6.4 shows Bayes' theorem applied to model selection and equation 6.5 defines $p(\Omega \mid H_i)$, called the marginal likelihood or evidence for H_i . Following certain assumptions about the weights, the evidence for H_i integral calculation can also be

efficiently approximated.

$$p(H_i | \Omega) = \frac{p(\Omega | H_i)p(H_i)}{p(\Omega)} \quad (6.4)$$

$$p(\Omega | H_i) = \int p(\Omega | w, H_i)p(w | H_i)dw \quad (6.5)$$

A caveat in a neural network approach is that its performance is highly dependent on the training data available. For this research, this would imply images from a large number of camera models which can present a practical challenge. A possible solution to this would be to use simulated data during the training phase or restricting the performance of the network to a subset of camera models or conditions. Nonetheless, a neural network based machine learning approach could be investigated in the future as an alternative or complement to this research.

6.3 Summary

The flexibility of neural networks presents an attractive alternative for this research. Although flexible, this also opens many possibilities and alternative approaches that could each be a full research endeavor in itself and for that reason, although initially evaluated as a solution, it was moved to a potential future work opportunity to develop different solutions to the statistical inference problem.

7 CRF Estimation Experimental Results

7.1 Introduction

Throughout the evolution of this research, continuous evaluation of CRF estimation results was carried out at the different stages of progress as well as analysis of the impact in the implementation of an IDT in a ACES workflow. The work completed is briefly summarized next followed by descriptions of the results obtained at each stage of the research.

1. The first results obtained in this research are related to the expansion of Lin's algorithm to include multiple frames vs. a single frame CRF estimation. This work is published in the 2013 SMPTE Annual Technical Conference Proceedings [28].
2. Following the expansion of Lin's methodology to include multiple frames, the Bayesian MCMC methodology described in chapter 3 was implemented to estimate the posterior probability distribution of the CRF for edge color distribution image features. The work covering the estimation of the CRF through this Bayesian MCMC methodology utilizing edge color distribution image features showed estimation improvements over the previous implementation and is published in the August 2019 issue of the SMPTE Motion Imaging Journal [40].
3. The Bayesian MCMC methodology was again implemented to estimate the posterior probability distribution of the CRF with a novel likelihood probability loss function based on noise image features as described in

section 3.2. These results showed to be comparable to the Bayesian MCMC methodology utilizing edge color distributions.

4. Update of DoRF to include additional camera CRF data.
5. IDT computation color error analysis. This analysis showed the expected perceived color error in the computation of an IDT based on different levels of CRF linearization error.

7.2 Camera, Images and Setup

Image data was collected from an ARRI D21 camera, which is a good representation of typical cameras used in the motion picture industry. The data collected consisted of both actual response measurements utilizing an ISO CRF measurement chart, as well as camera footage including different motion sequences. Figures 7.1 and 7.2 show the measured CRF for the ARRI D21 motion picture camera under two different settings. Figure 7.1 shows the CRF setting for typical HD video gamma encoding, and figure 7.2 shows a more “filmic” encoding labeled as LogC by ARRI, which increases the camera’s dynamic range capabilities at capture. The CRF under this setting allows the camera to capture highlights beyond the maximum allowable for the HD video gamma encoding while still not reaching the maximum encoded code value. These are examples of different CRF options to the user of a motion imaging camera.

Figure 7.1: Arri D-21 Gamma Encoding Response Function
Theoretical versus measured CRF for an ARRI camera.

Figure 7.2: Arri D-21 LogC Encoding Response Function
Comparison of standard video CRF and extended dynamic range CRF for ARRI camera.

7.3 Multiple Frames MAP CRF Estimation

Estimation of the inverse CRF for the ARRI D21 professional motion picture camera was done for the HD video gamma encoding setting and this served as the main benchmark testing set for all other methodologies implemented in this research, although additional cameras and settings were also tested. This setting makes the camera output high definition motion images following the ITU Rec.709 video standard. Figure 7.3 shows a frame from the sequence of images captured with the ARRI D21 in this setting.

For this part of the research, additional frames from a sequence of images were included to improve the estimation of the inverse CRF with the goal of reducing the overall impact of noise in the edge color distributions. The set of color triplets described in equation 3.16, Ω_c , was formed from 20 x 20 pixel windows that contained edge colors between two uniform regions. The windows selected were non-overlapping and centered on edges identified by the Canny edge detection algorithm as described in section 3.4.2.

To determine the validity and quality of a color triplet, the edges se-

Figure 7.3: Sample ARRI D21 Frame
Sample image from sequence captured with ARRI camera.

lected were analyzed by the detection algorithm and the mean color and color variance were computed for the two regions on each side of the edge. If both regions had a color variance below a specified low threshold, 4% for most experiments, then the regions were considered uniform. Also, to guarantee that the distance measure in equation 3.16 was not dominated by image noise, the mean colors had to lie a specified distance from each other, at least 8% color difference in this implementation. The edge colors also had to lie within the range delimited by the two regions where the edge is. This requirement helped exclude edges with ringing that would affect the distance calculation.

A comparison was done with Lin's single frame implementation and the results are presented in figure 7.4, showing our estimation with 10 and 20 frames and Lin's single frame algorithm, versus the aim for the ARRI D21. The aim curves were created by fitting a five degree polynomial to the aim

Figure 7.4: Camera Response Function Estimation Plot for 10 and 20 Frames
CRF estimation comparison for multiple frames versus single frame edge color distributions.

	Red	Green	Blue
20 frames	.077	.047	.044
10 frames	.067	.059	.040
Single frame	.072	.051	.063

Table 7.1: RMS Error

RMS error between estimated CRF and theoretical CRF for single frame and multiple frame estimation methods.

data collected by photographing a CRF measurement chart.

The RMS error between the estimations and the aim curves for each color channel was computed and is presented in table 7.1, showing modest improvement by using multiple frames.

7.4 CRF Estimation Using Edge Color Distributions

In addition to expanding Lin’s algorithm to include multiple frames, a Bayesian MCMC algorithm was also implemented using edge color distributions to estimate the posterior distribution of the CRF following the methodology presented in section 3.4, but with an expansion to estimate the posterior probability function instead of the MAP approach. This approach used a single frame captured under the same conditions described in the previous sections.

The Metropolis-Hastings algorithm, as described in [31] and section 4.3.10, was implemented to sample the space of the inverse CRF g and approximate the posterior distribution. The CRF model presented in section 3.2.1 was used with 5 PCA components, making the sampling a 5-dimensional problem. In this implementation of the algorithm, the MCMC sampling algorithm was optimized by looking at the range covered by the PCA coefficients based on the DoRF to optimize the CRF PCA model sampling space. Also equal CRFs were assumed for the red, green and blue channels. Figure 7.5 shows the CRF (inverse of g) expected value, compared with the actual measured CRF for the camera and the estimations presented in the previous section. RMS errors between the MCMC CRF estimation and the aim curves were .028, .021, .021 for the red, green and blue channels respectively, showing improvement over the MAP method. Figure 7.6 shows the MCMC approximation of the PCA coefficients CRF model target functions that represent the 5 dimensional posterior probability function of g . Coefficients 1 and 2 show more complex and non-Gaussian shaped probability distributions that are better described by computing the expected value of the distribution instead of a single value like the mode. This is evidence for the improvement of these

Figure 7.5: MCMC CRF Estimation with Edge Color Distributions
CRF estimation comparison with posterior probability distribution sampling
and CRF expected value calculation, including edge color distribution image
features.

results over the MAP approach.

Figure 7.6: MCMC Approximation of Target Function for Coefficient 1 through 5

Histograms showing the approximate shape of the posterior probability distribution of the CRF coefficients, for a 5 dimensional PCA model.

7.5 CRF Estimation Using Noise Image Features

A Bayesian MCMC algorithm was also implemented to estimate the posterior distribution of the CRF following the methodology presented in section 4.2 using consecutive motion frames captured under the same conditions described in the previous sections. The Metropolis-Hastings algorithm, as described section 4.3.10, was again implemented to sample the space of the inverse CRF g and approximate the posterior distribution. The CRF model presented in section 3.2.1 was again used, but this time with 3 PCA components instead of 5 PCA components, reducing the dimensionality of the sampling space. Equal CRFs were also assumed for the red, green and blue channels as this is a common industry assumption. Figure 7.7 shows the CRF (inverse of g) expected value, compared with the actual measured CRF for the camera. RMS errors between the MCMC CRF estimation using the image noise features and the aim curves were .017, .020, .020 for the red, green and blue channels respectively, showing modest improvement over the MCMC CRF estimation using edge color distributions. Figure 7.8 shows the MCMC approximation of the PCA coefficients CRF model target functions that represent a 3 dimensional posterior probability function of g . These results show the viability of using noise image features in the methodology presented.

Figure 7.7: MCMC CRF Estimation with Noise Image Features
CRF estimation comparison with posterior probability distribution sampling
and CRF expected value calculation, including noise image features.

7.6 IDT Computation Color Error Analysis

As mentioned in previous sections, the optimization of the IDT matrix is a linear transformation that will only reach a minimum ΔE error value when the camera exposures have been properly linearized and the minimum error will depend on various factors, the most important being the spectral sensitivities of the capture camera and how close their design is to a linear combination of the RICD spectral sensitivities specified by ACES. This means that based on different sensor manufacturers, the error in the optimization will likely never be zero except in the case where the camera spectral sensitivities satisfy the condition mentioned above. In this example analysis, the minimum ΔE obtained when the data was perfectly linearized was 1.1. Non-linear data will

Figure 7.8: MCMC Approximation of Target Function - 3 Coefficients
 Histograms showing the approximate shape of the posterior probability distribution of the CRF coefficients, for a 3 dimensional PCA model.

Gamma	deltaE	RMS Error
.95	2.2	.014
.9	3.5	.029
.85	5.2	.044
.8	6.8	.061

Table 7.2: IDT deltaE Calculation Error
Resulting color error in the computation or implementation of an IDT as the image data deviates from being radiometrically linear.

only increase the minimum error achievable by the matrix transformation, and it is important to again highlight that the minimum deltaE achievable, even when the image data is perfectly linearized, depends on the spectral sensitivity of the camera sensor and how similar or different it is to the RICD spectral sensitivity, as discussed in chapter 5. To illustrate this point, table 7.2 presents the impact of not estimating the CRF accurately, deviating from a perfect linearization of the image data. This was obtained by modifying linear data and encoding it non-linearly before the calculation of the IDT following the procedure in 5. The non-linear deviations were implemented as gamma functions of .95, .9, .85 and .8, where a gamma of 1 would represent original linear data or perfect linearized data after implementing an inverse CRF.

These results demonstrate how a slight deviation from linear correction causes an increase in the color difference error during the optimization of the IDT coefficients and how important it is to accurately estimate the CRF so that correct linearization of the image data leads to the least possible color difference error in the calculation or the implementation of the IDT matrix in post-production. The RMS error for the associated CRF estimation deviation is also included in the table as a comparison reference to the RMS errors presented in the previous two sections. Larger color difference errors in the IDT

computation will directly transfer to more time spent in post-production doing manual correction and adjustments to match images. Achieving a deltaE color error of around 2 or below, will produce satisfactory results in the ACES encoding of the images as discussed in section 5.2.1. A deltaE of 1-2 is considered a JND and is an appropriate aim threshold to achieve so images are considered encoded correctly into ACES via the IDT.

Looking at results in the previous sections, RMS errors were reduced from a maximum of around .07 with Lin's methodology, to RMS errors of around .02 implementing the Bayesian framework presented in this dissertation. This RMS error reduction translated to a color error reduction from a deltaE of approximately 7 to around 4 with the estimation of the posterior probability distribution using edge color distributions for the cameras considered, and subsequently to a deltaE of 2 with the noise image features approach. This overall reduction of 5 deltaE units represented a dramatic improvement in the performance of the IDT after proper linearization of the camera images to be transformed by the IDT. This error reduction is an example of the improvements produced by computing an expected value of the posterior probability distribution instead of the mode of the distribution. Coefficient 1 and coefficient 2 in figure 7.6 are examples of more complex parameter posterior probability distributions where the mode is not a sufficient descriptor and an expected value of the estimated posterior probability distribution produces more accurate results. This color error reduction also represents a potential reduction in the time spent in post-production working on technical grading of images from different camera sources, leaving more time to be spent in creative grading of the images. Although part of the post-production time savings de-

depends on the performance of the IDT, the exact total time also depends on the ability of the post-production colorist performing technical grade operations. With the color error reduction obtained, a conservative minimum 50% time savings could be achieved in technical grading, based on various conversations with industry professionals about the potential impact of this work. As part of the future work, measurements of the actual time savings are planned during testing of the algorithm in post-production.

In addition, the inclusion of a seventh parameter in the IDT calculation optimization, as described in section 5.2.1, showed that the IDT matrix optimization can simultaneously compute the exponent, β_7 , for a power function model of the CRF. As a reminder, this is an optional approach for the calculation of an IDT, but not possible during the implementation of an IDT. For implementation of an already calculated IDT, the linearization of the camera images would be performed following the methodology in this dissertation.

7.7 Methodology Limitations Summary

Results obtained confirmed that a Bayesian approach estimating the posterior probability distribution through an MCMC algorithm, and the subsequent calculation of an expected value from the posterior distribution, is a viable method for estimating the CRF from images. This methodology showed improvement over ML and MAP while maintaining some computational practicality which can be further improved with future algorithm optimizations. Nonetheless, there are certain limitations worth noting regarding the various methods presented in this work.

- For methods using edge color distributions, color edges satisfying the conditions in section 3.4.3 are necessary to avoid potential ambiguity in the estimation of the CRF. Neutral color edges will produce estimation ambiguities and should not be analyzed with this technique.
- For methods using edge color distributions, high levels of image noise caused by any kind of underexposures to the sensor, will result in non-optimum edge color distributions deviating from the conditions presented in section 3.4.3 and affecting the distance calculation in equation 3.16 and subsequent estimation of the CRF.
- For methods using noise image features, high spatial image processing will affect the distance calculation in equation 4.10. High noise reduction will affect the ability to measure the photon/read noise at different non-textured areas of the image and consequently affect the accurate estimation of the CRF.
- For methods using noise image features, very low dynamic range images, e.g. very dark or very bright images, will present a challenge in calculating an accurate estimate of data like shown in figure 4.5 for the images in question. The linearity of the data presented on figure 4.5 is best obtained and visualized with code values across the full dynamic range of the camera system. Reducing the number of measurements to points grouped together in any single area of the camera dynamic range, will affect the accurate calculation of the correlation coefficient and distance calculation in equation 4.10.
- Finally, for methods using noise image features, highly textured image

content will present a challenge in the identification of enough non-textured image areas to compute the total noise in equation 4.9.

Although these limitations must be taken into account, motion picture industry images provide the characteristics necessary for both approaches, using edge color distributions or noise image features, to be successful the majority of the times, as images tend to be composed of a wide dynamic range, in most cases have been minimally post-processed. Also, only in very particular situations, images are not captured in color or contain a majority of neutral tones. Because of this, the methods presented here are a general viable solution to the problem of estimating a CRF.

7.8 Hardware Platform Details

All instances of the algorithm were implemented using MATLAB as the programming platform. The latest implementation was done in MATLAB R2018a. The hardware utilized for most experiments consisted of a MacBook Pro with a 2.8 GHz Intel Core i7 CPU and 16 GB of RAM and including a AMD Radeon R9 M370X with 2GB of RAM. For tests on a faster platform, experiments were done on a Dell Precision 5820 with a 3.7 GHz Intel Xeon W2145 CPU and 32 GB of RAM, where full algorithm implementation times averaged 5-10 minutes for both the edge color distribution approach and for the noise features approach. It's believed that these times can be reduced with further algorithm optimization.

8 Summary

8.1 Conclusions

A Bayesian approach as described in this work, is a viable solution to estimating an electronic camera CRF from captured digital images. Again, this is an important computation when the CRF is not known from metadata or from a direct measurement. In many cases, only captured digital images, and not the camera or accompanying metadata, are available when incorporating said images into a motion picture post-production workflow. A CRF will typically represent a non-linear relationship between the irradiance from the scene and the measured pixel code value for benefits including encoding efficiency and perceptual accuracy as described in this research. Understanding this non-linearity is also important to ensure the proper calculation or implementation of color space transformations like the IDT. The calculation of the IDT assumes a linear relationship between the exposures captured by the camera system and the measured pixel code values and any non-linearity still present in the image pixel code values at the time of calculating or implementing an IDT, will result in a color error that will prevent a post-production workflow, like ACES, from being efficient in its implementation.

Although radiometric calibration methods exist to estimate a CRF, the majority of these methods require image capture techniques that restrict exposure or position of the camera at the moment of capture. These restrictions are impractical in any motion picture production workflow. In addition, many methods are limited in the CRF mathematical model implemented. To avoid these limitations, this work focused on expanding the methodology presented

by Lin et al. based on single image radiometric calibration. This presented a viable practical method where captured images could be analyzed to estimate the CRF of the camera in question, and proper image linearization could be performed before implementing any color transform requiring linear pixel code values. This presented a potentially viable methodology where the radiometric calibration implementation would only depend on one image or a series of images, without any strict restrictions on how those images were captured. This flexibility was key when it came to developing a methodology that would be practical for implementation in motion picture post-production workflows.

Lin's methodology centered around the use of edge color distributions to serve as the image feature that helps identify the linearity or non-linearity of pixel code values. This image feature was useful in any color image where edge image processing would not interfere with the likelihood function distance measure used in their MAP Bayesian inference approach. This meant that any image highly image processed, like with large edge enhancements, or with high levels of image noise, could present a problem for the CRF estimation with that approach. To overcome this, several possible improvements to their methodology were investigated. The first extension to the work focused on expanding the number of edge color distributions included in the data set used to compute the distance measure that was to be minimized during the MAP Bayesian inference. This expansion was carried out by including various ranges of frames in the optimization and it showed modest improvement over the existing method by providing more data that satisfied the statistical requirements for the edge color distributions considered fit for use in the optimization.

The next extension of the work focused on improving the Bayesian framework by going from a single value estimation, as is the case with MAP, to a full posterior probability distribution estimation of the parameters in the CRF model. This approach took into account the whole posterior distribution and allowed then the computation of an expected value based on the full posterior probability. The benefits of this approach were clearly seen in situations where the parameter probability distribution was more complex than a Gaussian shaped distribution, like coefficient 2 in figure 7.6. This approach produced noticeable improvements that were measured by computing the RMS errors between the estimated CRF and the actual measured CRF for a series of images. This RMS error was further correlated to perceived color differences that would be noticed by industry professionals manipulating said images. By reducing the RMS error, the expected perceived color error was also reduced, aiming at avoiding time spent in technical grading vs. creative grading in a motion picture post-production workflow. To expand the methodology as mentioned above, and estimate a full posterior probability distribution, an MCMC sampling method was used to find a compromise between computational accuracy and computational practicality. Calculating the exact posterior probability distribution can be computationally expensive or simply an intractable problem due to high parameter dimensionality. By implementing an MCMC sampling algorithm, an estimated posterior probability distribution could be obtained by analyzing the stability of the resulting statistics. This in turn provided a potentially viable method that could then be incorporated into post-production workflows.

Furthermore, by analyzing potential image artifacts affecting the edge

color distributions used in the likelihood function distance measure, different noise components, and their signature on linear vs. non-linear pixel code values, proved to be a novel image feature to be used as a key metric in the likelihood probability distribution. The noise image feature, based on photon and read noise that exist in all pixel code values, presented a second option to the posterior probability distribution estimation method and produced comparable results to the color edge distributions technique. This is a novel image feature and a key contribution to the radiometric calibration field which includes other applications like remote sensing imaging. Although some limitations also impact this approach, these only occur on very specific cases which tend to not be typical, or the majority of the images expected in motion picture workflows. In most cases, images that don't include the limitations will be accessible for analysis even when images with limitations also exist.

The Bayesian framework presented in this work can also be simplified by implementing a CRF model with fewer parameters, or a simpler model altogether, while still taking advantage of the edge color distributions or noise image features used for distance measures in the likelihood distribution function. This provides further research opportunities looking into the optimum implementation of this algorithm in different post-production workflows.

In conclusion, a robust method for radiometric calibration in motion picture workflow was developed and tested along with expansions to other existing methods. The methods rely solely on images captured during normal production workflow scenarios and utilizes image features that are easily measurable, to obtain an accurate approximation of the CRF. The Bayesian framework proved to be an appropriate methodology as initially suggested and

also keeps the door open to further experimentation and analysis of new image features that could potentially expand this research into the analysis of other image capture processes that could be inferred from image data.

8.2 Future Work

Finally, it's important to realize that the limitations each approach presents serve as motivation to keep improving the CRF estimation methodologies presented in this work. New capture technologies will open opportunities to test and expand the methodologies presented here with those new capture systems and images. Also, optimization of the algorithm to reduce implementation times, will increase the practicality and viability to become an integral part of motion picture post-production workflows like ACES. Research should continue into identifying additional image features that offer statistical information valuable not only for radiometric calibration, but for the estimation of other capture parameters that can help in the post-production of motion images in addition to simpler CRF models that can be implemented to optimize the performance of the algorithm.

Focusing on the practicality of the algorithm for use in post-production, three areas of further research can be prioritized. First, the combination of both image features into the Bayesian framework presented has the potential of improving results by overcoming some of the limitations suffered by each method on its own. Measuring noise and dynamic range of the image in question, can provide information to decide which image feature has the potential to produce better results in a real time implementation. Low dynamic range might mean that the noise image feature might be too localized and not provide sufficient points to obtain a reliable correlation measure of linearity, indicating the use of the edge color distributions in that case. High noise across the whole image might indicate that edge color distributions might be affected in their distance measure for linearity, and that no reliable non-textured areas can be

selected for the noise image feature approach, resulting in a rejection of that image data for reliable analysis.

Second, as new camera models get introduced, their CRFs can be included in the DoRF. Furthermore, the DoRF can be customized to just include a subset of the CRFs which can all be for example from a specific camera manufacturer. This has the potential of reducing the parameter space for that specific camera manufacturer. Several assumptions are made for this to be possible, these being that the camera manufacturer and designers keep consistency in their sensors' spectral sensitivity as well as the range of CRFs implemented. If these assumptions are correct, the DoRF can then be customized by camera manufacturer and a principal component model and Gaussian mixture model can be computed for each customized database, potentially reducing the parameter space. This would in turn potentially allow for a complete description of the posterior probability distribution, or a faster and more efficient MCMC implementation.

Third, at the same time that the DoRF can be customized, analysis of simpler CRF models could give an insight into faster alternatives for CRF estimation of certain camera image data. Although many camera manufacturers deviate from simple gamma or logarithmic CRF implementations like mentioned before, other camera manufacturers do implement these type of functions. Furthermore, the same camera manufacturer can implement a different CRF depending on the output format settings, allowing for a simpler CRF model under certain settings, and more complex CRFs under other settings. This could be studied in combination with image features that provide information about the output format. One example of these type of features

is the image size, where an image size of 1280x720 or 1920x1080 might indicate a High Definition standard output, where a gamma CRF model could be sufficient for linearization of the data.

By focusing the immediate future work in these three areas, improvements could be obtained in the implementation speed of the algorithm, further improving its practicality in motion picture post-production workflows and allowing the collection of feedback from colorists to quantify actual time savings in technical grading operations.

Last, at a more early stage but following the alternative problem solution direction discussed in chapter 6, a neural network based approach could be studied and combined with a Bayesian framework for network training and model selection. Neural networks open the door to numerous approaches including possible novel ways to learn image features affected by the non-linearity introduced by a CRF. Neural networks require a lot of data for training, and motion images are an attractive source of these large amounts of data. This approach would entail a rethinking of how the problem is solved and might focus on network training of image features that signal a radiometric non-linearity as a first step, and subsequently focus on network prediction of a simpler CRF model based on image pixel data and learned image features.

Index

1931 CIE tristimulus values, 85

A

Academy Color Encoding System, 3,
29

B

Bayes' theorem, 41, 49, 70

C

Camera Response Function, 2, 13

Canny edge detection, 44, 69

chromatic adaptation, 84

CINEON, 3, 27

D

Database of Response Functions, 38

Digital Intermediate, 26

E

Empirical Model of Responses, 38

F

fixed pattern noise, 59

G

gamma correction, 19

Gaussian Mixture Model, 51

I

Input Device Transform, 4

irradiance, 15

ISO speed, 17

L

lightness, 20

M

Markov Chain Monte Carlo, 75

Maximum A Posteriori, 49

Maximum Likelihood, 49

Metropolis-Hastings, 77

N

neural network, 92

P

Pearson correlation coefficient, 67

photon noise, 59

pixel response non-uniformity noise, 59

product rule of probability, 40

Q

quantization error, 59

R

radiance, 15

radiometric calibration, 33

read noise, 59

S

sum rule of probability, 40

T

thermal noise, 59

W

Weber's Law, 20

A APPENDIX

Example Published Camera Response Functions

Figure A.1: ARRI ISO Dependent CRFs

Figure A.2: Sony CRFs

Figure A.3: Canon CRFs

B APPENDIX

Camera Response Function Post-Production Workflow Diagram

Figure B.1: CRF Post-Production Workflow Diagram

This diagram shows a visual approximation of how the inverse CRF impacts the encoding of images into the ACES color space

C APPENDIX

ACES Workflow Description Diagram

References

- [1] Academy Color Encoding Specification (ACES). *SMPTE ST 2065-1:2012*, pages 1–23, April 2012.
- [2] M.D. Grossberg and S.K. Nayar. Determining The Camera Response from Images: What Is Knowable? *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 25(11):1455 – 1467, nov. 2003.
- [3] S. Lin, Jinwei Gu, S. Yamazaki, and Heung-Yeung Shum. Radiometric Calibration from a Single Image. In *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on*, volume 2, pages II–938 – II–945 Vol.2, june-2 july 2004.
- [4] ISO. *ISO 12232 Photography - Electronic still-picture cameras - Determination of ISO speed*, 2019 edition.
- [5] Charles Poynton. *Digital video and HDTV : Algorithms and Interfaces*. Morgan Kaufmann Publishers, 1st edition, 2003.
- [6] Michael H. Brill and Robert C. Carter. Does Lightness Obey a Log Or a Power Law? Or Is That The Right Question? *Color Research and Application*, 39(1):99–101, 2014.
- [7] Charles Poynton, Jeroen Stessen, and Rutger Nijland. Deploying Wide Color Gamut and High Dynamic Range in HD and UHD. *Motion Imaging Journal, SMPTE*, 124(3):37–49, April 2015.

- [8] Jan van Rooy and Rob Voet. Twelve Bit Acquisition, the Next Step in Digital Broadcast Cameras. In *Technical Conference Technical Papers Program, 138th SMPTE*, pages 1–14, Oct 1996.
- [9] Richard Patterson. Evaluating Density Metrics for Scanning Motion Picture Negatives. *Motion Imaging Journal, SMPTE*, 117(4):31–38, May 2008.
- [10] Steven J. Powell, Richard C. Sehlin, Roland J. Zavada, and Mitchell J. Bogdanowicz. Interface of Motion Picture Films and Video. *SMPTE Journal*, 95(6):614–623, June 1986.
- [11] David Snider, Glenn Kennel, Ken Curry, and Michael McCrackan. Digital Moving-Picture Exchange: File Format and Calibration. *SMPTE Journal*, 102(8):712–714, Aug 1993.
- [12] ISO. *ISO 14524 - Photography - Electronic still-picture cameras - Methods for measuring opto-electronic conversion functions (OECFs)*, 2009 edition.
- [13] T. Mitsunaga and S.K. Nayar. Radiometric Self Calibration. In *Computer Vision and Pattern Recognition, 1999. IEEE Computer Society Conference on.*, volume 1, pages 2 vol. (xxiii+637+663), 1999.
- [14] Paul E. Debevec and Jitendra Malik. Recovering High Dynamic Range Radiance Maps from Photographs. In *Proceedings of the 24th annual conference on Computer graphics and interactive techniques, SIGGRAPH '97*, pages 369–378, New York, NY, USA, 1997. ACM Press/Addison-Wesley Publishing Co.

- [15] C. Manders, C. Aimone, and S. Mann. Camera Response Function Recovery from Different Illuminations of Identical Subject Matter. In *Image Processing, 2004. ICIP '04. 2004 International Conference on*, volume 5, pages 2965 – 2968 Vol. 5, oct. 2004.
- [16] Seon Joo Kim, J.-M. Frahm, and M. Pollefeys. Joint Feature Tracking and Radiometric Calibration from Auto-Exposure Video. In *Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on*, pages 1 –8, oct. 2007.
- [17] H. Farid. Blind Inverse Gamma Correction. *Image Processing, IEEE Transactions on*, 10(10):1428 –1433, oct 2001.
- [18] M. Diaz and P. Sturm. Radiometric Calibration Using Photo Collections. In *Computational Photography (ICCP), 2011 IEEE International Conference on*, pages 1 –8, april 2011.
- [19] K. Shafique and M. Shah. Estimation of the Radiometric Response Function of a Color Camera from Differently Illuminated Images. In *Image Processing, 2004. ICIP '04. 2004 International Conference on*, volume 4, pages 2339–2342 Vol. 4, Oct 2004.
- [20] Sujit Kuthirummal, Aseem Agarwala, Dan B Goldman, and Shree K Nayar. Priors for Large Photo Collections and What They Reveal About Cameras. In *Computer Vision–ECCV 2008*, pages 74–87. Springer, 2008.
- [21] A. Badki, N. Khademi Kalantari, and P. Sen. Robust radiometric calibration for dynamic scenes in the wild. In *2015 IEEE International*

- Conference on Computational Photography (ICCP)*, pages 1–10, April 2015.
- [22] H. Lin, Seon Joo Kim, S. Süsstrunk, and M. S. Brown. Revisiting radiometric calibration for color computer vision. In *2011 International Conference on Computer Vision*, pages 129–136, Nov 2011.
- [23] J. Lee, Y. Matsushita, B. Shi, I. S. Kweon, and K. Ikeuchi. Radiometric calibration by rank minimization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(1):144–156, Jan 2013.
- [24] M.D. Grossberg and S.K. Nayar. Modeling the Space of Camera Response Functions. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 26(10):1272 –1282, oct. 2004.
- [25] M Jordan and J Kleinberg. Information science and statistics.
- [26] J. Canny. A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-8(6):679–698, 1986.
- [27] Philip Resnik and Eric Hardisty. Gibbs sampling for the uninitiated. Technical report, MARYLAND UNIV COLLEGE PARK INST FOR ADVANCED COMPUTER STUDIES, 2010.
- [28] Ricardo R. Figueroa and Jinwei Gu. Camera Radiometric Calibration from Motion Images. In *Annual Technical Conference Exhibition, SMPTE 2013*, pages 1–10, Oct 2013.

- [29] Emil Martinec. Noise, dynamic range and bit depth in digital slrs, May 2008.
- [30] M. Stojmenovic, A. Nayak, and J. Zunic. Measuring linearity of a finite set of points. In *Cybernetics and Intelligent Systems, 2006 IEEE Conference on*, pages 1 –6, june 2006.
- [31] Christophe Andrieu, Nando De Freitas, Arnaud Doucet, and Michael I Jordan. An introduction to mcmc for machine learning. *Machine learning*, 50(1-2):5–43, 2003.
- [32] John Kruschke. *Doing Bayesian data analysis: A tutorial with R, JAGS, and Stan*. Academic Press, 2014.
- [33] AMPAS Science and Technology Council. AMPAS P-2013-001, Recommended Procedures for the Creation and Use of Digital Camera System Input Device Transforms (IDTs). Technical report, Academy of Motion Picture Arts and Sciences, 2015.
- [34] J. Pytlarz, E. Pieri, and R. Atkins. Objectively evaluating high dynamic range and wide color gamut color accuracy. *SMPTE Motion Imaging Journal*, 126(2):27–32, March 2017.
- [35] Edward J. Giorgianni and Thomas E. Madden. *Digital Color Management: Encoding Solutions*. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 1998.
- [36] I.A Basheer and M Hajmeer. Artificial Neural Networks: Fundamentals, Computing, Design, and Application. *Journal of Microbiological Methods*, 43(1):3 – 31, 2000. Neural Computing in Micrbiology.

- [37] David Silver, Julian Schrittwieser, Karen Simonyan, Ioannis Antonoglou, Aja Huang, Arthur Guez, Thomas Hubert, Lucas Baker, Matthew Lai, Adrian Bolton, Yutian Chen, Timothy Lillicrap, Fan Hui, Laurent Sifre, George van den Driessche, Thore Graepel, and Demis Hassabis. Mastering the game of go without human knowledge. *Nature*, 550:354 EP –, 10 2017.
- [38] Ricardo R. Figueroa. Neural Network Subpixel Classification and Estimation of Multispectral Images. Master’s thesis, University of Puerto Rico at Mayaguez, 1998.
- [39] Christopher Bishop. *Pattern recognition and machine learning*. Springer, 1st ed. 2006. corr. 2nd printing 2011 edition, 2013.
- [40] R. R. Figueroa and R. Li. Bayesian methods for radiometric calibration in motion picture encoding workflows. *SMPTE Motion Imaging Journal*, 128(7):25–30, Aug 2019.