3D Multispectral Colorimetry

Anjali Thakrar
Ren Ng, Ed.
Austin Roorda, Ed.

Electrical Engineering and Computer Sciences
University of California, Berkeley

Technical Report No. UCB/EECS-2024-133
http://www2.eecs.berkeley.edu/Pubs/TechRpts/2024/EECS-2024-133.html

May 17, 2024

3D Multispectral Colorimetry

by

Anjali Thakrar

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Master of Science

in

Electrical Engineering and Computer Sciences

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Ren Ng, Chair
Professor Austin Roorda

Spring 2024

3D Multispectral Colorimetry

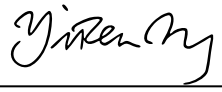
by Anjali Thakrar

Research Project

Submitted to the Department of Electrical Engineering and Computer Sciences, University of California at Berkeley, in partial satisfaction of the requirements for the degree of **Master of Science, Plan II**.

Approval for the Report and Comprehensive Examination:

Committee:



Professor Ren Ng
Research Advisor

May 17, 2024

(Date)

* * * * * * *

Professor Austin Roorda
Second Reader

May 17, 2024

(Date)

3D Multispectral Colorimetry

Abstract

3D Multispectral Colorimetry

by

Anjali Thakrar

Master of Science in Electrical Engineering and Computer Sciences

University of California, Berkeley

Professor Ren Ng, Chair

We introduce Multispectral Colorimetry, a novel technique for image color correction that uses 3D reconstruction techniques, object color theory, and photos captured by multiple cameras to generate a more accurate scene representation. Images captured by consumer cameras typically look similar to, but don't quite match the colors seen in the real world. In fact, there are many objects that have distinct spectral reflectances in the physical world but appear to have the same color when captured through a camera. This discrepancy is due to fundamental differences between the camera capture pipeline and that of the human eye: cameras only capture a subset of the colors humans can see, and any subsequent image processing introduces further error by approximating its output colors. In this work, we extend and improve current camera processing mechanisms to correct the color representation of any image, making its colors appear more similar to those found in the real world. We use images captured of the same scene by multiple cameras with slightly different response functions to extrapolate multispectral information within the visible spectrum and find a more accurate color mapping. To achieve this, we extend Gaussian Splatting to reconstruct a multispectral 3D scene using RAW captures from a set of cameras. This allows us to flexibly capture input images, integrate spectral samples from all cameras in 3D, and then generate multispectral images from arbitrary new views. We use the generated multispectral images to map colors between camera captures with pixel-perfect accuracy. We then use information about the spectra that can be functionally captured by each of the cameras and the average human eye in order to construct a mapping between each color value in the image and a set of candidate, "real" colors that it may represent in the world. As we introduce more cameras into this pipeline, the set of candidate colors becomes more constrained, and thus more precise. We produce a dataset of spectral response curves and color-corrected images for machine learning researchers, and an underlying processing pipeline that can be used by photographers.

To my family

Contents

Contents **ii**

List of Figures **iii**

1 Introduction **1**
1.1 Human Color Model 1
1.2 Camera Color Model 3
1.3 Object Color Solids 4
1.4 Multispectral Capture 5

2 Related Work **10**
2.1 Existing Color Correction Methods 10
2.2 Metamer Mismatch Volume 10
2.3 Multispectral Camera Capture 11

3 Volumetric Multispectral Inference **12**
3.1 Multispectral Gaussian Splatting using RAW images 12
3.2 Camera Processing Pipeline 13

4 Applying Color Theory to Reduce Metameric Error **18**
4.1 Metamer Mismatch Volume 18
4.2 Luther Condition 20

5 Discussion **22**
5.1 Analysis of Results 22
5.2 Future work 23

6 Conclusion **25**

Bibliography **26**

List of Figures

1.1 Cone fundamentals for the L, M, and S cone types, provided by Stockman and Sharpe [1] [2]. These cone fundamentals are an estimate of each cone's sensitivity to capture light at each wavelength of the visible spectrum. The sensitivity is defined as the probability that a photon is absorbed by the cone cell. These sensitivity values are normalized to be in the range [0,1]. 2
1.2 The spectral locus plotted in LMS space and CIE 1931 XYZ space. Left top: the spectral locus plotted in LMS space. Left bottom: the spectral locus plotted in CIE 1931 XYZ space. Right: The spectral locus scaled by X + Y and projected onto the Z axis, constructing a chromaticity diagram. 3
1.3 Diagram of Bayer Pattern. Top row: Diagram of mosaiced bayer sensor pattern and demosaiced RGB image. Bottom row: Cartoon of a RAW image with bayer tiling and the same image after converting it into a 3-channel image and demosaicing. 7
1.4 Left: 3D SRCs with LMS and a camera's curves overlayed in XYZ space. Because the SRCs differ, they cover different regions of the color gamut 8
1.5 The object color solid of a trichromatic human observer, using cone fundamentals measured by Stockman and Sharpe [1] [2]. 9

3.1 3D processing pipeline to convert RAW captures into an interactive, multispectral scene 14
3.2 The same image rendered of the same scene from three different cameras: Sony a7s, Canon EOS 5Dsr, and iPhone 13 Pro. This is generated by taking a multispectral capture in the 3D environment and separating its channels into its representative cameras' images. 15
3.3 PCA visualization of a multispectral reconstruction to show what channels contribute the most to each region of the scene 16
3.4 Visualization of the camera image processing pipeline for a given image 17

4.1 Visualization of metamer mismatch between the LMS and single-camera space . 19
4.2 Demonstration of how pixels of an image are mapped to a metamer mismatch. Each pixel color will generate a new metamer mismatch region that exists within the LMS object color solid 20

5.1 Two test cases to confirm the implementation works as expected. Left: source SRC and target SRC are both LMS, and for a given color, the mismatch volume is a point. This means it is a perfect 1:1 mapping between the solids, which is expected. Right: Metamer mismatch volume is larger towards the center and smaller towards the black and white points. 23
5.2 A set of charts demonstrating how the Vora Value increases as new spectral response curves are added to the Q matrix. The initial set of SRCs are from the Canon EOS 5DsR camera, and have a baseline Vora Value of 0.921. In the top row, each channel of the Canon EOS R camera is added to the set of SRCs. The bottom row adds the SRCs of a third camera, the Nikon D7100. 24

Acknowledgments

I would like to thank my advisor, Prof. Ren Ng, for his guidance, support, and encouragement over the past few years. Prof. Ren Ng's excitement for research is infectious, and conversations with him have taught me to always stay curious, think outside of the box, and embrace project pivots. I have grown an unbelievable amount as a researcher and a leader in the time that I have worked with Prof. Ng, and I am grateful to have had him as a mentor.

Much of the work described in this thesis is a joint effort with Riley Peterlinz. This research would not be possible without his efforts and insights towards developing technically-interesting and applicable research. I'm grateful to have learned from him and alongside him, and am lucky to call Riley my close friend.

I'd like to extend a special thanks to Atsu Kotani and Jessica Lee for the many discussions about color science, vision science, and photography that provided much of the background necessary for this project. Their guidance has been invaluable as the project developed and changed. I'd also like to thank James Fong, Alexander Belsten, Hannah Doyle, Congli Wang, Justin Kerr, and Brent Yi, for their input and suggestions about the project.

Finally, I would like to thank my family for their unwavering support and encouragement. Their support has given me the confidence to do anything I set my mind to, and has been a driving force behind my academic career thus far.

Chapter 1

Introduction

Cameras are built to capture and represent the world in a way that is as similar to the human experience as possible. Images are considered **colorimetric** if they represent the spectral appearance of the world in the same way it is captured by the eye. Human color perception is defined by the wavelengths of light that our eyes are biologically able to accept and process, which is its color **gamut**. Due to physical design constraints, each camera captures a slightly different gamut than the human eye does, leading to different color representations. In order to build accurate systems that improve the color representation of images, it is vital to precisely model color differences between capture devices. This section goes into detail about how color perception is modeled in both the human eye and cameras in terms of optical design and color theory.

1.1 Human Color Model

The human eye captures light through a set of photoreceptive cone cells that each capture different subsets of the visible light spectrum. Most humans are trichromats, meaning that they have three distinct classes of cones cells - long (L), medium (M), and short (S), which accept long, medium, and short wavelengths of light, respectively. These cones capture light in accordance with their cone fundamentals, which are spectral responses describing the cells' peak sensitivity to certain wavelengths. These fundamentals are shown in Figure 1.1.

The combination of the light that is captured by these cells make up the color that humans perceive in the world. When plotted in 3D space, these fundamentals construct a volume referred to as the **spectral locus**. This locus shows human cone cells' response to monochromatic light ranging from 380nm to 700nm and defines the extent of possible colors given the LMS response curves. When transformed into CIE 1931 color space, scaled, and projected along the Z axis, this locus becomes the chromaticity diagram, as shown in Figure 1.2.

When viewing a scene, we compute the LMS cone activity as follows:

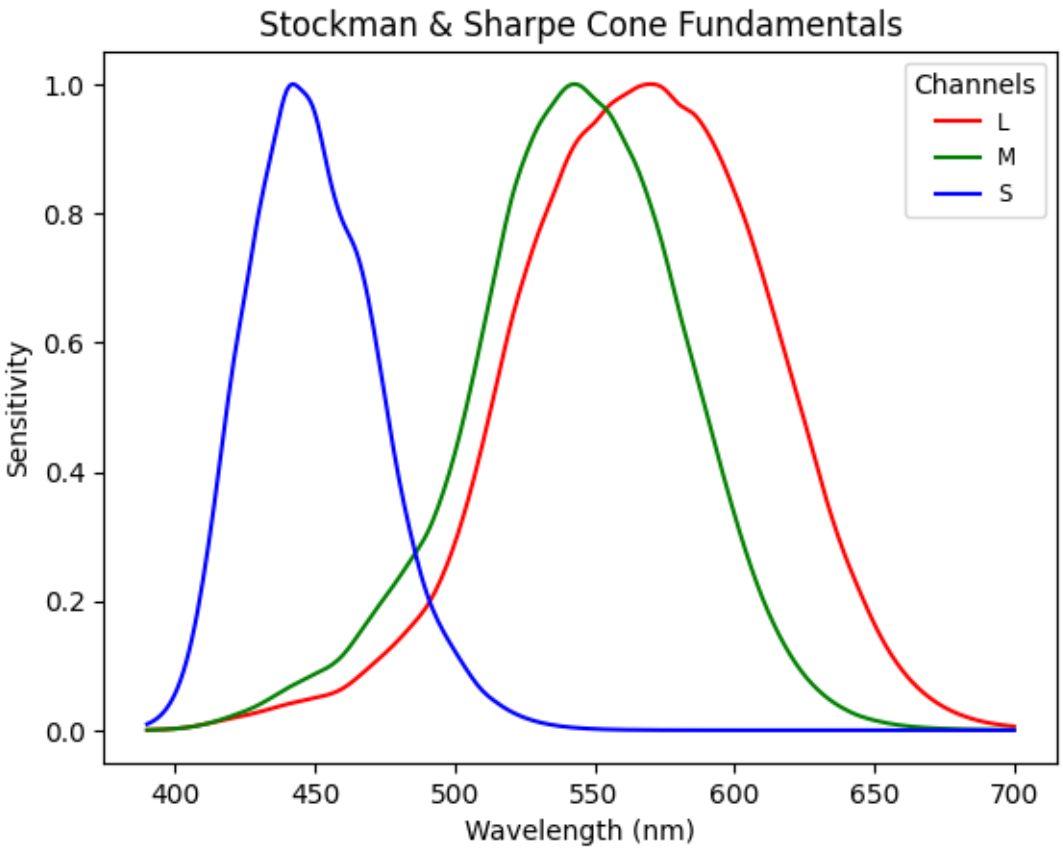


Figure 1.1: Cone fundamentals for the L, M, and S cone types, provided by Stockman and Sharpe [1] [2]. These cone fundamentals are an estimate of each cone's sensitivity to capture light at each wavelength of the visible spectrum. The sensitivity is defined as the probability that a photon is absorbed by the cone cell. These sensitivity values are normalized to be in the range [0,1].

$$L = \int F_L(\lambda)\Phi(\lambda)d\lambda$$

$$M = \int F_M(\lambda)\Phi(\lambda)d\lambda$$

$$S = \int F_S(\lambda)\Phi(\lambda)d\lambda$$

where $F_L(\lambda)$, $F_M(\lambda)$, and $F_S(\lambda)$ are the LMS cone fundamentals and $\Phi(\lambda)$ is the spectral power distribution (SPDs) of the scene.

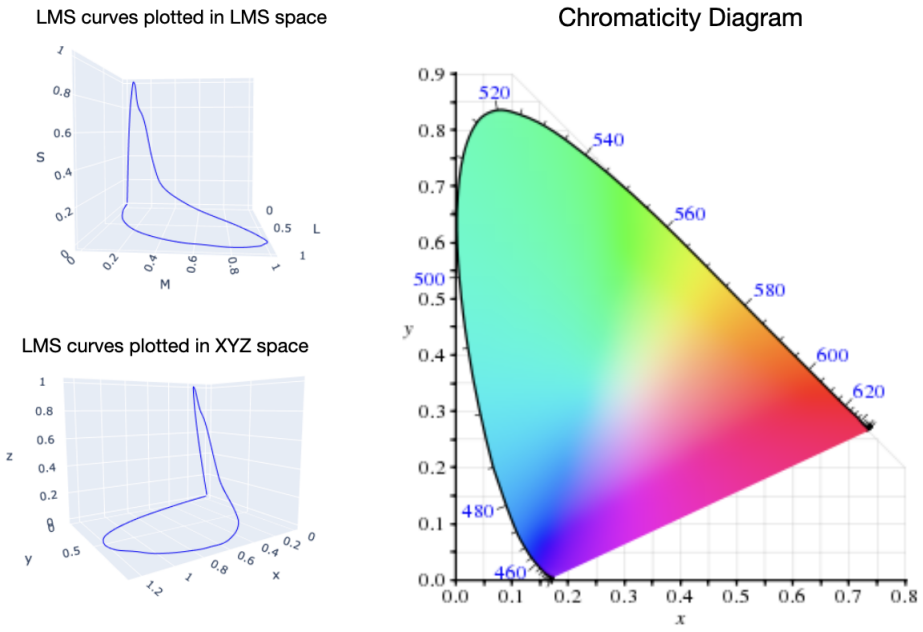


Figure 1.2: The spectral locus plotted in LMS space and CIE 1931 XYZ space. Left top: the spectral locus plotted in LMS space. Left bottom: the spectral locus plotted in CIE 1931 XYZ space. Right: The spectral locus scaled by X + Y and projected onto the Z axis, constructing a chromaticity diagram.

The light that L, M, and S cone cells capture are defined by their spectral power distributions (SPDs), which represents the amount of light accepted at each wavelength in Figure 1.1. In this work, we use the cone fundamentals measured by Stockman and Sharpe [1] [2].

1.2 Camera Color Model

Cameras are built to mimic the human visual system, and thus have similar capture mechanisms. The camera captures light on its image sensor, similar to the human retina. This sensor is typically constructed as a filter pattern called the **bayer mosaic**. This is a pattern of photosites, which are light-sensitive elements that convert the light that lands on them into digital intensity values. Each of the photosites has a spectral response curve that defines that spectra it can capture. These are typically classified as approximately capturing the red, green, and blue values in the scene, and each of their spectral response curves (SRCs) are approximately similar to the L, M, and S response functions, respectively. Once light hits the sensor, it is filtered through the mosaic and produces a tiled image of light intensities. This is called the RAW image, which is then processed to generate the image rendered on the camera screen. This process is shown in Figure 1.3.

The light intensity at each pixel on a RAW Bayer image is described by

$$R = \int R(\lambda)L(\lambda)S_d^R(\lambda)d\lambda$$

$$G = \int R(\lambda)L(\lambda)S_d^G(\lambda)d\lambda$$

$$B = \int R(\lambda)L(\lambda)S_d^B(\lambda)d\lambda$$

where R, G, and B are computed for each pixel with the corresponding red, green, and blue filter. $R(\lambda)$ is the reflectance of the scene at the captured point, $L(\lambda)$ is the illumination of the environment, and $S_d^R(\lambda)$, $S_d^G(\lambda)$, and $S_d^B(\lambda)$ are the set of SRCs of the capture device d.

As shown in 1.4, the Canon 5DsR's color gamut encompasses slightly different regions of the spectrum than the spectral locus. Notably, certain regions of the chromaticity diagram are left uncovered. This means that the camera is unable to capture, and thus unable to represent certain colors that humans can see. Since the camera is unable to accurately capture those colors, it represents them as the closest approximation it is able to capture. This approximation collapses all values from the trichromatic human observer's gamut into the closest value within the camera gamut, which means many visible colors collapse to the same color on the camera sensor and information is lost. A direct result of this information loss is **metameric error**, which arises when certain colors in the real world look distinctly different to the human eye but appear to be the same color when captured by a camera.

Camera lenses and sensor hardware are notably different from the human lens and retina in terms of refractive qualities and spatial organization. Cameras are developed to optimize for high signal-to-noise ratio and visual clarity, leading to design trade-offs that shift its SRCs away from the LMS SRCs. As the camera SRCs shift away from the LMS SRCs, the camera becomes less colorimetric which leads to metameric error. In this work we analyze metameric error using the geometry of object color solids.

1.3 Object Color Solids

Object color solids are a core theoretical framework used in metameric error analysis. A *colorsolid* is an N-dimensional convex body representing every possible color, chromaticity, and hue created by a set of 3 or more SPDs. An object color solid is a type of zonotope, which is defined as the Minkowski sum of the full color basis. This essentially combines all possible combinations of spectral vectors. Formally, it is defined as $S = \Phi(\mathcal{X})$, where $\mathcal{X}$ is the set of the object stimuli, $\mathcal{X} = \{x \in L^\infty : 0 \leq x(\lambda) \leq 1\}$, and Φ is the **color signal map** [3]. This map is a set of linearly independent spectral sensitivity functions $\phi_i(\lambda)$ that describe the space of basis functions in the color space given a range of wavelengths. The object color solid's maximal point is the white point and is typically normalized to $\vec{1} = (1, 1, ..., 1)$.

The origin is the black point, $\vec{0} = (0, 0, ..., 0)$. We refer to the achromatic segment between these points as the luminance axis. The most simple object color solid formulation assumes that the illumination of the viewing condition is fixed. For the purposes of this analysis, the illumination is assumed to be D65, daylight illumination.

In our implementation, we represent the object color solid as just its surface, the optimal object stimuli, in order to more efficiently compute the volume of intersection between multiple solids. The optimal object stimuli represent the colors that are the most luminous for each chromaticity [4]. The set of optimal object stimuli can be derived from any n functions from the linear subspace of spectral sensitivity functions, $Lin(\phi_1, ..., \phi_n)$. These functions are evaluated for all wavelengths, $(\lambda_{min}, \lambda_{max})$ where $\lambda_{min}, \lambda_{max} \in \mathbb{R}$. Since it is computationally expensive to perform these operations on continuous functions, the model is discretized using transition functions.

A **transition function** is the spectral response function, masked by an indicator function at each wavelength. The degree of the transition function is described by how many times the indicator switches from 0 to 1 or 1 to 0 in the length of the indicator string, which ranges from $(\lambda_{min}, \lambda_{max})$. For example, a set of 1-transition functions can be defined as

$$x_1(\lambda; \lambda_1) = \begin{cases} 0, \text{if } \lambda < \lambda_1 \\ 1, \text{if } \lambda \geq \lambda_1 \end{cases}$$

For a trichromatic human observer, the transition wavelengths of all optimal stimuli can be explicitly found by the equation below:

$$\xi_1 g(\lambda; \lambda_L^{max}) + \xi_2 g(\lambda; \lambda_M^{max}) + \xi_3 g(\lambda; \lambda_S^{max}) = 0$$

where ξ_1, ξ_2, $\xi_3 \in \mathbb{R}$, $g(\lambda; \lambda_L^{max})$, $g(\lambda; \lambda_M^{max})$, $g(\lambda; \lambda_S^{max})$ are the spectral sensitivities of the LMS cone pigments, and λ_L^{max}, λ_M^{max}, and λ_S^{max} are the peak absorbance wavelengths for each cone. The object color solid for the trichromatic observer is shown in Figure 1.5. This formulation can be applied to any set of spectral response curves.

Object color solids can be represented in n-dimensional space given a set of n spectral response curves. This generalization allows us to generate object color solids using multiple cameras' spectral response curves and compute metameric error using multispectral data. We apply object color theory described by Logvinenko [5] and Urban [6] to find the intersection of a set of object color solids from different capture devices in order to compute the metameric error of a color in an image.

1.4 Multispectral Capture

A multispectral image is an image that contains more than 3 channels, and thus, more than 3 spectral response curves. Multispectral images are often captured using a multispectral camera, but may also be generated as an aggregate of multiple images or extrapolated from a trichromatic image using statistical analysis. In most literature, multispectral imaging

refers to camera systems that capture light both inside and outside of the visible spectrum. However, because the application of this framework is to make captures more colorimetric, we limit the images to the visible spectrum. For the purposes of this thesis, a multispectral image refers to an image with 4 or more spectral channels, captured by multiple cameras with spectra ranging from 380nm to 780nm.

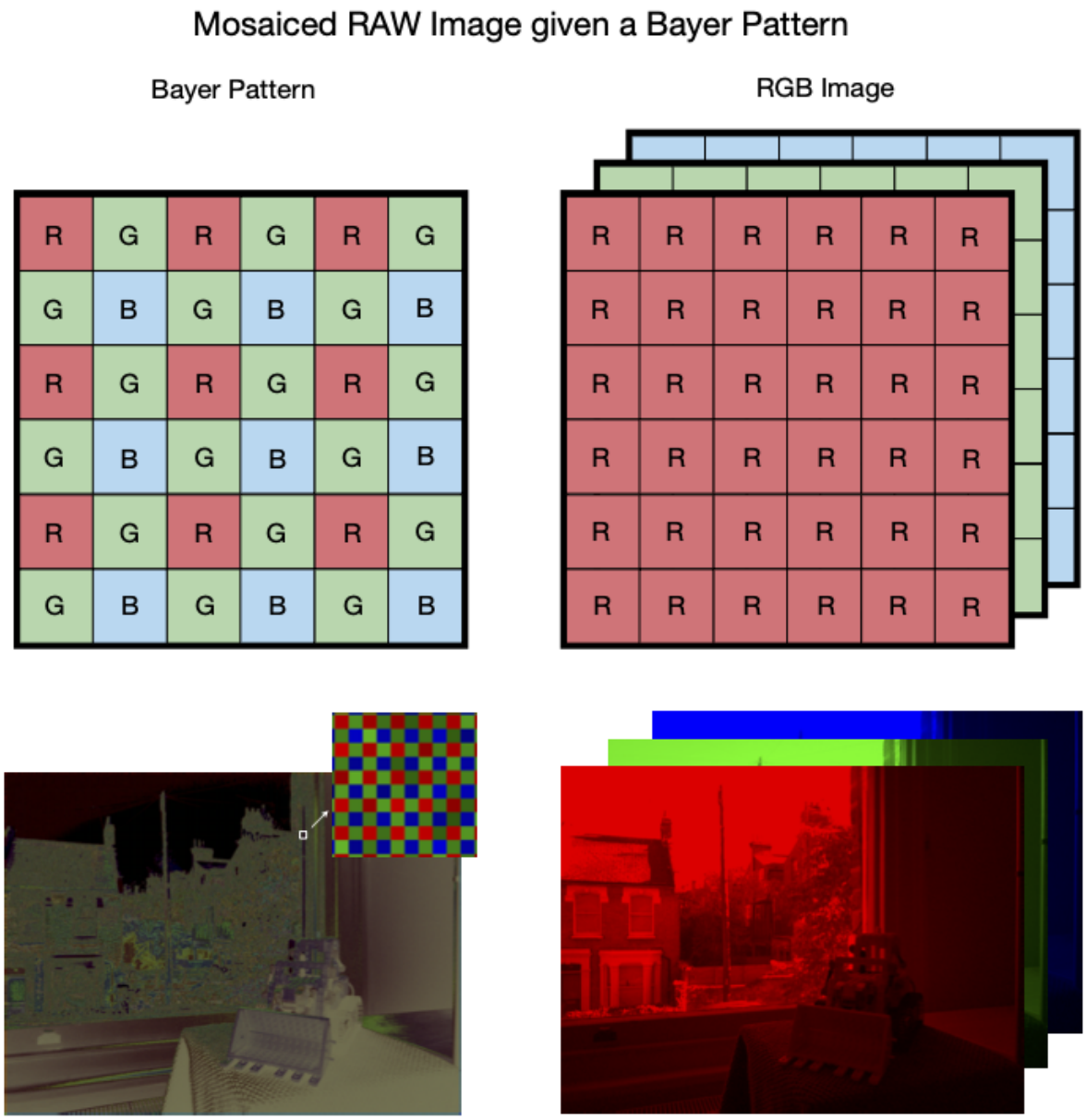


Figure 1.3: Diagram of Bayer Pattern. Top row: Diagram of mosaiced bayer sensor pattern and demosaiced RGB image. Bottom row: Cartoon of a RAW image with bayer tiling and the same image after converting it into a 3-channel image and demosaicing.

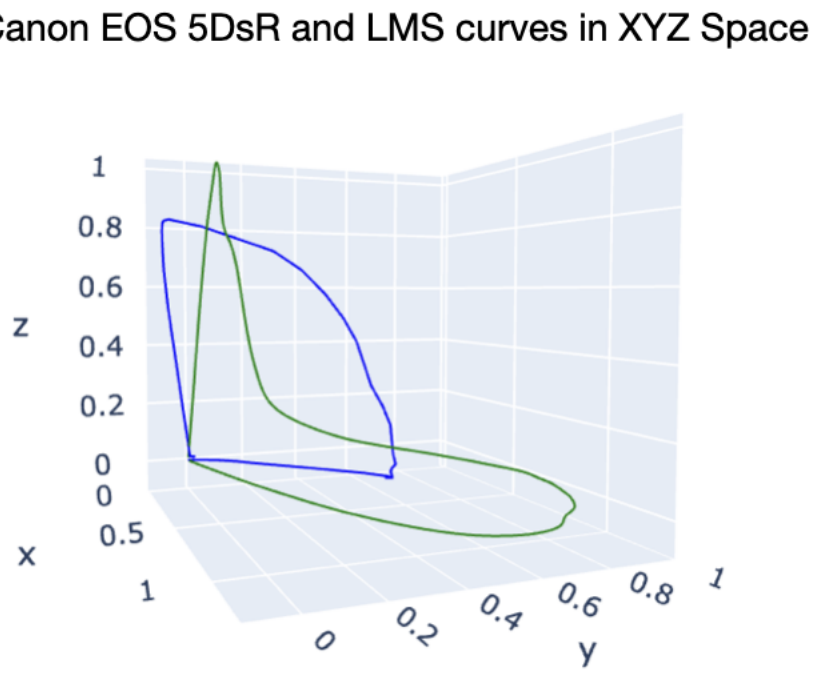


Figure 1.4: Left: 3D SRCs with LMS and a camera's curves overlayed in XYZ space. Because the SRCs differ, they cover different regions of the color gamut

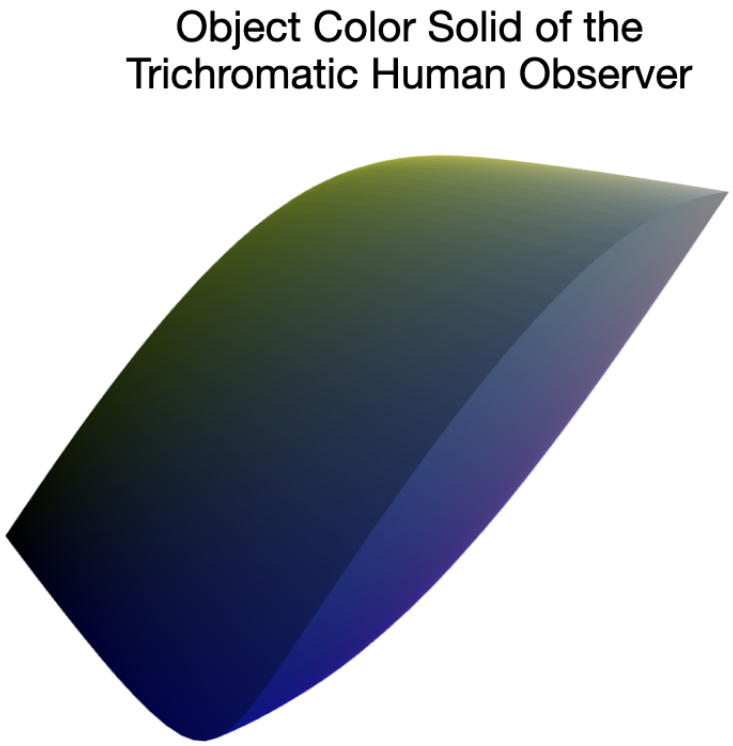


Figure 1.5: The object color solid of a trichromatic human observer, using cone fundamentals measured by Stockman and Sharpe [1] [2].

Chapter 2

Related Work

2.1 Existing Color Correction Methods

Baseline color correction in consumer cameras involve both linear and non-linear operations like white balancing, gamma correction, and tone-mapping. However, these do not lead to perfectly colorimetric color reproductions. To approach this problem, researchers have developed methods that are both statistical and model-based.

There are many proposed statistical methods that have been applied towards color correction, such as root polynomial regression, principal component analysis, and neural networks, to name a few. However, statistical models generate or approximate a reasonable candidate color based on patterns in the data priors, introducing bias based on the input data. Because the method's result is largely dependent on the training colors in the datasets, if the dataset is biased or is not fully-inclusive of all colors in the target gamut, these methods may produce large error rates.

There is also a wide set of literature that approaches the problem of color correction using principles of color theory. This literature includes, but is not limited to methods like the principal eigenvector method, local wiener-hopf equations, and the metamer mismatch volume [6], [5], [7]. Many of these techniques also require images containing a MacBeth Color Checker [8] in order to establish a known color value to match after color correction. These correction methods require the ColorChecker which constrains the types of images that can be color corrected. In this work, we apply the metamer mismatch volume formulation to real-world, multispectral image inputs to achieve improved color correction.

2.2 Metamer Mismatch Volume

The metamer mismatch volume characterizes the metameric error as a subset of the 3-dimensional LMS object color solid. To illustrate the concept of metameric error, Logvinenko describes a set of distinct color signals under illumination 1 that all appear the same under illumination 2, which means the set of reflectances in the scene are metameric under

illumination 2 [3]. In 3D space, this means that a color signal under illumination 2 maps to a set of color signals under illumination 1. Given a constant set of spectral response functions, two distinct object color solids can be constructed where each represents the range of all possible colors under a different illuminant. This subspace of colors under illumination 1 is called the metamer mismatch volume.

This framework can be adapted to computing the metamer mismatch between different observer spectral response functions under the same lighting conditions, since the illumination and spectral response curves are linearly independent in the spectral response formulation, $c_i = \int R(\lambda)L(\lambda)S_i(\lambda)$, where c_i is the output color for a pixel in SRC channel i, $R(\lambda)$ is the scene reflectance, $L(\lambda)$ is the scene illuminant, and $S_i(\lambda)$ is the spectral response curve for channel i [5]. Thus, there is a non-injective mapping between colors captured by a camera, in the camera RGB color space, and colors seen in real-life, in the LMS color space. This framework has been applied towards color correction across illuminants [9], but to the best of our knowledge, we are the first to apply this theory to improve camera color representation. Our method uses multispectral images to introduce further constraints to this formulation, ultimately producing a more colorimetric color correction estimation.

2.3 Multispectral Camera Capture

This work seeks to use high quality and low barrier-to-entry multispectral capture set-ups in order for our method to be easily accessible to both researchers and recreational users. In this section, we discuss the most widely-available multispectral capture options: using a multispectral camera, using a 3-channel camera with lens filters, and generating multispectral from RGB images using statistical techniques.

Though multispectral cameras produce the most reliable and precise data, they are relatively expensive. Configuring a traditional camera to be multispectral requires removing the UV and IR cut filters inside of the camera and stacking bandpass filter lenses to capture the desired spectrum. This can be achieved using products like the Kolari full-spectrum filter sets. However, since most multispectral capture techniques require highly specialized and expensive equipment, they may be inaccessible to casual photographers or researchers. A simple multispectral capture system can be built upon a StereoPi system by using a raspberry pi with two cameras on board, but these images are lower quality than those of an off-the-shelf multispectral camera. Statistical techniques require minimal equipment but introduce precision errors due to incorrect color extrapolation, so they are an not an ideal choice for color correction analysis. Specifically, precision errors may be introduced by biased datasets, the use of cameras with different spectral responses in the data, and overfitting. The 3D reconstruction-based technique described in this thesis seeks to circumvent economical barriers and precision errors by using any available camera to construct a multispectral image with physically-based accuracy metrics. While this thesis focuses on multispectral capture within the visible spectrum, this technique can be extended to any set of spectra.

Chapter 3

Volumetric Multispectral Inference

Our technique leverages modern breakthroughs in computer vision and graphics to reconstruct multispectral, 3D representations of photographed scenes. By generating a 3D reconstruction for each camera's captures, we achieve pixel alignment between photographs without having the location of each camera's captured images being exactly the same. This set-up allows us to more easily compare spectral data. The 3D environment also allows us to generate a large dataset of our extrapolated spectral response, since we are able to virtually capture new images of the scene after it has been reconstructed and post-processed. In this work, we present a novel Gaussian Splatting method that uses RAW images from multiple cameras to produce a multispectral scene. We also create an operation-by-operation, customizable image processing pipeline within Nerfstudio.

3.1 Multispectral Gaussian Splatting using RAW images

Gaussian Splatting is a reconstruction method used convert a sparse set of photographs of a scene from continuously-varying viewpoints into a virtual 3D environment [10]. It represents scenes as a set of 3D gaussians of varying sizes and orientations. Gaussian splatting is most suitable for this application because it achieves faster render times and comparable accuracy to implicit-function-based techniques like Neural Radiance Fields (NeRF) [11], Mip-NeRF [12], and Zip-Nerf [13], and has higher accuracy than grid-based techniques like Instant-NGP [14].

Our model is a flexible extension of Gaussian Splatting, allowing users to capture and reconstruct scenes with any number of spectral channels. We integrate this implementation into Nerfstudio [15] and extend the 3-dimensional gaussian splat into an n-dimensional gaussian splat, meaning the model is trained over n input spectral channels. During training, these channels are jointly optimized with a single, common opacity value for each 3D gaussian. We further extend Gaussian Splatting to train on RAW images, taking inspiration from RawNeRF [16]. It is necessary to use RAW images because converting the RAW image

into a JPEG, PNG, or any other human-viewable format is a lossy process - information is lost due to the approximated conversion matrices, value clipping, and reduced bit depth of the image. Thus, it is important to use the RAW image capture when editing an image in order to preserve the greatest amount of spectral information. Similarly, it is vital to use RAW images when analyzing and color correcting images.

To configure the Gaussian Splatting model to run RAW images, we first run COLMAP [17] to find the camera poses of each input image and generate a 3D point cloud. Since COLMAP uses feature matching to run a structure-from-motion algorithm, it is agnostic to color differences between the images. Thus, we use images captured by multiple cameras in COLMAP to generate a point cloud informed by all of the data. Because COLMAP is currently not compatible with RAW images and is unaffected by color, we run COLMAP on the images in PNG format. We then use the COLMAP output and RAW images, labelled by spectrum, as input to the Gaussian Splat model to generate an N-dimensional multispectral 3D scene. This pipeline is described in Figure 3.1. To the best of our knowledge, this is the first instance of training N-dimensional Gaussian Splats as well as using RAW images in Gaussian Splats.

To improve the reconstruction performance, we use the loss function introduced by RawNerf [16], which is an approximation of an L2 loss over a tonemapped image. This loss function is defined as $\tilde{L}_{\psi}(\hat{y}, y) = \sum_{i=1}^{n} \frac{\hat{y}_i - y_i}{sg(\hat{y}_i)+\epsilon}$, where ψ is a tone curve, $\hat{y}$ is the rendered estimated color, y is the noisy observed intensity that has aψ applied to it, $sg()$ is a stop-gradient, and $\epsilon = 10^{-3}$. Using this loss qualitatively improved our results over the baseline L2 loss.

Once the scene is generated, our virtual set-up allows us to generate multi-spectral, physically-accurate 3D scenes, in which we can set up a virtual camera and quickly capture a large number of images from different cameras at the same extrinsic location as shown in Figure 3.2. This allows us to have pixel-perfect mappings between the colors in each of our images, which produces additional spectral data for the analysis.

3.2 Camera Processing Pipeline

When an image is taken in RAW format, there are a number of steps that are taken to convert it into a common file format like JPEG or PNG:

1. **Linearization**: The image is first normalized using the black point and white point in the image. Specifically, this is defined as $\frac{color - blacklevel}{whitelevel - blacklevel}$, where the blacklevel is the signal the camera sensor reads when when no light is present and the whitelevel is the highest possible value that the sensor image can read. This operation linearizes and normalizes the pixel values to the [0.0, 1.0] range. Due to sensor error, there may be values that exist either below 0 or above 1, which are then clipped. The linearization operation ensures that the R, G, and B channels are all contributing equally to the image, rather than the captured light intensity being biased towards the SRCs with the largest area under its curve.

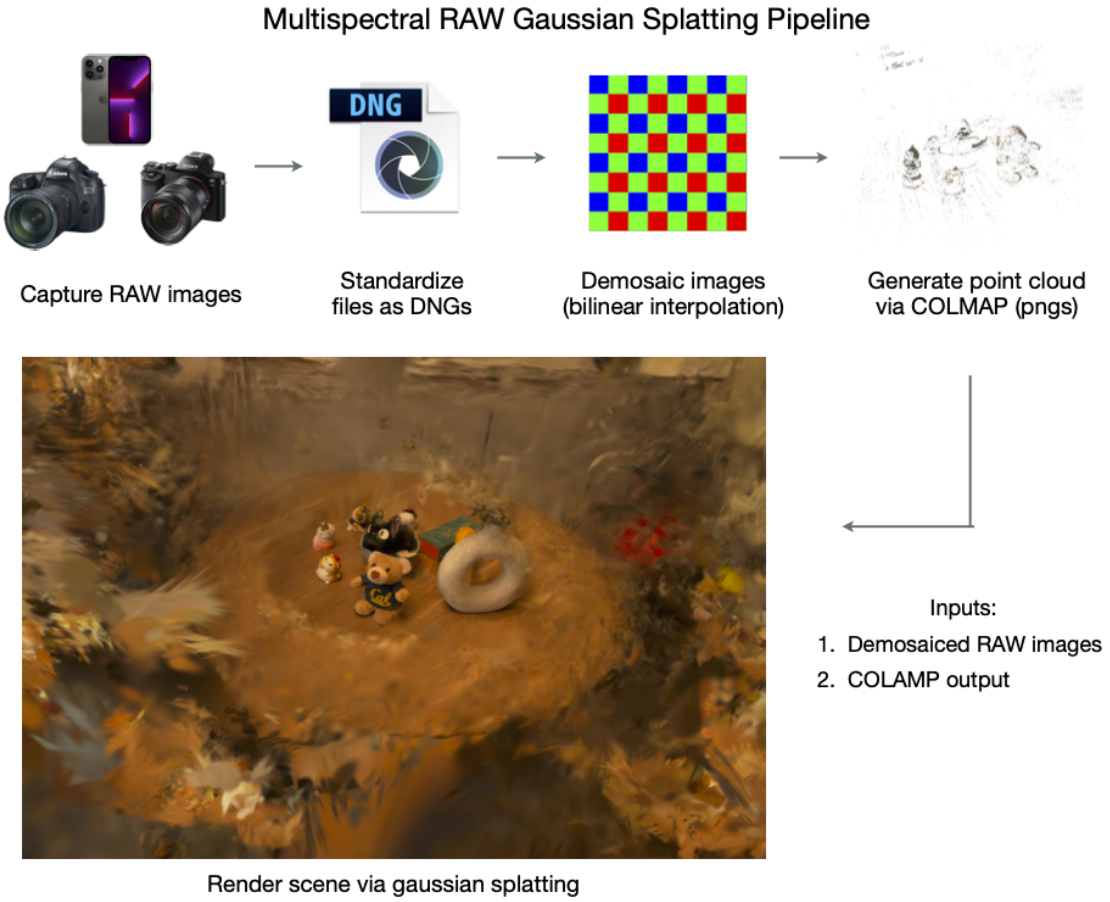


Figure 3.1: 3D processing pipeline to convert RAW captures into an interactive, multispectral scene

2. **White Balancing**: This operation rescales the channels in an image by making a white pixel in the image represented as closely to white as possible. The white balancing parameter in the Adobe DNG specification, 'AsShotNeutral' [18], is a linear normalization factor that is computed at the time of capture. Since this is computed at time of capture, the real illumination is unknown, and is thus approximated as an interpolation of the Standard Illuminant A and D65 Illuminant (Daylight). This means white balancing is an imprecise process, which leads to color reproduction error.

3. **Demosaicing**: The image is interpolated so it is no longer organized as a 1-channel bayer pattern, but a 3-channel set of stacked RGB pixels. While there are many complex techniques to interpolate these values by taking into account lens parameters and sensor design, the naive demosaicing implementation is simply bilinear interpolation.

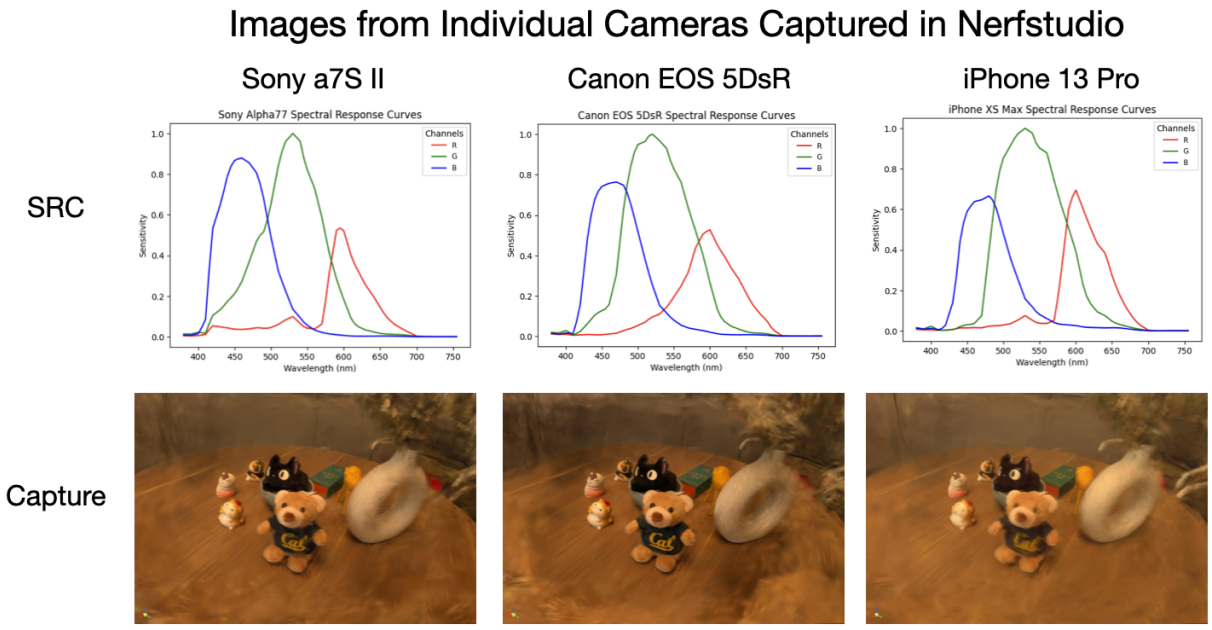


Figure 3.2: The same image rendered of the same scene from three different cameras: Sony a7s, Canon EOS 5Dsr, and iPhone 13 Pro. This is generated by taking a multispectral capture in the 3D environment and separating its channels into its representative cameras' images.

In this pipeline, we implement demosaicing as a bilinear interpolation for efficiency and simplicity.

4. **Color Space Conversion (Camera RGB → CIE 1931 XYZ → sRGB)**: The pixel values are transformed from the camera's color space to sRGB, using CIE 1931 as an intermediary color space. Every camera stores at least one color transformation matrix in its RAW image metadata, and this matrix will transform the image from camera color space to CIE 1931 color space. This matrix is referred to as 'ColorMatrix1', and optionally there is a second 'ColorMatrix2'. Both of these matrices perform the same color conversion, but ColorMatrix1 is computed assuming the scene is under Standard Illuminant A, and ColorMatrix2 is computed for use under D65 Illumination [18]. After the image is in CIE 1931 color space, it is transformed to sRGB color space using transformation matrices defined by the International Electrotechnical Commission [19]. The **sRGB** color space is the color gamut of most common display systems, making it an important and necessary color space to operate in for the final steps of this process.

5. **Gamma Correction**: This is a non-linear operation that rescales camera tonal values

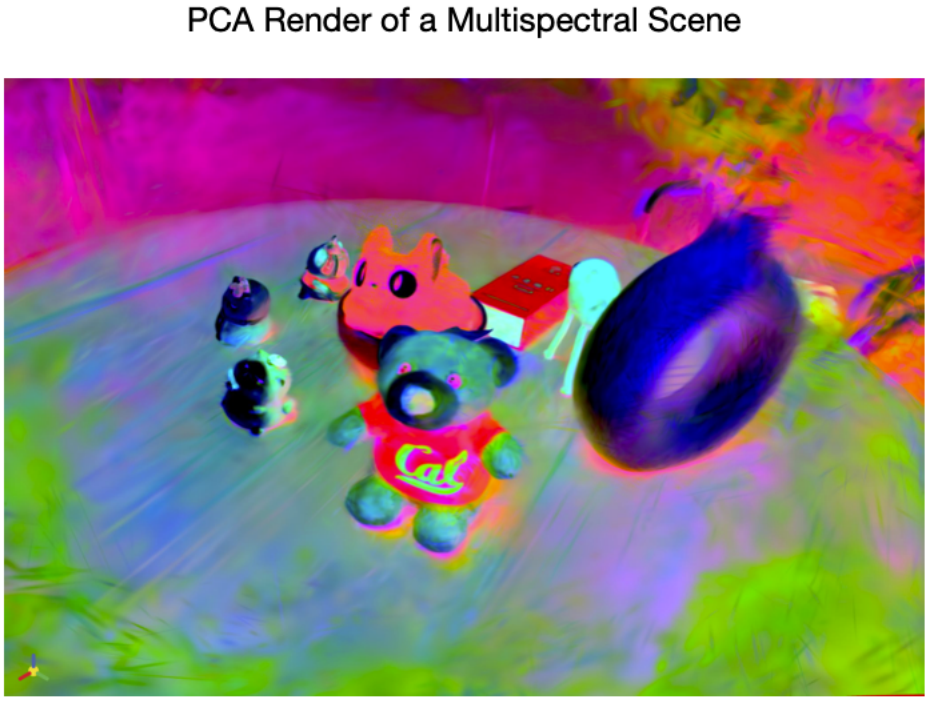


Figure 3.3: PCA visualization of a multispectral reconstruction to show what channels contribute the most to each region of the scene

to make the image more perceptually uniform.

6. **Tone Mapping**: This operation modifies tonal values to approximate high dynamic range (HDR) for more colorimetric and visually-striking renderings on digital displays.

A visualization of this full pipeline is shown in 3.4.

As described above, many of the camera and capture-specific information necessary to process the image are stored in the RAW file's metadata. Thus, it is vital to store RAW images with its metadata in order to preserve shot-specific processing information.

It is important to note that while ColorMatrix and ColorMatrix2 in the Color Space Conversion step are commonly used and lead to approximately correct solutions, they are approximations computed by a small set of physical experimental results [20]. Our method is a color correction framework that converts from linearized color values to sRGB can replace the existing white balancing and color conversion operations to reduce additional error introduced from approximation and improve overall image colorimetry. This new color correction framework is described in Chapter 4.

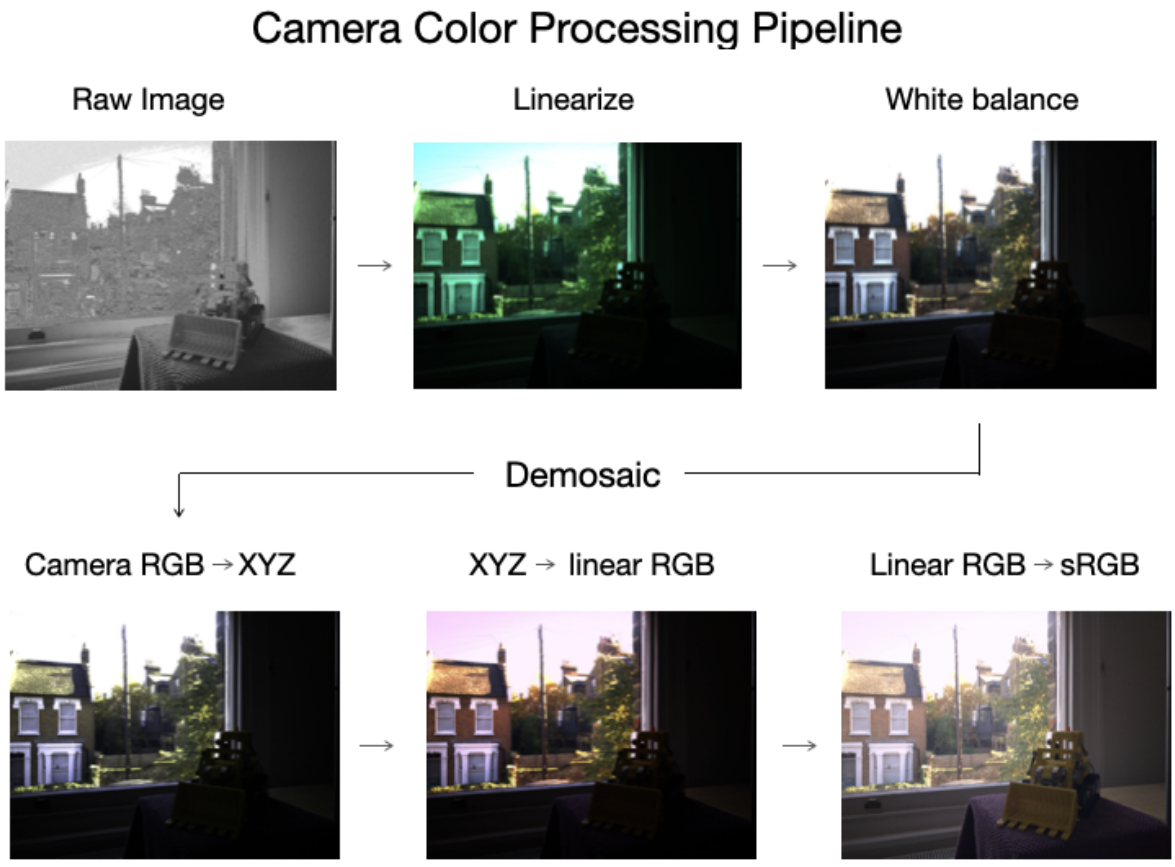


Figure 3.4: Visualization of the camera image processing pipeline for a given image

Chapter 4

Applying Color Theory to Reduce Metameric Error

This work finds the candidate color using only the information provided by a multispectral scene capture, applying first principles to produce an accurate color correction with no bias. Specifically, we apply Logvinenko's formulation of the metamer mismatch volume to two sets of different spectral response curves and extends it to real-world data [5]. We use camera spectral response data recorded by Image Engineering in a publicly-available dataset and images captured by Canon EOS 5DsR, Canon EOS R, and the Nikon D7100.

4.1 Metamer Mismatch Volume

The metamer mismatch volume is a color error metric that quantifies all possible colors that may be metamerisms when comparing two spectral capture devices. Using Logvinenko's general formulation [5], we construct an object color solid in N-dimensional space where each dimension represents a different spectral response curve from the set of cameras and LMS cone fundamentals. In this formulation, if we are finding the metamer mismatch for a color captures by a camera with three spectral response curves, we would represent the volume as a 6-dimensional object color solid where three of the dimensions are from the camera and three of the dimensions are the L, M, and S axes. However, as the number of captured spectral response curves increases, the computational complexity of generating and processing this volume also increases as a result of the curse of dimensionality. To efficiently compute the mismatch volume in any higher dimensional space, the N-dimensional solid is constructed as N 3-dimensional solids in L, M, S space, where each camera spectral response is computed as a new color solid. This mechanism produces (N-3) 3D color solids in LMS space. The volume of intersection of these solids is the metamer mismatch, which is the set of all probable LMS values for the input color value. For example, to find the mismatch volume of the color (0.2, 0.3, 0.1) captured by a Canon 5DsR, there would be 3 object color solids plotted in LMS space - one representing the R = 0.2 value, one representing the G =

0.2 value, and one representing the B = 0.1 value. The region of intersection of these three volumes is the metamer mismatch space, as shown in Figure 4.1

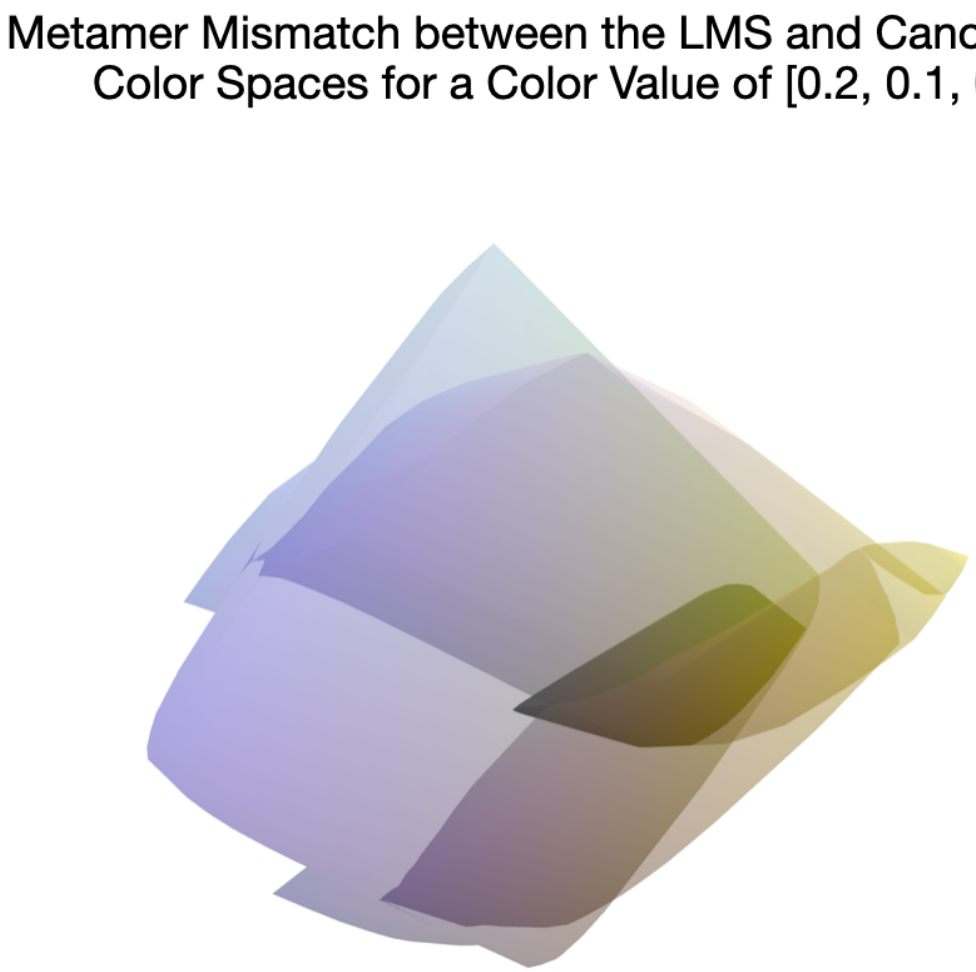


Figure 4.1: Visualization of metamer mismatch between the LMS and single-camera space

The full metamer mismatch region is an equally-probable colorimetric solution given the input information. We select the center of mass of the mismatch volume as the initial solution, as proposed by Finlayson [21]. We can measure the colorimetric improvement by capturing an image with a known reflectance, like a Macbeth color checker, and compare the captured color and the corrected color from the mismatch volume to the target LMS color. When we represent this in CIELAB, we can use Euclidean distance compute the perceptual difference between these terms to quantify the improvement of our method.

When evaluating the colors from the full image, we map each pixel in an image to a metamer mismatch in the LMS space, generating a set of mismatch volumes that represent the possible space of "correct" values in an image. This set of mismatch volumes for pixel values of (0.01, 0.01, 0.01) and (0.2, 0.2, 0.2) for a given input image is shown in Figure 4.2.

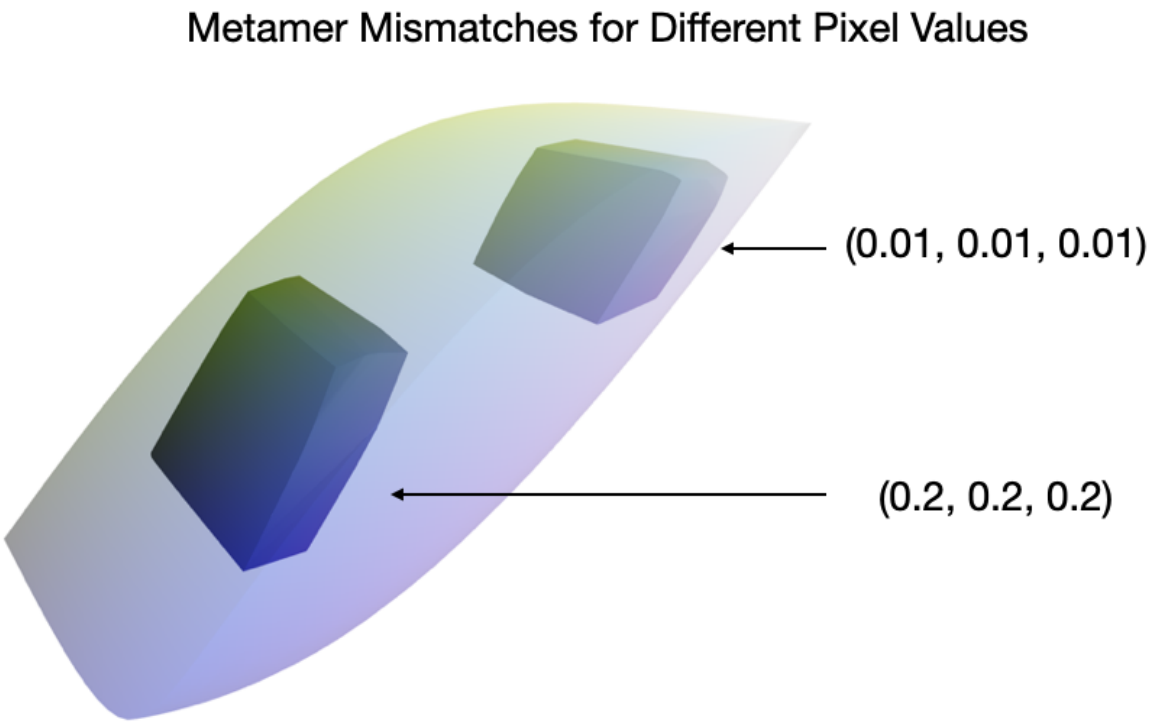


Figure 4.2: Demonstration of how pixels of an image are mapped to a metamer mismatch. Each pixel color will generate a new metamer mismatch region that exists within the LMS object color solid

4.2 Luther Condition

The Luther condition is a theoretical measure of the linearity of color reproduction systems, and is often used when evaluating if a camera is colorimetric. The Vora Value is a metric used to quantify the Luther Condition and quantifies how colorimetric a set of spectral response curves are [22]. The Vora-Value ranges from [0, 1], where 0 is the least colorimetric and 1 is the most colorimetric. Given a set of camera spectral response curves Q and the trichromatic human SRCs X, the Vora-Value is can be defined as $v(Q, X) = \frac{1}{3}trace(P\{Q\}P\{X\})$, where P is a least-squares matrix projector, $P\{Q\} = Q(Q^TQ)^{-1}Q^T$ [23]. This work uses Vora Values

to validate the behavior of the metamer mismatch volume as spectral response curves input changes.

Chapter 5

Discussion

5.1 Analysis of Results

To validate the application of Logvinenko's theory to real images captures by different cameras, we use properties of the metamer mismatch volume, the value of the corrected color, and the computed vora values.

We call the source spectra the camera spectral response, and the target spectra the trichromatic human spectral response. When the source and target spectra are different, the metamer mismatch is a non-zero volume, which means one value in source space maps to a set of values in target space. However, when the source and target spectra are the same, we expect there to be a perfect 1:1 mapping between colors in both spaces. This means the metamer mismatch volume is a point. We demonstrate this behavior in Figure 5.1.

When regarding mismatch volumes found along the luminance axis, we also expect the mismatch volumes to be the largest in volume when located near the center of the target object color solid, and to decrease in size as the mismatches approach the poles of the axis. Since there are no metamerisms on the object color solid boundary, it is expected that the subspace of possible metamerisms decreases as the color approaches this region, and thus the metamer mismatch volumes, will decrease.

As expected, as we increase the number of different spectral response curves included in our computation, the vora value decreases. Since the volume of the metamer mismatch is also a measure of being colorimetric, we expect the mismatch volume to decrease in size. This property is shown in Figure 5.2, which shows Vora Value improvement when additional SRCs are included. In the top row, the computed Vora Value of 0.92 from the Canon EOS 5DsR increases by 0.27 with the addition of the R, G, and B channels from the Canon EOS R camera. The bottom row shows that the Vora Value continues to increase as unique spectral response curves are included in the input. This demonstrates that there is colorimetric improvement as more SRCs are used.

Finally, we can measure the distance between the captured color in the original images, the proposed corrected color, and the real color using Euclidean distance in CIELAB space.

The original color is the color generated from the original camera capture. The proposed corrected color is computed as the center of mass of the mismatch volume. We measure distance in CIELAB because it is a perceptually-uniform color space. Computing the distance between points in CIE 1931 XYZ space is will lead to non-linear perceptual distances, and thus is not an accurate measure of distance for a perceptual metric of an image such as its color quality. We establish a set of real colors in the scene by capturing a ColorChecker in our dataset, where each tile on the checker has a known reflectance function [8]. This allows us to directly measure the color improvement.

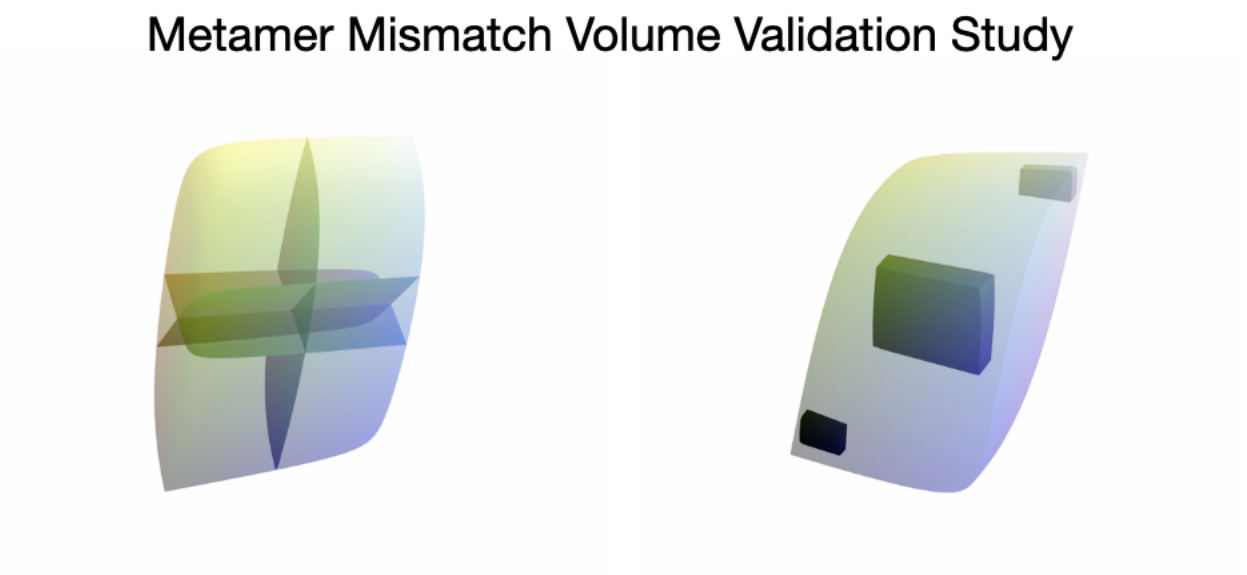


Figure 5.1: Two test cases to confirm the implementation works as expected. Left: source SRC and target SRC are both LMS, and for a given color, the mismatch volume is a point. This means it is a perfect 1:1 mapping between the solids, which is expected. Right: Metamer mismatch volume is larger towards the center and smaller towards the black and white points.

5.2 Future work

This color correction and analysis framework opens research directions in a variety of spaces, including photography, machine learning, and imaging. Most directly, it can be used as an editing tool by photographers and videographers to make their shots more colorimetric prior to doing any additional editing. This color correction method is complementary to traditional image processing pipelines and can be easily integrated into photographer workflows. The data captured and corrected can also be used to generate high quality, color-corrected

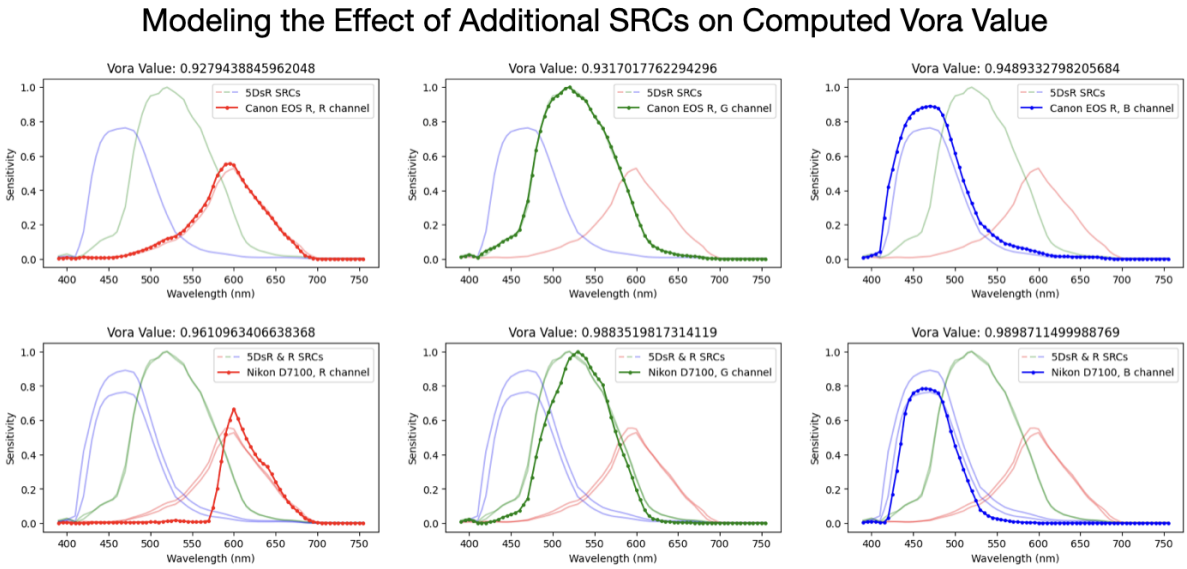


Figure 5.2: A set of charts demonstrating how the Vora Value increases as new spectral response curves are added to the Q matrix. The initial set of SRCs are from the Canon EOS 5DsR camera, and have a baseline Vora Value of 0.921. In the top row, each channel of the Canon EOS R camera is added to the set of SRCs. The bottom row adds the SRCs of a third camera, the Nikon D7100.

datasets for machine learning applications. Correcting the color values has been shown to improve model performance [24]. The RAW Gaussian Splatting technique itself can be extended to generate high quality hyperspectral datasets from any set of camera responses and filters. Capturing hyperspectral images requires expensive equipment which have multi-minute capture times, which largely constraints the nature of the objects that can be captured. From a multispectral dataset ranging many bands, it may be possible to generate a set of plausible hyperspectral images, which would be a valuable contribution to the biology, robotics, and imaging communities. The metamer mismatch technique can also be modified to be illumination agnostic [9], which would allow for a more generalized color-correction method that does not assume a scene illuminant. This technique is generalizable and can be applied to any number of spectral response curves. As such, this framework will unlock analysis techniques for tetrachromatic human vision that has never been explored before.

Chapter 6

Conclusion

In this work, we have introduced a method for image color correction using photos from multiple cameras, Gaussian Splatting, and metamer mismatch theory. We present a novel method to generate multispectral images using Gaussian Splatting and Nerfstudio. We introduce Multispectral-Splat, an N-dimensional gaussian splat reconstructed from a set of continuously-varying images captured by two or more cameras. This formulation sets up a pixel-perfect mapping between captures, by which we can create a multispectral image as a post-processing step. We further extend the Multispectral-Spat to run on RAW images, which preserves more spectral information and provides greater flexibility in post-capture color processing.

Using this tool, we adapt Logvinkenko's metamer mismatch theory to compare individual pixels' color values across the multispectral channels and estimate a more colorimetric color than the original pixel. We validate the application of this approach to real data captured by 3 cameras with slightly different color gamuts. We capture scenes containing a Macbeth ColorChecker in order to validate our color correction results. Our method enables new multispectral and hyperspectral data creation techniques and opens up possibilities for color-correction processes with unprecedented precision.

Bibliography

[1] Lindsay T. Sharpe Andrew Stockman. The spectral sensitivities of the middle- and long-wavelength-sensitive cones derived from measurements in observers of known genotype. 2000.

[2] C. Fach A. Stockman, L. T. Sharpe. The spectral sensitivity of the human short-wavelength sensitive cones derived from thresholds and color matches. *Vision Res.*, 39: 2901–2927, 1999.

[3] Vladmir L. Levin Alexander B. Logvinenko. *Foundations of Colour Science: From Colorimetry to Perception*. Wiley, 2022.

[4] Erwin Schrödinge. Theorie der pigmente von grösster leuchtkraft. *Annalen der Physik*, 367(15):603–622, 1920.

[5] Brian Funt Alexander D. Logvinenko and Christoph Godau. Metamer mismatching. *IEEE Transactions on Image Processing*, 23(1):34–43, 2013.

[6] Philipp Urban and Rolf-Rainer Grigatn. Metamer density estimated color correction. *Signal, image and video processing*, 3:171–182, 2009.

[7] Brian Funt Xiandou Zhang and Hamidreza Mirzaei. Metamer mismatching in practice versus theory. *JOSA A*, 33(3):A238–A247, 1973.

[8] D. Pascale. Rgb coordinates of the macbeth colorchecker. *The BabelColor Company*, 6, 2006.

[9] Alexander D. Logvinenko. An object-color space. *Journal of Vision*, 9(11):5–5, 2009.

[10] Thomas Leimkühler Bernhard Kerbl, Georgios Kopanas and George Drettakis. 3d gaussian splatting for real-time radiance field rendering. *ACM Transactions on Graphics*, 42(4):1–14, 2023.

[11] Matthew Tancik Jonathan T. Barron Ravi Ramamoorthi Ben Mildenhall, Pratul P. Srinivasan and Ren Ng. Nerf: Representing scenes as neural radiance fields for view synthesis. *Communications of the ACM*, pages 99–106, 2021.

[12] Matthew Tancik Peter Hedman Ricardo Martin-Brualla Jonathan T. Barron, Ben Mildenhall and Pratul P. Srinivasan. Mip-nerf: A multiscale representation for anti-aliasing neural radiance fields. *IEEE/CVF International Conference on Computer Vision*, pages 5855–5864, 2021.

[13] Dor Verbin Pratul P. Srinivasan Jonathan T. Barron, Ben Mildenhall and Peter Hedman. Zip-nerf: Anti-aliased grid-based neural radiance fields. *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 19697–19705, 2023.

[14] Christoph Schied Thomas Müller, Alex Evans and Alexander Keller. Instant neural graphics primitives with a multiresolution hash encoding. *ACM transactions on graphics (TOG)*, 41(4):1–15, 2022.

[15] Evonne Ng Ruilong Li Brent Yi Terrance Wang Alexander Kristoffersen et al Matthew Tancik, Ethan Weber. Nerfstudio: A modular framework for neural radiance field development. *ACM SIGGRAPH 2023*, pages 1–12, 2023.

[16] Ricardo Martin-Brualla Pratul P. Srinivasan Ben Mildenhall, Peter Hedman and Jonathan T. Barron. Nerf in the dark: High dynamic range view synthesis from noisy raw images. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 16190–16199, 2022.

[17] Johannes L. Schonberger and Jan-Michael Frahm. Structure-from-motion revisited. *IEEE conference on computer vision and pattern recognition*, pages 4104–4113, 2016.

[18] D. Andrew Rowlands. Color conversion matrices in digital cameras: a tutorial. *Optical Engineering*, 59(11):110801–110801, 2020.

[19] British Standards Institution. Multimedia systems and equipment: Colour measurement and management. colour management: Default rgb colour space: Srgb. *British Standards Institution*, 2000.

[20] P. M. Hubel. Matrix calculations for digital photography. *Color and Imaging Conference*, 5:105–111, 1997.

[21] G. Finlayson and P. Morovic. Metamer constrained color correction. *J. Imag. Sci. Technol.*, 44(4):1–16, 2000.

[22] Poorvi L. Vora and H. Joel Trussell. Measure of goodness of a set of color-scanning filters. *OSA A*, 10(7):1499–1508, 1993.

[23] Yuteng Zhu and Graham D. Finlaysonr. A mathematical investigation into the design of prefilters that make cameras more colorimetric. *Sensors*, 20(23), 2020.

[24] Shreyank N. Gowda and Chun Yuan. Colornet: Investigating the importance of color spaces for image classification. *Computer Vision–ACCV 2018: 14th Asian Conference on Computer Vision*, IV(14):581–596, 2019.