# Incorporating Physical Constraints for Color Correction and Spectral Recovery

**Abdullah Kucuk**

A thesis submitted for the degree of

Doctor of Philosophy

School of Computing Science

University of East Anglia

**September 2025**

# *Abstract*

RGB cameras sense the colour signal with three device-dependent measurements that differ from the standard Human Visual System (HVS). While standard human colour vision is mediated by cone cells, each RGB camera uses its own device-specific spectral sensitivity functions. Therefore, a *colour correction* step is necessary to convert these values into standard colour spaces. Sampling light with only three values also loses spectral information about the scene, which is important in many scientific fields. To address this, *spectral reconstruction* algorithms have been developed to estimate spectra solely from RGB data. However, hyperspectral or multispectral cameras, which capture more comprehensive spectral information, are preferable. Unfortunately, these cameras often encounter resolution limitations due to the constraints of sensing hardware. Traditional *pan-sharpening* techniques are designed to enhance the spatial resolution of these images by leveraging their higher-resolution panchromatic counterparts. RGB-guided approaches, however, fuse high-resolution RGB images instead of panchromatic images, offering the promise of improved spectral estimation.

This thesis explores the topics of colour correction and spectral recovery algorithms for RGB, multispectral, and hyperspectral cameras by incorporating physical constraints. The contributions are presented in four main chapters as follows.

We first examine colour correction by reviewing existing algorithms for RGB cameras and comparing traditional regression methods with neural network (NN) solutions. Our findings indicate that while NNs outperform simple linear regression, polynomial and root-polynomial regressions still surpass current NN approaches. We then address the issue of exposure invariance in colour correction algorithms and propose an exposure-invariant neural network solution.

Next, we enhance spectral recovery algorithms by incorporating physical constraints. Revisiting the Matrix-R algorithm, we illustrate that any spectrum can be decomposed into its fundamental metamer and metameric black components. The fundamental metamer, which the camera detects, can be directly calculated from RGB values given the camera's sensitivity functions. The metameric black component, orthogonal to the camera sensor space, does not affect the RGB values. Despite the ability to calculate the correct fundamental metamer for a given RGB, existing spectral reconstruction and pan-sharpening algorithms still exhibit errors in the fundamental metamer part of the signal. The Matrix-R post-processing framework ensures that any spectral reconstruction algorithm has the correct fundamental metamer. Here, we further develop the framework for pan-sharpening methods and include the generalisation for multispectral images. Additionally, we extend the Matrix-R approach and demonstrate that representing spectra in lower dimensions

further enhances algorithm performance.

Furthermore, we present another novel pan-sharpening post-correction least-squares solution by upsampling hyperspectral pixels and merging them with high-resolution RGB data. We prove that these post-corrected spectra have the built-in Matrix-R feature and consistently have the correct fundamental metamer. The standalone version of this algorithm also achieves competitive performance compared to the latest state-of-the-art deep learning pan-sharpening solutions.

Finally, we investigate a dual-camera solution for colour correction where high-resolution RGB and corresponding low-resolution multispectral pixels are available, similar to the setup for the pan-sharpening problem. We propose a novel colour correction method using linear regression by integrating upsampled multispectral and the corresponding RGB pixels. In experiments, the technique delivers superior performance than the classical RGB-only colour correction algorithms.

# *Acknowledgements*

First and foremost, I would like to thank my PhD supervisor, Prof. Graham Finlayson, for giving me the brilliant opportunity to work in colour and imaging science. It has been a unique privilege to benefit from his knowledge and experience over the past four years, from mathematical concepts to publishing research. I am grateful for his constant support and motivation, which has continually pushed me to challenge my limits.

I am also grateful to Spectricity for making this research possible by funding it, providing an exciting PhD internship opportunity, and allowing me to gain hands-on experience with camera algorithms. I would like to thank everyone on the team for their support and assistance, especially Dr. Bart Goris, Dr. Alex Borgo, and Dr. Jonathan Borremans.

I also want to thank my colleague Dr. Ethan (Yi-Tun) Lin for his advice, support, and friendship. It has been a pleasure to work with you.

I consider myself fortunate to have had several great teachers and professors who changed my life. I want to send special thanks to them, in chronological order: my primary school teacher Meral Gürsoy, who discovered my potential and directed my energy towards science until I became the top student in school; my high school history teacher Berrin Malatyalı, whose teaching made history one of my favorite subjects; Prof. Şahin Emrah, who taught me in both high school and university and fostered my love for math and algorithms; and Dr. Hacer Yalım Keleş, who encouraged me to pursue computer vision and AI during my undergraduate studies. Finally, a big thanks again to Prof. Graham Finlayson, from whom I have learned so much and who opened the door to the world of colour and imaging science for me.

Finally, and most importantly, I would like to thank my family for allowing me to be who I am today. I am especially grateful to my mother for her unconditional love and support at every step of my life, and to my father, who always supported me in all my choices. He will forever inspire me with his career, knowledge, and wisdom.

# *List of Publications*

**Journal Publications**

(Drafted) Kucuk, A., & Finlayson, G. D. A Novel Approach to Color Correction.

(Drafted) Finlayson, G. D., & Kucuk, A. Post-processing Pan-sharpening.

(Submitted) Finlayson, G. D., Lin, Y. T., & Kucuk, A. Matrix-R Theory: A Simple Generic Method to Improve All RGB-Guided Spectral Recovery Algorithms. Sensors.

Kucuk, A., Finlayson, G. D., Mantiuk, R., & Ashraf, M. (2023). Performance Comparison of Classical Methods and Neural Networks for Colour Correction. Journal of Imaging.

**Conference Publications**

Lin, Y. T., Finlayson, G. D., & Kucuk, A. (2023). An Optimality Property of Matrix-R Theorem, Its Extension, and The Application to Hyperspectral Pan-sharpening. In Color and Imaging Conference (CIC).

Kucuk, A., Finlayson, G. D., Mantiuk, R., & Ashraf, M. (2022). An Exposure Invariant Neural Network for Colour Correction. In Color and Imaging Conference (CIC).

Kucuk, A., Finlayson, G., Mantiuk, R., & Ashraf, M. (2022). Comparison of Regression Methods and Neural Networks for Colour Corrections. In London Imaging Meeting (LIM).

# Contents

# List of Tables

# List of Figures

# Chapter 1

# Introduction

Digital imaging has become an integral part of modern life, encompassing everything from smartphone photography to complex industrial applications. While early digital imaging systems were designed primarily for visual reproduction and documentation, modern applications have expanded to treat cameras as sophisticated measurement devices, underpinning technologies such as autonomous navigation, satellite remote sensing, and automated industrial inspection. Among various imaging systems, RGB cameras are the most popular solution due to their ability to capture sufficient colour information through three channels that broadly mimic the trichromatic nature of the Human Visual System (HVS).

Despite their ubiquity, RGB cameras face two significant challenges that we will investigate in this thesis. First, the spectral sensitivity functions that RGB cameras rely on are inherently device-specific and differ substantially from the cone cells in the human retina responsible for colour perception. Theoretically, for a camera to capture colour exactly as a human does, its sensors would need to satisfy the *Luther condition*, meaning they would need to be a linear transformation of the human cone fundamentals. However, due to manufacturing constraints and design choices focused on signal-to-noise ratios, this condition is rarely met in practice. This discrepancy creates a mismatch between the camera's raw output and how humans perceive colour.

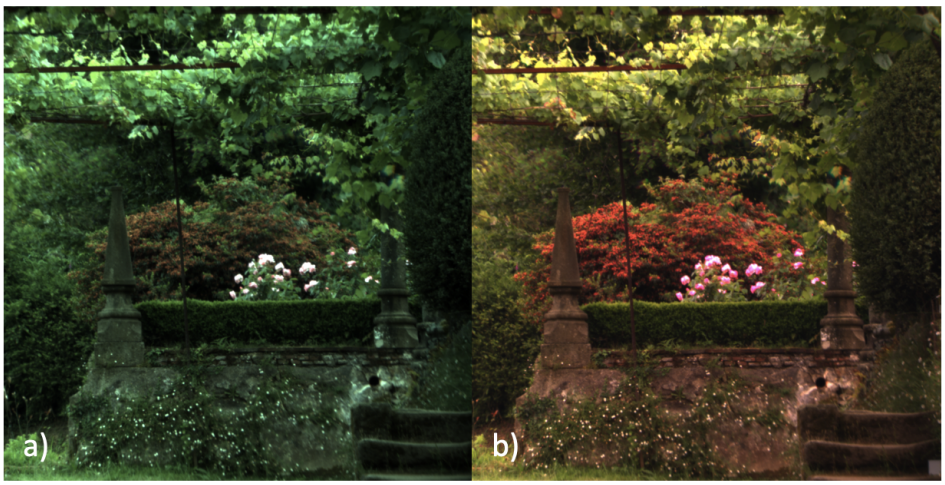


**Figure 1.1:** *Illustration of the colour correction process. The images are generated from David Foster's hyperspectral reflectance dataset (Foster et al., 2006) with Nikon D5100 camera responses and D65 illumination. The left image demonstrates the RAW RGB image, while the right image represents the colour-corrected sRGB image.*

To address this, the camera pipelines have a process called *colour correction*, which transforms device-specific RGB values into a standard colour space that more closely aligns with human vision, such as CIE XYZ or sRGB. This transformation is non-trivial; establishing an accurate mapping between the camera's unique signal space and these universal standards often requires complex non-linear optimizations. This process is illustrated in Figure 1.1, where the left-hand image shows the raw, camera-specific RGB data, and the right-hand image demonstrates the corrected version in the sRGB colour space.

Accurate colour reproduction is critical for various applications. However, specific fields, such as food processing, environmental monitoring, medical diagnostics, or cultural heritage preservation, often require not just accurate colours but also detailed spectral information. For instance, identifying the freshness of produce via chlorophyll absorption, detecting oxygenation levels in blood, or estimating the precise spectral power distribution of ambient light for robust colour constancy requires resolving specific wavelengths. This requirement highlights the second major limitation of RGB cameras: they provide only three broad spectral channels, which fail to capture these fine spectral details. Mathematically, this represents a drastic dimensionality re-

duction where the infinite-dimensional continuous spectrum is projected onto a three-dimensional subspace, causing a massive loss of information.

A common solution is to estimate spectral information from RGB data, namely *spectral reconstruction*, but this is an inherently ill-posed problem. Since multiple spectral distributions can map to the same RGB triplet, there is no unique mathematical solution without the introduction of strong prior knowledge. While recent algorithms have improved spectral recovery, their performance may still fall short for demanding industrial or scientific applications.

To overcome this limitation, hyperspectral and multispectral cameras have been developed. These devices capture a broader range of wavelengths, providing rich spectral data to reveal the detailed 'spectral fingerprint' of materials that can be pivotal in applications requiring high spectral fidelity. However, these systems are not without drawbacks. Due to hardware constraints, hyperspectral and multispectral cameras often suffer from limited spatial resolution, creating a trade-off between spectral and spatial detail. While high-resolution spectral capture is possible using scanning mechanisms or complex beam-splitters, such systems are often bulky, expensive, or unsuitable for dynamic scenes. Conversely, compact single-shot sensors must partition the sensor area to accommodate multiple spectral bands, inherently sacrificing spatial resolution to gain spectral depth.

To address this problem, the concept of *pan-sharpening* has emerged. As shown in Figure 1.2, pan-sharpening involves fusing a low-resolution hyperspectral image with corresponding high-resolution RGB data to produce an output that is both spectrally and spatially detailed. Traditional pan-sharpening techniques use high-resolution panchromatic images to guide the fusion process, while modern approaches leverage high-resolution RGB data instead. By exploiting the high-frequency spatial details (edges and textures) from the RGB image and injecting them into the spectral bands, these advancements aim to enhance both spatial and spectral fidelity, bridging the gap

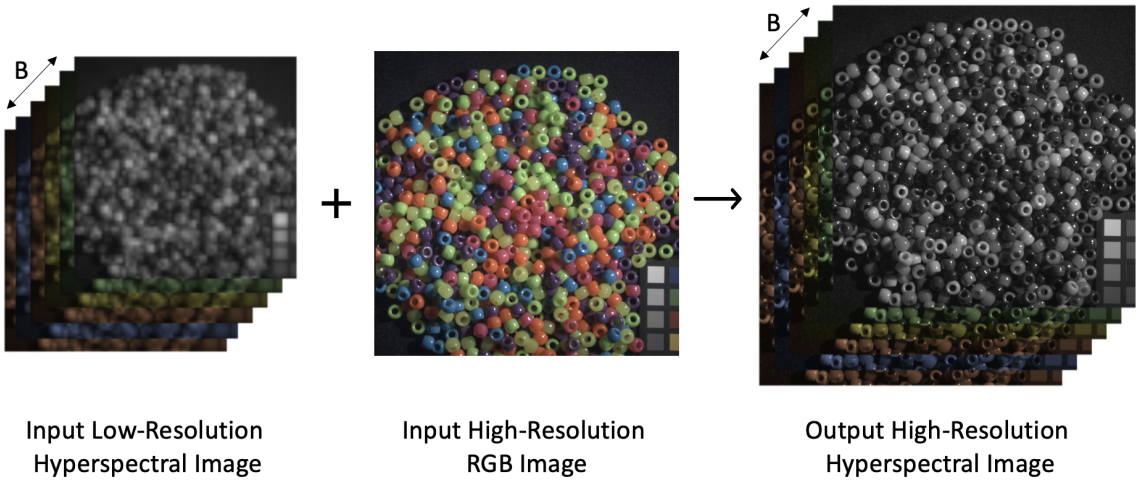


**Figure 1.2:** *An illustration of the RGB-based pan-sharpening process. The images are sourced from the CAVE Dataset (Yasuma et al., 2010). The left image shows the low-resolution hyperspectral image, the centre shows the high-resolution RGB image, and the right shows the pan-sharpened high-resolution hyperspectral image.*

between the capabilities of hyperspectral imaging and the needs of practical applications.

In this thesis, we address both colour correction and spectral recovery topics by effectively combining the strengths of RGB and spectral imaging systems and attempting to mitigate their limitations by incorporating physical constraints. Rather than treating these problems as purely statistical mappings, we ground our solutions in the physics of image formation, ensuring that our computational recoveries remain consistent with the sensor's actual response mechanisms.

## 1.1 Contributions

The contributions of this research are as follows:

**Advancing Colour Correction Algorithms:** We begin by reviewing existing colour correction methods for RGB cameras, comparing traditional regression techniques with neural network (NN) solutions. Our investigation reveals that while NNs offer improvements over simple linear regression, advanced polynomial and root-polynomial regressions outperform current NN approaches in colour accuracy. This comparative analysis highlights that higher

model complexity does not always yield better generalization in colourimetry. Recognising the impact of exposure variations on colour correction - a factor often overlooked in standard fixed-exposure datasets - we introduce an exposure-invariant neural network model that maintains performance across different light level conditions. This capability is particularly vital for concrete applications such as autonomous driving and mobile photography, where lighting environments and exposure parameters fluctuate dynamically. By ensuring consistent colour fidelity regardless of signal magnitude, our research bridges the gap between laboratory-constrained colourimetry and the reliability required in uncontrolled real-world environments.

**Enhancing Spectral Recovery with Physical Constraints:** The Matrix-R theory states that any spectrum can be decomposed into a fundamental metamer, directly detectable by the camera, and a metameric black component, which is invisible to the camera sensors. Despite this knowledge, existing spectral reconstruction and pan-sharpening algorithms often yield errors in the fundamental metamer, effectively 'hallucinating' spectra that are physically impossible given the observed RGB values. The Matrix-R post-processing method is proposed to solve this problem for spectral reconstruction methods. Here we generalise this framework and make it suitable for all spectral recovery attempts (not only for spectral reconstruction but also for pan-sharpening). In addition to its use in hyperspectral imaging, we also demonstrate how to adapt it for use with multispectral cameras. Extending this approach, we show that representing spectra in lower-dimensional spaces can further improve recovery accuracy.

**Developing a Novel Pan-sharpening Post-Correction Technique:** We introduce a least-squares post-processing solution for pan-sharpening that upsamples low-resolution hyperspectral pixels and fuses them with high-resolution RGB data. We mathematically prove that this method inherently possesses the Matrix-R property, guaranteeing that the fused spectra have the correct fundamental metamer. Our technique also achieves competitive perfor-

mance compared to state-of-the-art deep learning solutions when it is used standalone, offering a more computationally efficient alternative suitable for resource-constrained environments.

**Improved Colour Correction with Multispectral Capture:** In scenarios where both high-resolution RGB and corresponding low-resolution multispectral data are available, a typical setup in pan-sharpening, we explore a novel colour correction method. This time, we regress the fused data (combination of RGB and upsampled spectral data) to the standard colour space through linear regression, and achieve superior performance. This approach leverages the complementary information from both cameras, addressing the limitations inherent in single-camera systems.

This thesis holds significant value as it aims to enhance the performance of multiple imaging systems - RGB, multispectral, and hyperspectral - together in a coupled way in both colour correction and spectral recovery fields. By addressing the fundamental issues in these fields, we contribute solutions that rely on physical constraints. The methods and insights presented here aim to enhance future imaging technologies, contributing to more accurate and effective ways of capturing and reproducing the visual world.

## 1.2 Thesis Outline

This thesis is structured as follows:

**Chapter 1** introduces the challenges in digital colour imaging and sets the stage for the subsequent discussions on colour correction and pan-sharpening. The chapter also includes a summary of the contributions and the thesis outline.

In **Chapter 2**, we set the scene with the theoretical background of colour perception and image formation, explaining how cameras and the HVS capture colour. We review the related linear algebra concepts that describe image formation and provide an overview of camera processing pipelines and the key

approaches in the literature.

**Chapter 3** focuses on colour correction algorithms. We critically evaluate various approaches, including linear regression, polynomial methods, and neural networks, discussing their strengths and limitations. We introduce our exposure-invariant neural network model and compare its performance against existing methods.

**Chapter 4** explores spectral reconstruction and pan-sharpening techniques. We revisit the Matrix-R algorithm and demonstrate how incorporating physical constraints can enhance the accuracy of spectral recovery. We propose the generalised version of the Matrix-R post-processing framework and discuss its impact on existing algorithms.

**Chapter 5** presents our novel pan-sharpening post-correction method. We detail the development of the least-squares solution, its inherent Matrix-R property, and its advantages over current approaches.

**Chapter 6** investigates the dual-camera solution for colour correction. We introduce integrating multispectral data with RGB images and achieving superior colour correction performance through linear regression.

**Chapter 7** concludes the thesis, summarising our contributions in the fields of colour correction and spectral recovery, and outlining the limitations and potential future directions for research in physics-based computational photography.

# Chapter 2

# Background and Literature Review

This chapter provides a review of the fundamental concepts, established technologies, and relevant literature that form the basis for the research presented in this thesis. The aim is to equip the reader with the necessary theoretical and practical understanding of colour science, digital imaging systems, colour correction methodologies, and spectral recovery techniques. We begin by exploring the fundamental aspects of the Human Visual System (HVS) and then transition to the mechanisms of digital image capture, detailing sensor technologies, colour filter arrays, and the distinct characteristics of RGB, multispectral, and hyperspectral cameras. This includes a discussion of the physical factors influencing image formation: illuminants and surface reflectances. Subsequently, we examine the standard colour spaces and perceptual difference metrics essential for colour communication and evaluation. A significant portion of the chapter is dedicated to reviewing the digital camera processing pipeline and the critical task of *colour correction*. Following this, we explore the domain of *spectral recovery*, encompassing the challenges and algorithms for both spectral reconstruction and pan-sharpening techniques. Finally, we provide a detailed exposition of the *Matrix-R framework*, exploring its theoretical underpinnings related to spectral ambiguity and its role in developing

physically-constrained recovery algorithms. This broad background serves to contextualise the specific problems addressed and the novel solutions proposed in the subsequent methodology chapters (Chapters 3-6).

## 2.1 Human Visual System

A foundational understanding of how humans perceive colour is indispensable for designing imaging systems and algorithms intended to capture and reproduce visually meaningful information. The process of human vision begins when light interacts with the eye, specifically the retina, which converts this light energy into neural signals. The journey of light into the eye, passing through the pupil and lens to be projected onto the retina, initiates phototransduction, the conversion of photons into electrochemical signals by millions of photoreceptor cells (Dowling, 1987). These photoreceptors are primarily of two types:

- **Rods:** These photoreceptors are characterised by their high sensitivity to light, enabling vision in low-illumination (scotopic) environments. While numerous, rods become saturated in brighter (photopic) conditions and do not play a significant role in colour discrimination, primarily contributing to achromatic vision (Lamb, 2016).

- **Cones:** Functioning optimally under well-lit (photopic) conditions, cones are responsible for colour perception and the high spatial acuity associated with central vision, particularly within the fovea. Humans typically possess three types of cones, each containing a distinct photopigment that confers differential sensitivity to various parts of the visible light spectrum (Lamb, 2016).

The existence of these three cone classes underpins the trichromatic theory of colour vision, a concept initially postulated by Thomas Young (Young, 1802) and further developed by James Clerk Maxwell (Maxwell, 1857). The distinct

spectral absorption profiles of the L-cones (peak sensitivity ≈ 560-580 nm, primarily detecting longer wavelengths), M-cones (peak sensitivity ≈ 530-545 nm, for medium wavelengths), and S-cones (peak sensitivity ≈ 420-440 nm, for shorter wavelengths), as shown in Figure 2.1, enable the visual system to differentiate between colours (Stockman et al., 1993).

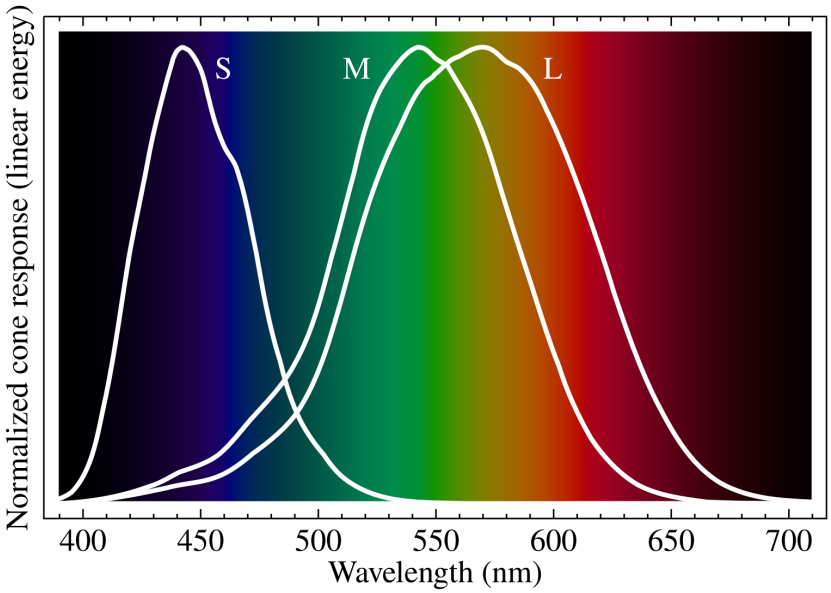


**Figure 2.1:** *The normalized spectral response of the cone cells. This image is taken from https://www.wikiwand.com/en/articles/cone_cell.*

The response of each cone type to an incident spectral power distribution (SPD) $E(\lambda)$ across the visible spectrum $\omega$ (typically 400 nm to 700 nm) is determined by integrating the product of $E(\lambda)$ and the respective cone fundamental ($\bar{l}(\lambda)$, $\bar{m}(\lambda)$, or $\bar{s}(\lambda)$):

$$L = \int_{\omega} E(\lambda)\bar{l}(\lambda)d\lambda, \quad M = \int_{\omega} E(\lambda)\bar{m}(\lambda)d\lambda, \quad S = \int_{\omega} E(\lambda)\bar{s}(\lambda)d\lambda. \tag{2.1}$$

The resultant $[L, M, S]^T$ triplet forms the initial physiological encoding of the colour stimulus. For computational purposes, these continuous functions are discretised. If $\underline{e}$ represents the $n$-dimensional sampled vector of $E(\lambda)$, and $\mathbf{H}$ is the $n \times 3$ matrix containing the discretised cone fundamentals, the HVS response vector $\underline{c}$ can be expressed as $\underline{c} = \mathbf{H}^T\underline{e}$.

While trichromacy explains the initial photoreceptor stage, subsequent

neural processing within the retina and higher visual centres transforms these LMS signals. Ewald Hering's opponent-process theory proposed that colour information is further encoded into opponent channels: Red vs. Green, Blue vs. Yellow, and an achromatic White vs. Black (luminance) channel (Hering, 1964). The theory, which was verified by later studies (Hurvich and Jameson, 1957; De Valois et al., 1966), explains phenomena such as negative afterimages and the perceptual impossibility of certain colour combinations (e.g., reddish-green). Modern theories of colour vision integrate both trichromatic capture and opponent processing as sequential stages in constructing our perception of colour and its various attributes, including hue, saturation, brightness, and more complex appearance effects like colour constancy (Fairchild, 2013).

### 2.1.1 Standardized Colour Specification

To establish an objective and universally understood framework for colour measurement and communication, bridging the gap from individual, subjective human perception, standardised systems have been thoroughly developed. The Commission Internationale de l'Éclairage (CIE) has been the foremost international authority in defining these standards (Smith and Guild, 1931).

#### Colour Matching Experiments

The quantitative foundation of modern colorimetry rests on psychophysical colour matching experiments, with seminal work conducted by W. D. Wright (Wright, 1929) and J. Guild (Guild, 1931). In these experiments, an observer visually matches a monochromatic test light at each wavelength $\lambda$ by additively mixing adjustable amounts of three fixed primary lights (e.g., Red, Green, and Blue) (Fairman et al., 1997). The set of three intensity values required to match the test light across the entire visible spectrum constitutes the Colour Matching Functions (CMFs) (Fairman et al., 1997). A critical finding from this work was that for certain test colours, particularly in the blue-green region, a

direct match was impossible. To achieve a visual match, one of the primaries (often red) had to be added to the test light's field, which is mathematically represented by negative values in the corresponding CMF (Fairman et al., 1997).

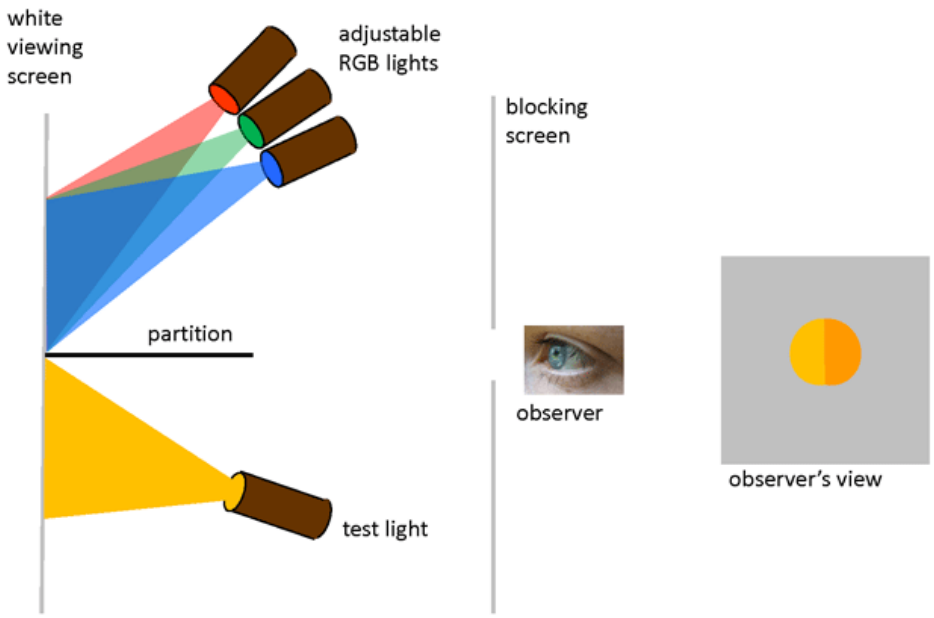


**Figure 2.2:** *The illustration of the colour matching experiments. An observer adjusts the intensities of three fixed monochromatic primary lights, Red (700 nm), Green (546.1 nm), and Blue (435.8 nm), until their additive mixture visually matches a given test light. This image is taken from https://www.oceanopticsbook.info/view/photometry-and-visibility/chromaticity.*

### The CIE 1931 Standard Colourimetric System

Leveraging the extensive data from Wright, Guild, and others, the CIE formally defined the CIE 1931 Standard Colourimetric Observer in 1931 (Smith and Guild, 1931). This standard observer represents the colour matching characteristics of an average human for a 2° angular field of view, chosen to isolate foveal vision - the perception mediated by the fovea, a small region in the retina densely packed with cone cells responsible for sharp, detailed colour vision.

To create a mathematically robust and universally applicable system, and eliminate the negative values present in experimental RGB CMFs, the CIE defined a new set of CMFs, $\bar{x}(\lambda)$, $\bar{y}(\lambda)$, and $\bar{z}(\lambda)$, through a linear transformation

of the averaged experimental RGB CMFs (Fairman et al., 1997). These new CMFs correspond to a set of non-physical, imaginary primaries denoted X, Y, and Z. This transformation was specifically designed to ensure several key properties:

- All values of $\bar{x}(\lambda)$, $\bar{y}(\lambda)$, and $\bar{z}(\lambda)$ are non-negative across the entire visible spectrum.
- The $\bar{y}(\lambda)$ function was defined to be identical to the CIE 1924 photopic luminous efficiency function. This ensures that the Y tristimulus value is directly proportional to the luminance of the colour stimulus, integrating photometry with colourimetry (Smith and Pokorny, 1975).
- For an equal energy spectrum (a theoretical spectrum with constant power at all wavelengths), the X, Y, and Z tristimulus values are equal.

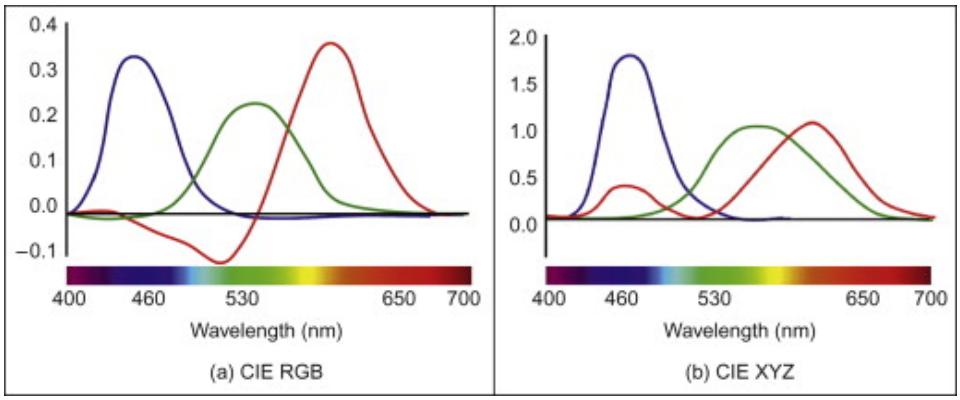


**Figure 2.3:** *The comparison of the CIE RGB and CIE XYZ colour matching functions. This image is taken from https://www.sciencedirect.com/topics/engineering/color-matching-function.*

The XYZ tristimulus values for any given spectral power distribution $E(\lambda)$ are then calculated by integrating the product of $E(\lambda)$ with each of the XYZ CMFs over the visible spectrum $\omega$ (typically 400–700 nm):

$$X = k \int_{\omega} E(\lambda)\bar{x}(\lambda)d\lambda, \quad Y = k \int_{\omega} E(\lambda)\bar{y}(\lambda)d\lambda, \quad Z = k \int_{\omega} E(\lambda)\bar{z}(\lambda)d\lambda. \tag{2.2}$$

The normalization constant $k$ is typically defined so that $Y = 100$ for a perfect reflecting diffuser when illuminated by a chosen reference illuminant $I(\lambda)$; a common formulation is $k = 100/\int_\omega I(\lambda)\bar{y}(\lambda)d\lambda$. When working with discretized spectra, where $\underline{e}$ is the $31\times1$ sampled vector of $E(\lambda)$ and $\mathbf{X}$ is the $31\times3$ matrix of the sampled CMFs $[\bar{\mathbf{x}}, \bar{\mathbf{y}}, \bar{\mathbf{z}}]$, the tristimulus vector $\underline{x} = [X, Y, Z]^T$ is given by $\underline{x} = k\mathbf{X}^T\underline{e}$.

For situations involving larger visual stimuli (viewing angles greater than 4°), where parafoveal cones and rod activity play a more significant role and the effective density of macular pigment differs, the CIE defined the 1964 Standard Supplementary Colourimetric Observer (the 10° observer). This standard utilizes a distinct set of CMFs, $\bar{x}_{10}(\lambda), \bar{y}_{10}(\lambda), \bar{z}_{10}(\lambda)$, derived from colour matching experiments with a 10° field (Stiles and Burch, 1959).

### 2.1.2 Physical Components of Image Formation

The spectral radiance signal $\underline{e}$ that constitutes the input to an imaging sensor is determined by the interplay of the incident illumination and the reflective properties of the surfaces within the scene.

#### Illuminants

As discussed in the context of CIE standards, the illuminant's SPD, $I(\lambda)$ describes the distribution of light energy emitted by the source across the spectrum. Natural daylight exhibits significant spectral variation due to factors like solar elevation (time of day, season, latitude), atmospheric conditions (e.g., Rayleigh scattering causing blue skies, Mie scattering from dust, smoke and clouds) (Pastilha and Hurlbert, 2022). Artificial light sources, including incandescent, halogen, fluorescent, and LED lamps, each possess distinct SPDs governed by their respective light generation mechanisms (e.g., thermal radiation, gas discharge, electroluminescence) (Weisbuch, 2018). Characterising and, where necessary, estimating $I(\lambda)$ is crucial for colour-critical applications

such as colour constancy (computational adjustment for illuminant changes) and physically accurate rendering.

### Surface Reflectance

The interaction of light with a surface is characterised by its Surface Spectral Reflectance Factor (SRF), $S(\lambda)$ (discretised as vector $\underline{s}$). For a given illumination and viewing geometry, $S(\lambda)$ defines the ratio of reflected radiant flux to incident radiant flux at each wavelength (Palmer, 2010).

For many common surfaces, $S(\lambda)$ is determined by the selective absorption and scattering of light by pigments, dyes, or the intrinsic molecular structure of the material. The value of $S(\lambda)$ for opaque surfaces generally lies between 0 (a perfect absorber, or blackbody) and 1 (a perfect reflector, or ideal white diffuser). The specific shape of the $S(\lambda)$ curve across the visible spectrum is what imparts an object with its characteristic colour when viewed under a particular illuminant.

It is well established that the spectral reflectances of many natural and man-made objects are typically smooth functions of wavelength, lacking sharp discontinuities (Vrhel et al., 1994). This property is fundamental to the observation that these spectra can often be accurately represented by low-dimensional linear models (Maloney, 1986). Standardised collections of measured SRFs, such as the Munsell chips (Parkkinen et al., 1989), the Macbeth ColorChecker chart (McCamy et al., 1976), and databases containing spectra of natural objects, textiles, and other materials (Barnard et al., 2002), are indispensable resources for research in colour science. The spectral radiance $E(\lambda)$ reflected from such a surface and reaching a sensor is then, for a simple imaging geometry, given by the product $E(\lambda) = I(\lambda)S(\lambda)$.

## 2.2 Digital Imaging Systems and Processing Pipeline

Understanding the journey of light from the scene to a final digital image requires an appreciation of the underlying hardware components and the sequence of processing stages involved. These elements collectively define the characteristics and limitations of the captured image data, providing crucial context for the development of colour correction and spectral recovery algorithms.

### 2.2.1 Image Sensor Technologies

At the heart of every digital camera lies an image sensor, a solid-state device that converts incident photons into an electrical signal, typically charge or voltage. The two dominant technologies for image sensors have been Charge-Coupled Devices (CCDs) and Complementary Metal-Oxide-Semiconductor (CMOS) sensors.

- **Charge-Coupled Device (CCD):** Originating from Bell Labs in 1969 with the work of Boyle and Smith (Boyle and Smith, 1970), Charge-Coupled Devices (CCDs) operate by accumulating photogenerated electrical charge in discrete packets within individual pixel wells during an exposure period. Following exposure, these charge packets are systematically and sequentially transferred across the sensor, often row by row, to a common output structure where the charge is converted into a voltage signal, typically then amplified and digitised by an off-chip Analogue-to-Digital Converter (ADC).

- **Complementary Metal-Oxide-Semiconductor (CMOS) Active Pixel Sensor (APS):** CMOS image sensor technology, particularly the Active Pixel Sensor (APS) design where each pixel contains its own amplifier (and sometimes other circuitry like reset transistors or correlated

double sampling logic), gained traction in the 1990s and has since become dominant in most imaging applications (El Gamal and Eltoukhy, 2005).

### 2.2.2 Colour Filter Arrays (CFAs) and Demosaicing

Most digital cameras employ a Colour Filter Array (CFA) to capture colour information. A CFA is a mosaic of microscopic optical filters placed directly over the sensor's pixels, making each pixel sensitive to only a specific portion of the visible spectrum. Typically, these filters are designed to transmit broad regions of light that correspond to what we perceive as red (long wavelengths), green (medium wavelengths), and blue (short wavelengths).

The most widely adopted CFA is the Bayer filter pattern, invented by Bryce E. Bayer at Eastman Kodak in the 1970s (Bayer, 1976). As shown in Figure 2.4, this pattern typically consists of a repeating $2 \times 2$ unit cell containing one red filter, one blue filter, and two green filters (often arranged as RGGB or similar permutations). The higher proportion of green filters (50%) compared to red (25%) and blue (25%) is designed to mimic the HVS's greater sensitivity to green light, which contributes significantly to perceived luminance and spatial detail (Lukac and Plataniotis, 2005).

Since each pixel under a CFA records only one colour channel, the remaining two missing colour values at that pixel location must be estimated from the data captured by neighboring pixels with different colour filters. This process of reconstructing a full-colour image from the spatially undersampled CFA output is known as demosaicing or CFA interpolation (Li et al., 2008). Numerous algorithms have been developed for the problem such as simple interpolation (Malvar et al., 2004), frequency domain (Dubois, 2005), learning-based methods (Kokkinos and Lefkimmiatis, 2019).

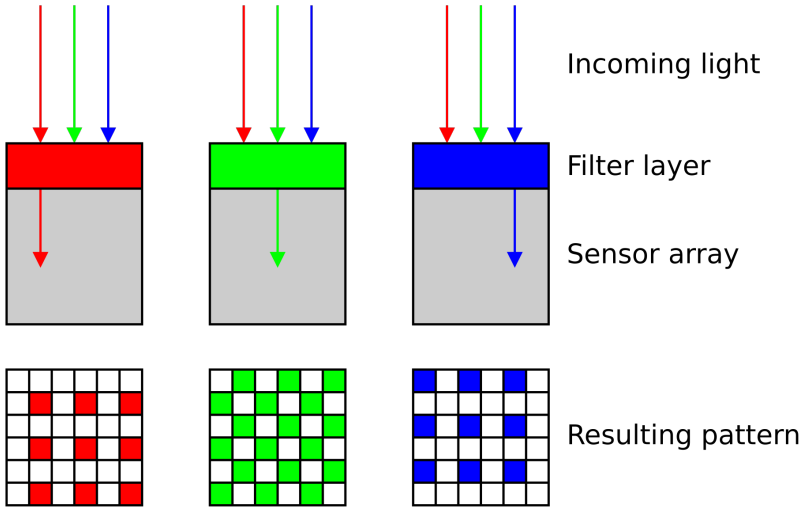


**Figure 2.4:** *The common Bayer filter array (CFA) pattern, showing the repeating* $2 \times 2$ *mosaic composed of 50% green, 25% red, and 25% blue filters. This image is taken from https://en.wikipedia.org/wiki/Bayer_filter.*

### 2.2.3 Multispectral and Hyperspectral Capture

Beyond the three broad spectral channels captured by conventional RGB cameras, Multispectral and Hyperspectral imaging systems are designed to acquire more detailed spectral information from a scene. The primary distinction lies in the number and nature of the spectral bands they capture:

- **Multispectral Imaging (MSI):** Typically captures a discrete, relatively small number of spectral bands (commonly ranging from 4 to about 20 bands). These bands can vary in width; some MSI systems utilize relatively wide-band filters, potentially with significant spectral overlap between adjacent bands, similar in characteristic (though more numerous) to RGB sensor sensitivities. Other MSI sensors may employ narrower bandpass filters for more precise spectral sampling. Technologies include:
  - *Filter Wheels:* This approach involves a monochrome camera equipped with a mechanically rotated wheel holding multiple (typically 5 to 20) different bandpass filters. Images are captured sequentially as each filter is positioned in front of the sensor (Mansouri et al., 2005).

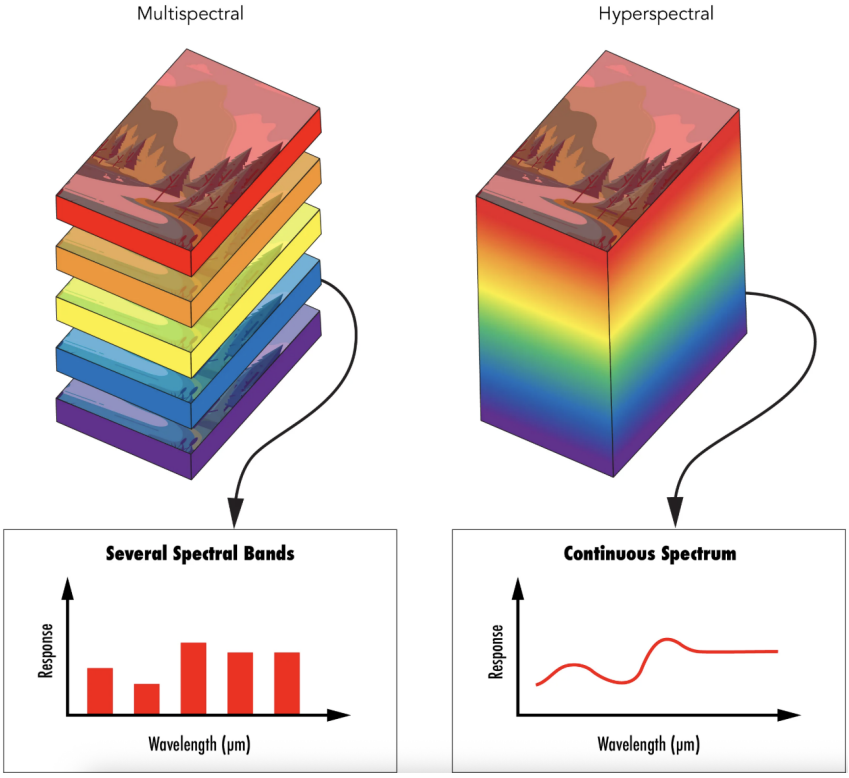


**Figure 2.5:** *Conceptual illustration comparing Multispectral and Hyperspectral imaging. Multispectral systems capture several, usually wider spectral bands (note that while depicted here as discrete for clarity, these bands can be broad and overlapping, similar to RGB sensors). Hyperspectral systems, in contrast, capture data in many narrow, contiguous bands, forming a nearly continuous spectrum for each pixel. This image is taken from https://www.edmundoptics.co.uk/knowledge-center/application-notes/imaging/hyperspectral-and-multispectral-imaging/.*

This provides flexibility in selecting spectral bands but is inherently slow, limiting its use to static scenes and stable illumination.

– *Snapshot Capture:* Analogous to CFAs for RGB imaging, Spectral Filter Arrays (SFAs) are mosaics of different spectral filters (more than three types) tiled over a single sensor array. This allows for "snapshot" multispectral capture, where all selected bands are acquired simultaneously in a spatially interleaved manner (Monno et al., 2015). The captured raw image then requires a specialized demosaicing algorithm to reconstruct the full multispectral data cube for each pixel. This approach trades spatial resolution per band for snapshot capability.

- **Hyperspectral Imaging (HSI):** Aims to acquire data in many (of-

ten hundreds) narrow, contiguous (or nearly contiguous) spectral bands across a defined wavelength range (e.g., the visible and near-infrared spectrum). This dense sampling results in a detailed, almost continuous spectrum for each pixel, forming a 3D hyperspectral data cube (two spatial dimensions, one spectral dimension). This high spectral resolution allows for fine discrimination between materials based on subtle differences in their spectral signatures. The acquisition of such detailed hyperspectral data cubes has traditionally been achieved through three main scanning methodologies:

- *Line-Scan (Pushbroom Imagers):* An entrance slit isolates light from a single line in the scene. This line of light is then dispersed by the grating/prism, projecting its spectrum along one dimension of the 2D sensor array (Goetz et al., 1985). The other sensor dimension corresponds to the spatial dimension along the slit. To form a 2D spatial image with spectral information at each pixel (a hyperspectral cube), relative motion between the sensor and the scene is required (e.g., an aircraft or satellite moving forward, or an object moving on a conveyor belt).
- *Point-Scan (Whiskbroom Imagers):* Collects light from a single point in the scene at a time. This light is dispersed, and its spectrum is measured by a 1D array of detectors (or a single detector scanned across wavelengths). A 2D spatial image is built up by mechanically scanning the sensor's field of view across the scene in two dimensions. This method is slower than pushbroom scanning.
- *Snapshot Hyperspectral Imaging:* In contrast to scanning methods, snapshot techniques are designed to capture the entire 3D hyperspectral data cube in a single exposure, enabling the capture of video and dynamic scenes. These often rely on computational imaging principles. A prominent example is Coded Aperture Snapshot

Spectral Imaging (CASSI), where a coded aperture (a patterned mask) is placed in the optical path before a dispersive element (Gehm et al., 2007). The resulting 2D sensor image is a compressed, overlapping projection of the 3D data cube. A reconstruction algorithm, often based on compressive sensing, is then required to computationally recover the full hyperspectral cube from this single measurement (Gehm et al., 2007).

The choice of a specific MSI or HSI technology is highly dependent on the application's unique demands for spectral detail (number of bands, bandwidth, contiguity), spatial resolution, temporal resolution (for dynamic scenes), sensitivity, and practical constraints such as system cost, size, weight, and power.

### 2.2.4 Image Signal Processor (ISP) Pipeline

The raw data acquired by a digital camera's image sensor is typically not directly viewable or suitable for most applications. It must undergo a sequence of processing steps, collectively known as the Image Signal Processor (ISP) pipeline, to transform it into a visually pleasing and standardized image format (e.g., JPEG). While the exact architecture and order of operations can vary significantly between manufacturers and device types (e.g., DSLRs, smartphones, industrial cameras), a general ISP pipeline incorporates several key stages. Understanding these stages is crucial as they influence the final colour and quality of the image, and the RAW data (if available) often bypasses many of these.

1. **Preprocessing Stage:** Initial corrections are applied directly to the raw sensor data to address inherent limitations of the imaging system. This typically involves compensating for defective pixels by interpolating their values from neighbours (El-Yamany, 2017), removing sensor-generated thermal noise (dark current) by subtracting a 'dark frame' (Healey and

Kondepudy, 2002), and correcting for optical effects such as lens shading (vignetting) (Zheng et al., 2008). These fundamental steps are important for producing a clean and accurate representation of the captured scene before subsequent processing.

2. **White Balance (WB) / Chromatic Adaptation:** This crucial stage aims to remove unwanted colour casts caused by the spectral properties of the scene illuminant, ensuring that objects perceived as white by humans are rendered as white (or neutral grey) in the image. This process emulates the chromatic adaptation capabilities of the HVS (Stockman et al., 2010).

   Accurate WB typically involves two steps: (1) estimating the colour of the scene illuminant, and (2) applying a transformation to correct the image data. The transformation often follows the Von Kries adaptation model (Fairchild, 2020), which assumes independent scaling of the camera's captured RGB response vector $[R_c, G_c, B_c]^T$ as:

$$\begin{bmatrix} R_a \\ G_a \\ B_a \end{bmatrix} = \begin{bmatrix} g_R & 0 & 0 \\ 0 & g_G & 0 \\ 0 & 0 & g_B \end{bmatrix} \begin{bmatrix} R_c \\ G_c \\ B_c \end{bmatrix} \tag{2.3}$$

   where $[R_a, G_a, B_a]^T$ is the desired response vector adapted to a reference illuminant (e.g., appearing as if captured under daylight). The $3 \times 3$ diagonal matrix contains the adaptation gains $g_R, g_G, g_B$. These scalar gains are calculated based on the estimated chromaticity of the scene illuminant relative to the target reference illuminant. The challenge in automatic white balancing then largely becomes accurately estimating the source white point from the image data. Common illuminant estimation algorithms include:

   - *Gray World:* Assumes that the average reflectance in a scene is

spectrally neutral (achromatic). The gains are set to equalize the average R, G, and B values of the image (Buchsbaum, 1980). Simple, but fails for scenes dominated by specific colours.

- *White Patch / Max RGB:* Assumes that the brightest pixel(s) or region in an image correspond to a perfectly reflecting (white) surface under the scene illuminant. The gains are derived from these maximum R, G, B values (Cardei and Funt, 1999). Susceptible to clipped specular highlights or intensely coloured bright objects. Although specular reflections theoretically represent the illuminant's colour, they frequently exceed the sensor's dynamic range; this saturation distorts the recorded RGB ratios, leading to incorrect gain estimation (Gijsenij et al., 2011).
- *Shades of Gray:* A generalization that uses the Minkowski p-norm of the image data for averaging, encompassing Gray World (p=1) and Max RGB (p=$\infty$) as special cases (Finlayson and Trezzi, 2004).
- *Learning-Based Methods:* Modern approaches employ machine learning, often using deep neural networks, to estimate illuminant chromaticity from image features, trained on large datasets of images with known ground-truth illumination (Afifi et al., 2022; Buzzelli and Bianco, 2024). These methods often achieve state-of-the-art performance.

3. **Demosaicing:** If a CFA is used, this step reconstructs a full-colour image by interpolating the missing colour values at each pixel location, as detailed in Section 2.2.2.

4. **Colour Correction / Colour Space Transformation:** After demosaicing and white balance, the device-specific RGB values are transformed into a standard, device-independent colour space like CIE XYZ, or directly into a target output/working space such as sRGB. This step,

detailed in Section 2.4, is crucial for ensuring consistent and accurate colour reproduction.

5. **Gamma Correction:** A non-linear mapping is applied to the linear light intensity values. For display-referred pipelines, this typically involves applying a power function with an exponent around $1/\gamma$ (e.g., 1/2.2 for sRGB) to encode the image data in a way that, when displayed on a device with a typical gamma of $\gamma$, results in a linear system response. This also compresses tonal range in a way that roughly aligns with perceptual lightness sensitivity (Bull and Zhang, 2014).

6. **Post-processing / Image Enhancement:** Various operations may be applied to improve the image's visual quality or aesthetic appeal.

    - *Tone Mapping:* Compresses the dynamic range of the scene, particularly important when mapping HDR sensor data to the LDR capabilities of typical displays or file formats. Tone mapping operators apply a non-linear curve to image content to preserve detail in both highlights and shadows (Kao et al., 2007).

    - *Sharpening / Edge Enhancement:* Enhances perceived sharpness by highlighting edges and fine details, often using techniques like unsharp masking or deconvolution (Singh et al., 2014).

    - *Noise Reduction:* Applies spatial filters to reduce random noise introduced by the sensor and electronics, especially prominent at high ISO settings or in low-light conditions (Tomasi and Manduchi, 1998).

    - *Colour Enhancement:* Adjustments to overall saturation, vibrancy, or specific hues based on predefined "picture styles" or user preferences, often aiming for a more "pleasing" rather than strictly accurate rendition.

The specific implementation and order of these ISP stages significantly influence the final image output and are often proprietary secrets of camera manufacturers. Access to RAW data allows bypassing much of this proprietary pipeline, offering more control for custom colour correction and processing.

## 2.3 Colour Spaces and Colour Difference Metrics

Representing and quantifying colour accurately requires standardized mathematical models.

### 2.3.1 Key Colour Spaces

- **CIE XYZ (1931):** The fundamental device-independent reference space based on the standard observer CMFs. All perceivable colours can be represented by positive XYZ values. It forms the hub for transformations between other spaces (Smith and Guild, 1931).

- **CIELAB (1976):** This colour space, recommended by the CIE in 1976, is a non-linear transformation of the CIE XYZ space designed to be more perceptually uniform, meaning that Euclidean distances within it correspond more closely to perceived colour differences (Hunt and Pointer, 2011). The transformation from XYZ tristimulus values $(X, Y, Z)$ to CIELAB coordinates $(L^*, a^*, b^*)$ is defined relative to the XYZ values of a specified reference white point $(X_n, Y_n, Z_n)$ as follows:

$$\begin{aligned} L^* &= 116 f(Y/Y_n) - 16 \\ a^* &= 500[f(X/X_n) - f(Y/Y_n)] \\ b^* &= 200[f(Y/Y_n) - f(Z/Z_n)] \end{aligned} \tag{2.4}$$

where the non-linear function $f(t)$ is defined as:

$$f(t) = \begin{cases} t^{1/3} & \text{if } t > \left(\frac{29}{6}\right)^3 \\ \frac{1}{3}\left(\frac{6}{29}\right)^2 t + \frac{4}{29} & \text{otherwise} \end{cases} \tag{2.5}$$

In this system, $L^*$ represents psychometric lightness (ranging from 0 for black to 100 for diffuse white), $a^*$ represents the red-green opponent axis (positive for red, negative for green), and $b^*$ represents the yellow-blue opponent axis (positive for yellow, negative for blue). CIELAB is widely used for colour difference calculation, device characterization, and in various colour management workflows.

- **sRGB:** A standard RGB working and output space for the web, consumer displays, and printers (Anderson et al., 1996). It defines specific R, G, B primary chromaticities, a D65 white point, and a non-linear gamma encoding. Its gamut is relatively limited compared to human vision.

### 2.3.2 Colour Difference Metrics ($\Delta E$)

While colour spaces like CIELAB are designed to be more perceptually uniform than CIE XYZ, quantifying the precise visual difference between two colours requires a specific mathematical formula known as a colour difference metric. The goal of such a metric is to produce a single number, denoted as $\Delta E$, that directly corresponds to the magnitude of the visual difference perceived by an average human observer. Notable examples of these formulas include:

- $\Delta E^*_{ab}$ **(CIE76):** Euclidean distance in CIELAB (Eq. 2.4). Simple baseline, but known limitations in uniformity.

$$\Delta E^*_{ab} = \sqrt{(\Delta L^*)^2 + (\Delta a^*)^2 + (\Delta b^*)^2}. \tag{2.6}$$

- $\Delta E_{00}$ **(CIEDE2000):** It was developed to provide a closer agreement with human perception of small colour differences for surface colours through refined lightness ($\Delta L'$), chroma ($\Delta C'$), and hue ($\Delta H'$) terms, comprehensive weighting functions ($S_L, S_C, S_H$) depending on location in CIELAB space, and a rotation term ($R_T$) for blue region interactions (Sharma et al., 2005).

$$\Delta E_{00} = \sqrt{\left(\frac{\Delta L'}{k_L S_L}\right)^2 + \left(\frac{\Delta C'}{k_C S_C}\right)^2 + \left(\frac{\Delta H'}{k_H S_H}\right)^2 + R_T \left(\frac{\Delta C'}{k_C S_C}\right)\left(\frac{\Delta H'}{k_H S_H}\right)}. \tag{2.7}$$

## 2.4 Colour Correction Methodologies

This section provides a more detailed survey of the key fundamental methods used to perform the core colour correction task. This process involves mapping the camera's raw, device-specific RGB response vector ($\underline{\rho}$) to a standard, device-independent colour representation, most commonly the CIE XYZ tristimulus values ($\underline{x}$).

### 2.4.1 Regression-Based Approaches

These methods model the transformation $f(\underline{\rho}) \approx \underline{x}$ using regression techniques trained on paired data. This training data consists of an $N \times 3$ matrix $\mathbf{P}$, where each row is a camera's measured RGB response vector ($\underline{\rho}_i$), and a corresponding $N \times 3$ matrix $\mathbf{X}$, where each row is the known ground-truth CIE XYZ tristimulus vector ($\underline{x}_i$) for each of the $N$ training samples.

#### Linear Regression

Models $f$ as a $3 \times 3$ matrix $\mathbf{M}$, found as:

$$\arg\min_{\mathbf{M}} \|\mathbf{PM} - \mathbf{X}\|_F \tag{2.8}$$

where $F$ denotes the Frobenius norm (Horn and Johnson, 1990), and the equation can be solved via the Moore-Penrose pseudo-inverse (Penrose, 1955):

$$\mathbf{M} = (\mathbf{P}^T\mathbf{P})^{-1}\mathbf{P}^T\mathbf{X} \tag{2.9}$$

The linear regression method is computationally simple and stable. It is also exposure-invariant, a critical property meaning that if the overall scene brightness changes by a certain factor, the output XYZ values scale by the same factor, preserving the colour appearance (Finlayson et al., 2015). However, its primary limitation is its struggle to accurately model the often non-linear relationship between the sensor responses and the target XYZ values. This non-linearity arises because a camera's spectral sensitivities are not a simple linear transformation of the human eye's sensitivities; if they were (a state known as the Luther-Ives condition (Sharma, 2017)), linear regression would be perfectly accurate. Because this condition is not met by real-world sensors, linear regression can only provide an approximation.

**Polynomial and Root-Polynomial Regression**

These enhance linear regression by first applying a non-linear feature expansion $f_{exp}(\underline{\rho})$ to create a $k$-dimensional feature vector, then applying a linear regression $\mathbf{M}$ $(k \times 3)$ to these features: $\mathbf{X} \approx f_{exp}(\mathbf{P})\mathbf{M}$.

- **Polynomial:** In addition to the original RGB values, it uses the cross and squared terms as input $R, G, B, RG, RB, BG, R^2, B^2, G^2$. Can achieve higher accuracy than linear by fitting non-linearities (Hong et al., 2001). However, the degree and specific terms need careful selection to avoid overfitting. Critically, it is not exposure-invariant, limiting its use in varying lighting conditions (Finlayson et al., 2015).

- **Root-Polynomial:** Uses terms like $R, G, B, \sqrt{RG}, \sqrt{RB}, \sqrt{GB}$ (Finlayson et al., 2015). This specific form ensures the homogeneity of the expansion, $f(c\underline{\rho}) = cf(\underline{\rho})$, resulting in an exposure-invariant colour cor-

rection while still capturing non-linear interactions between channels. As shown in Chapter 3, this often provides the best trade-off among classical regression methods.

### 2.4.2 Neural Network Approaches

Neural Networks (NNs) provide a data-driven methodology for colour correction, capable of learning complex, non-linear transformations directly from paired datasets of camera RGB responses and target colour values (MacDonald and Mayer, 2021). Common architectures like Multi-Layer Perceptrons (MLPs) typically process individual pixel information for colour mapping (Hwang et al., 2015; Adkins et al., 1993), whereas Convolutional Neural Networks (CNNs) can incorporate spatial context from image patches, which may be advantageous for certain correction tasks or joint image enhancements (Vlachos and Skarlatos, 2021; Afifi and Brown, 2020).

The training of these networks relies on optimizing their parameters over large datasets, frequently using loss functions based on perceptual colour difference metrics (e.g., $\Delta E_{00}$) to better align the outcomes with human visual assessment (MacDonald and Mayer, 2021). While NNs have a promising future, their practical performance is closely tied to the quality and scope of the training data. Challenges remain in data requirements, computational cost, generalization to unseen cameras/conditions, interpretability, and ensuring physical properties like exposure invariance (the focus of Chapter 3).

## 2.5 Spectral Recovery: Spectral Reconstruction and Pan-sharpening

Spectral recovery aims to estimate high-dimensional spectral signals from spectrally or spatially lower-dimensional measurements, addressing limitations of standard cameras.

### 2.5.1 Spectral Reconstruction (SR) from RGB Data

Spectral Reconstruction (SR) from RGB data aims to estimate a high-dimensional spectral signal, $\underline{e} \in \mathbb{R}^n$ (e.g., $n = 31$), from a low-dimensional camera RGB response, $\underline{\rho} \in \mathbb{R}^3$. This task is of significant practical interest because it allows for the potential recovery of rich spectral information using ubiquitous and cost-effective RGB cameras, thereby avoiding the need for specialised and often expensive hyperspectral or multispectral imaging systems. Furthermore, SR from RGB inherently bypasses challenges related to the spatial and temporal alignment that can arise when fusing data from multiple cameras or sequentially scanned spectral systems.

However, SR from RGB is a fundamentally ill-posed inverse problem (Zhang et al., 2022). The mapping from an $n$-dimensional spectrum to a 3-dimensional RGB response, $\underline{\rho} = \mathbf{Q}^T\underline{e}$ (where $\mathbf{Q}$ is the $n \times 3$ camera spectral sensitivity matrix), involves a significant loss of information. This leads to the phenomenon of metamerism, where multiple, spectrally distinct signals $\underline{e}_1 \neq \underline{e}_2$ can produce the identical RGB response $\underline{\rho}$ (Cohen and Kappauf, 1982). This inherent information loss means that unique spectral recovery is impossible without incorporating additional prior knowledge or learned statistical models, and the accuracy of any SR method is thus inherently constrained. Various approaches have been developed to tackle this challenge, broadly categorised below.

- **Linear Model-Based Methods:** These assume that spectra $\underline{e}$ can be approximated by a low-dimensional linear model, $\underline{e} \approx \mathbf{B}\underline{\alpha}$, where $\mathbf{B}$ is a $n \times m$ matrix of basis spectra (often derived via Characteristic Vector Analysis (CVA) (Maloney, 1986) from a spectral database, with $m \ll n$) and $\underline{\alpha}$ is a $m \times 1$ coefficient vector. The task is to estimate $\underline{\alpha}$ from $\underline{\rho}$ (where $\underline{\rho} \approx (\mathbf{Q}^T\mathbf{B})\underline{\alpha}$), typically via pseudo-inverse methods. While computationally simple, their accuracy depends heavily on the suitability of the fixed basis $\mathbf{B}$ for the target spectra (Maloney, 1986).

- **Regression Methods:** These techniques learn a direct mapping function from RGB responses $\underline{\rho}$ to the full spectrum $\underline{e}$. Methods including linear regression, polynomial and root-polynomial regression, discussed in the context of colour correction (Section 2.4), can be adapted for SR by changing the target output from XYZ values to the $n$-dimensional spectrum. Polynomial and root-polynomial methods expand $\underline{\rho}$ into higher-order features before linear regression (Connah and Hardeberg, 2005).

- **Dictionary Learning and Sparse Coding Approaches:** These methods operate on the premise that spectral signals can be represented as a sparse linear combination of elementary spectra (atoms) from an over-complete dictionary $\mathbf{D}$ (i.e., $\underline{e} = \mathbf{D}\underline{w}$, where $\underline{w}$ is a sparse coefficient vector). The dictionary $\mathbf{D}$ is often learned from a spectral database using techniques like K-SVD (Lin and Finlayson, 2023). The A+ algorithm, for example, employs locally linear regression within neighbourhoods defined by dictionary atoms (Aeschbacher et al., 2017). Building upon that, the A++ method further refines neighbourhood selection using initial spectral estimates instead of RGB points and has demonstrated improved accuracy for reconstructing hyperspectral images from RGB inputs (Lin and Finlayson, 2023).

- **Neural Network Approaches:** Various neural network architectures are employed for SR, leveraging their ability to learn complex, non-linear mappings from large datasets. Radial Basis Function Networks (RBFNs) map input RGBs to a feature space based on their proximity to a set of learned data centres (prototypes), using radial basis functions (e.g., Gaussians) as activations in a hidden layer. The output spectrum (or its coefficients) is then typically a linear combination of these hidden layer activations. RBFNs can model non-linearities effectively and are often trained in a two-stage process or end-to-end (Nguyen et al., 2014). Multi-Layer Perceptrons (MLPs) are foundational feedforward NNs, with

one or more hidden layers of neurons using non-linear activations (e.g., sigmoid, ReLU), and they can be trained to perform pixel-wise mapping from RGB to spectra (Batinić et al., 2020). More complex deep learning models, especially Convolutional Neural Networks (CNNs) (Shi et al., 2018), Generative Adversarial Networks (GANs) (Alvarez-Gila et al., 2017), and Transformers (Cai et al., 2022), often achieve state-of-the-art SR performance (Arad et al., 2022). These models learn hierarchical spatial-spectral features end-to-end from large datasets of paired RGB and hyperspectral image patches. By processing image patches rather than individual pixels, these deep learning methods can better leverage spatial context to help resolve metameric ambiguities. However, they are typically data-hungry, computationally intensive, can be challenging to generalise, and often lack transparency (Lin and Finlayson, 2020a, 2023).

### 2.5.2 Pan-sharpening Techniques

Pan-sharpening addresses the common trade-off in imaging systems between high spatial resolution and high spectral resolution. It is a fusion process that aims to combine a low-resolution (LR) multispectral (MS) or hyperspectral (HS) image, denoted $\mathbf{L}$, with a co-registered high-resolution (HR) guidance image, $\mathbf{H}$, to produce an HR MS/HS image estimate, $\widehat{\mathbf{C}}$. While traditional pan-sharpening in remote sensing often uses an HR panchromatic (PAN) image as guidance - a single-band, high-resolution grayscale image sensitive to a broad range of visible light wavelengths (Vivone et al., 2014) - modern approaches, and the focus relevant to this thesis, increasingly utilize an HR RGB image as the guidance $\mathbf{H}$ (which we can denote $\mathbf{P}$ for clarity when it is specifically RGB) (Fu et al., 2019). The overarching goal is to generate an image $\widehat{\mathbf{C}}$ that ideally possesses the spatial fidelity of $\mathbf{H}$ and the spectral fidelity of $\mathbf{L}$ (Vivone et al., 2014). This is crucial for applications requiring both fine spatial details and rich spectral information. Prominent examples include precision agriculture, where assessing crop health requires spectral precision at the individual plant

level, and remote sensing for urban planning, which demands the classification of specific surface materials within dense environments. Pan-sharpening methods can be broadly categorised:

- **Component Substitution (CS):** These methods typically upsample the LR MS/HS image $\mathbf{L}$ to the resolution of $\mathbf{H}$. Both $\mathbf{L}_{\text{up}}$ and $\mathbf{H}$ are transformed into a new space where spatial and spectral information are somewhat separated (e.g., Intensity-Hue-Saturation (IHS) space (Tu et al., 2001), or via Principal Component Analysis (PCA) (Chavez et al., 1991)). The component representing spatial detail from $\mathbf{L}_{\text{up}}$ is replaced by the corresponding component from $\mathbf{H}$ (or $\mathbf{H}$ itself, if panchromatic). An inverse transform then yields the pan-sharpened image $\widehat{\mathbf{C}}$. CS techniques are generally simple and computationally efficient. However, they can introduce significant spectral distortions if the replaced component also carries important spectral information or if the spectral range of $\mathbf{H}$ does not well represent the intensity of all MS/HS bands (Vivone et al., 2014).

- **Multiresolution Analysis (MRA):** MRA-based techniques operate on the principle of extracting high-frequency spatial details from the HR guidance image $\mathbf{H}$ and injecting them into the upsampled LR MS/HS image $\mathbf{L}_{\text{up}}$ (Vivone et al., 2014). This is achieved by decomposing $\mathbf{H}$ (and sometimes $\mathbf{L}_{\text{up}}$) into different spatial frequency levels using tools like the Laplacian pyramid (Addesso et al., 2021), or wavelet transforms (Nunez et al., 1999). The spatial detail extracted from $\mathbf{H}$ at each level is then appropriately scaled and added to the corresponding level of $\mathbf{L}_{\text{up}}$. While MRA methods generally achieve a superior balance between enhancing spatial detail and preserving spectral fidelity compared to CS approaches, they may occasionally introduce subtle artefacts such as ringing or aliasing. They can also be more computationally demanding.

- **Model-Based Approaches:** These methods aim to estimate the HR

MS/HS image $\widehat{\mathbf{C}}$ by optimising an objective function. This function typically includes data fidelity terms that model the relationship between $\widehat{\mathbf{C}}$ and the observed images (i.e., how $\widehat{\mathbf{C}}$ relates to $\mathbf{L}$ through blurring and downsampling, and how it relates to $\mathbf{H}$ through a spectral response function if $\mathbf{H}$ is RGB). It also incorporates regularisation terms or priors that impose desired properties on $\widehat{\mathbf{C}}$, such as smoothness, sparsity in a transformed domain, or low-rank structure (Wei et al., 2015). Matrix and tensor factorisation techniques model $\widehat{\mathbf{C}}$ as a combination of spectral endmembers and spatial abundance maps (Yokoya et al., 2011; Lanaras et al., 2015). Model-based approaches are principled and flexible, allowing for the incorporation of physical constraints, and often yield high spectral fidelity. However, they typically involve iterative optimisation procedures, which can be computationally intensive, and their performance depends on the accuracy of the underlying models and the chosen priors.

- **Deep Learning Approaches:** In recent years, deep learning, particularly using Convolutional Neural Networks (CNNs) (Yang et al., 2018; Dong et al., 2022) and more recently Transformers (Hu et al., 2022b; Zhang et al., 2023), has become a prominent approach for pan-sharpening. These methods learn an end-to-end mapping from the input LR MS/HS image and the HR guidance image to the target HR MS/HS image. Architectures are often designed with multiple branches to process spatial and spectral information effectively, using techniques like attention mechanisms to capture complex relationships (Zhang et al., 2023). Deep learning models have demonstrated state-of-the-art performance on many benchmark datasets in terms of quantitative metrics. However, they require large amounts of registered training data (HR MS/HS ground truth paired with LR MS/HS and HR guidance inputs), can be computationally expensive to train, may act as "black boxes"

limiting interpretability, and their generalization performance to new sensors or scenes significantly different from the training data can be a concern.

The post-processing methodologies developed in Chapters 4 and 5 of this thesis are designed to be applicable to the outputs $\widehat{\mathbf{C}}$ generated by any of these diverse pan-sharpening techniques, aiming to further refine their spectral accuracy.

## 2.6 Matrix-R Framework

A foundational concept in colour science is metamerism, the phenomenon where two physically different spectral signals produce the exact same colour response in a three-channel imaging system, such as the human eye or an RGB camera (Cohen and Kappauf, 1982). These spectrally distinct signals that appear as a colour match are known as metamers. The Matrix-R framework, originating from the study of this phenomenon (Wyszecki and Stiles, 2000; Cohen and Kappauf, 1982), provides essential insights into the information content of signals captured by imaging systems with fewer sensors than the dimensionality of the original signal. It is particularly relevant for understanding RGB imaging of spectral signals.

Consider the image formation equation $\underline{\rho} = \mathbf{Q}^T\underline{e}$, where $\mathbf{Q}$ is the $31 \times 3$ sensor matrix. The framework decomposes the spectrum $\underline{e}$ into two orthogonal components relative to the subspace spanned by the sensor sensitivities (the columns of $\mathbf{Q}$).

- **Fundamental Metamer ($\underline{e}_{fm}$):** This component is the orthogonal projection of $\underline{e}$ onto the sensor subspace spanned by $\mathbf{Q}$. It captures the entire part of the spectrum that the sensors can respond to. All the information needed to produce the response $\underline{\rho}$ resides in $\underline{e}_{fm}$. The projection matrix $\mathbf{R}$ performs this projection:

$$\mathbf{R} = \mathbf{Q}(\mathbf{Q}^T\mathbf{Q})^{-1}\mathbf{Q}^T. \tag{2.10}$$

Applying this to the spectrum gives the fundamental metamer:

$$\underline{e}_{fm} = \mathbf{R}\underline{e}. \tag{2.11}$$

Since $\mathbf{Q}^T\underline{e}_{fm} = \mathbf{Q}^T\mathbf{R}\underline{e} = \mathbf{Q}^T\underline{e} = \underline{\rho}$, the fundamental metamer $\underline{e}_{fm}$ produces the same sensor response as the original spectrum $\underline{e}$. Importantly, if $\mathbf{Q}$ and $\underline{\rho}$ are known, $\underline{e}_{fm}$ can be uniquely determined without knowing $\underline{e}$:

$$\underline{e}_{fm} = \mathbf{Q}(\mathbf{Q}^T\mathbf{Q})^{-1}\underline{\rho}. \tag{2.12}$$

All spectra $\underline{e}_i$ that are metameric (i.e., produce the same response $\underline{\rho}$) share the exact same fundamental component $\underline{e}_{fm}$.

- **Metameric Black ($\underline{e}_{mb}$):**

  This is the component of the spectrum $\underline{e}$ that is orthogonal to the space defined by the camera's sensor sensitivities $\mathbf{Q}$. In other words, it lies in the null space of $\mathbf{Q}^T$. By definition, the sensors are "blind" to this component: $\mathbf{Q}^T\underline{e}_{mb} = \underline{0}$. It represents the spectral information completely lost during the dimensionality reduction inherent in the image capture process. It is calculated as the residual, where $\mathbf{I}$ is the $31 \times 31$ identity matrix:

$$\underline{e}_{mb} = \underline{e} - \underline{e}_{fm} = (\mathbf{I} - \mathbf{R})\underline{e}. \tag{2.13}$$

  The matrix $(\mathbf{I}-\mathbf{R})$ is the orthogonal projector onto the null space of $\mathbf{Q}^T$. Any vector $\underline{e}'_{mb}$ from this null space can be added to the fundamental metamer $\underline{e}_{fm}$ to form a valid spectrum $\underline{e}' = \underline{e}_{fm} + \underline{e}'_{mb}$ that still produces the original response $\underline{\rho}$. This infinite set of possible metameric blacks is the source of ambiguity in spectral recovery.

The decomposition $\underline{e} = \underline{e}_{fm} + \underline{e}_{mb}$ is unique and orthogonal ($\underline{e}_{fm}^T\underline{e}_{mb} = 0$, $\|\underline{e}\|^2 = \|\underline{e}_{fm}\|^2 + \|\underline{e}_{mb}\|^2$). This framework clarifies that spectral reconstruc-

tion or pan-sharpening essentially involves estimating the unknown metameric black $\underline{e}_{mb}$, given that $\underline{e}_{fm}$ is determined by the sensor measurements $\underline{\rho}$ (assuming known $\mathbf{Q}$). Chapter 4 leverages this by proposing a post-processing step applicable to any spectral recovery algorithm. Since the fundamental metamer component $\underline{e}_{fm}$ is uniquely determined by the known camera RGB response, it represents a physical constraint that many algorithms violate, producing an estimate $\widehat{\underline{e}}$ with an incorrect fundamental component $\widehat{\underline{e}}_{fm}$. The proposed method corrects this by simply replacing the incorrect $\widehat{\underline{e}}_{fm}$ with the correctly computed $\underline{e}_{fm}$, while retaining the algorithm's original estimate of the metameric black, $\widehat{\underline{e}}_{mb}$. Chapter 5 proposes a data-driven regression that implicitly learns a mapping respecting this decomposition without requiring knowledge of $\mathbf{Q}$.

This chapter has laid out the essential background in human vision, colourimetry, digital imaging technology, standard processing pipelines, colour correction techniques, spectral recovery challenges, and the fundamental Matrix-R theory. This comprehensive foundation serves as the launching point for the specific contributions and methodologies presented in the following chapters.

# Chapter 3

# Colour Correction: Comparison of Classical Methods and Neural Networks

This chapter is based on a published journal paper (Kucuk et al., 2023).

Building on the theoretical foundations of image formation and camera processing pipelines established in Chapter 2, this chapter evaluates existing colour correction algorithms. To address the limitations of current approaches, we then introduce and evaluate a novel exposure-invariant neural network model.

Colour correction is the process of converting RAW RGB pixel values of digital cameras to a standard colour space such as CIE XYZ. A range of regression methods including linear, polynomial and root-polynomial least-squares have been deployed. However, in recent years, various neural network (NN) models have also re-emerged in the literature as an alternative to classical methods. The use of neural networks for colour calibration is not entirely new; early studies, such as (Kang and Anderson, 1992; Adkins et al., 1993), demonstrated the potential of NNs to approximate the non-linear mapping between device-dependent and device-independent colour spaces. While interest in these methods has fluctuated, the recent availability of larger spec-

tral datasets, improved optimization techniques, and increased computational power has reignited research into NN-based solutions, positioning them as a viable competitor to classical approaches. In the first part of this chapter, a leading neural network approach is compared and contrasted with regression methods. We find that, although the neural network model supports improved colour correction compared with simple least-squares regression, it performs less well than the more advanced root-polynomial regression. Moreover, the relative improvement afforded by NNs, compared to linear least-squares, is diminished when the regression methods are adapted to minimise a perceptual colour error. Problematically, unlike linear and root-polynomial regressions, the NN approach is tied to a fixed exposure (and when exposure changes, the afforded colour correction can be quite poor). We explore two solutions that make NNs more exposure-invariant. First, we use data augmentation to train the NN for a range of typical exposures and second, we propose a new NN architecture which, by construction, is exposure-invariant. Finally, we look into how the performance of these algorithms is influenced when models are trained and tested on different datasets. As expected, the performance of all methods drops when tested with completely different datasets. However, we noticed that the regression methods still outperform the NNs in terms of colour correction, even though the relative performance of the regression methods does change based on the train and test datasets.

## 3.1 Introduction

Colour correction algorithms usually convert camera-specific RGB values into camera-independent colour spaces such as sRGB (Anderson et al., 1996) or CIE XYZ (Hunt and Pointer, 2011). In Figure 3.1, we plot spectral sensitivity functions of the Nikon D5100 camera and CIE XYZ colour matching functions. If there existed a linear transform which took the Nikon (or any other camera) sensitivity curves so that they were equal to the XYZ matching function then

the same linear transform would perfectly correct the camera's RGB responses to the corresponding XYZ tristimuli. However, there are no commercial photographic cameras that meet this linear transform condition and so camera RGBs can only be approximately converted to XYZs.

An illustration of the colour correction problem is shown in Figure 3.2. Here, RAW RGB Nikon D5100 camera responses are converted by linear colour correction to the sRGB (Anderson et al., 1996) colour space. The image shown is drawn from the Foster et al. hyperspectral image set (Foster et al., 2006) with the RGB and sRGB images calculated by numerical integration. Both images have the sRGB non-linearity applied.

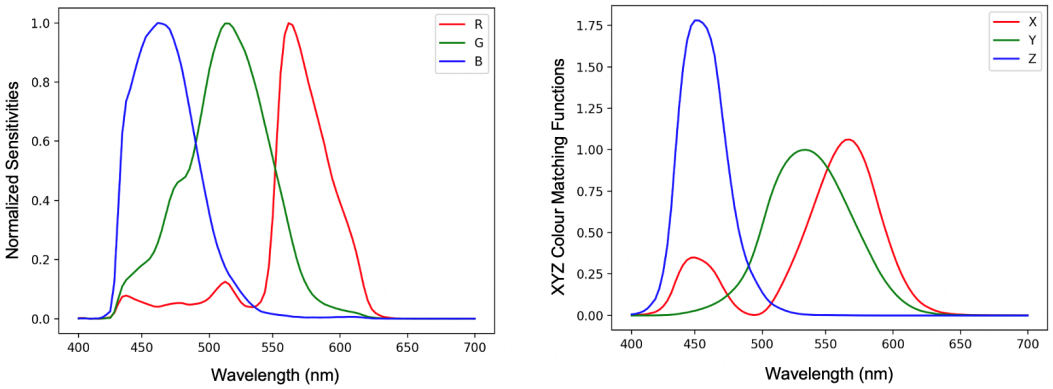


**Figure 3.1:** *Normalised sensitivity functions of Nikon D5100 camera (**left**) and the CIE XYZ standard observer colour matching functions (**right**). The XYZ matching curves are relative to 'Y' (green curve) which has a maximum response of 1.*

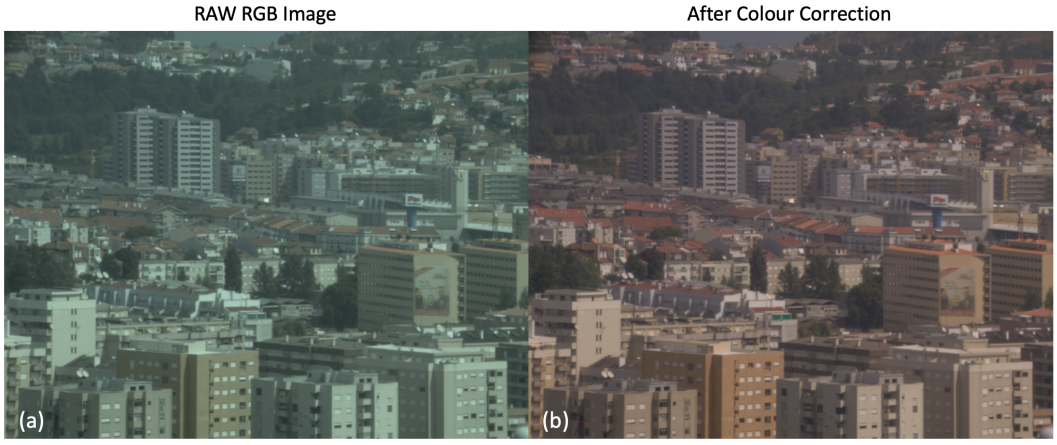


**Figure 3.2:** *An illustration of colour correction. The images are generated from David Foster's hyperspectral reflectance dataset (Foster et al., 2006) with Nikon D5100 camera responses and D65 illumination. The left one (a) demonstrates the RAW RGB image; the right one (b) represents the colour-corrected sRGB image.*

The most common approach to colour correction maps RGB data to corresponding XYZs using a $3 \times 3$ matrix (found by regression) such that:

$$\mathbf{M}\underline{\rho} \approx \underline{x} \tag{3.1}$$

where $\underline{\rho}$ and $\underline{x}$ represent the RAW RGB camera response vector and XYZ tristimulus, respectively. Polynomial (Hong et al., 2001) and root-polynomial (Finlayson et al., 2015) approaches can also be used for colour correction. In each case, the RGB values are expanded according to the order of the polynomial (normal or root) and a higher order regression is used to determine the regression transform. As an example, the second-order root-polynomial expansion maps $[R\ G\ B]^T$ to $[R\ G\ B\ \sqrt{RG}\ \sqrt{RB}\ \sqrt{GB}\ ]^T$ ($^T$ denotes transpose) and the correction matrix $\mathbf{M}$ is $3 \times 6$.

In addition to regression methods, neural networks have been used for colour correction. In the literature, there are several shallow network approaches (Kang and Anderson, 1992; Hwang et al., 2015; Fdhal et al., 2009; Adkins et al., 1993), and more recently, convolutional neural networks have been proposed (Li et al., 2020a; Vlachos and Skarlatos, 2021; Fu and Cao, 2020). However, the recent literature focuses on the problem of correcting the colours captured underwater where the correction problem is different to the one we address here (e.g., they deal with the attenuation of colour due to the distance between the subject and the camera).

Recently, MacDonald and Mayer (MacDonald and Mayer, 2021) designed a Neural Net (**NN**) with a fully connected Multilayer Perceptron (MLP) structure for colour correction and demonstrated that the network delivered colour correction that was better than the linear approach (Equation 3.1). In the first part of this chapter, we investigate the performance of **NN** algorithm versus regression methods more generally. Broadly, we confirm the finding that the **NN** approach is significantly better than linear regression but that the polynomial (Hong et al., 2001) and root-polynomial (Finlayson et al., 2015) regressions

actually deliver significantly better colour correction than the **NN**.

As well as delivering poorer performance than the best regression methods, the **NN** approach was also found not to be exposure invariant. That is, a network trained to map RGBs to XYZs for a given exposure level delivered relatively poor colour correction when the exposure level changed. The polynomial colour correction algorithm (Hong et al., 2001) suffers from the same exposure problem; polynomial regression works very well for a fixed exposure but less well when exposure changes (Finlayson et al., 2015). Indeed, the existence of this problem led to the development of the *root* polynomial correction algorithm (which, by construction, is exposure invariant) (Finlayson et al., 2015).

Let us now run a quick experiment to visually understand the problem of exposure in polynomial regression (we get similar results for the Neural Net). For the UEA dataset of spectral reflectance images (Hordley et al., 2004), we sampled four reflectances. The RAW RGB responses of the Nikon D5100 are shown at the top of Figure 3.3a. Next, (b), the actual true sRGB image, rendered for a D65 whitepoint, is shown. In (c), we render these reflectances using the Nikon camera sensitivities and correct the four RGBs to the corresponding sRGB values using a second-order polynomial expansion. In detail, the second-order expansion has nine terms, $[R^2\ G^2\ B^2\ RG\ RB\ GB\ R\ G\ B]^T$, and the colour correction matrix is $9 \times 3$. In both the sRGB and fitted camera image, the maximum over all pixel values (across all three colour channels) is scaled to 1.

Now, we multiply the linear Nikon RGBs and the corresponding linear ground-truth values by 7. As before, we calculate the second-order polynomial expansion of the RGBs and then apply the same colour correction matrix for the exposure $= 1$ condition. After colour correction, we again scale to make the exposure (across all three channels) equal to 1. The resulting colour patches are shown at the bottom of Figure 3.3. It is clear that the colours of the patches have changed significantly and that the colour correction is more accurate for

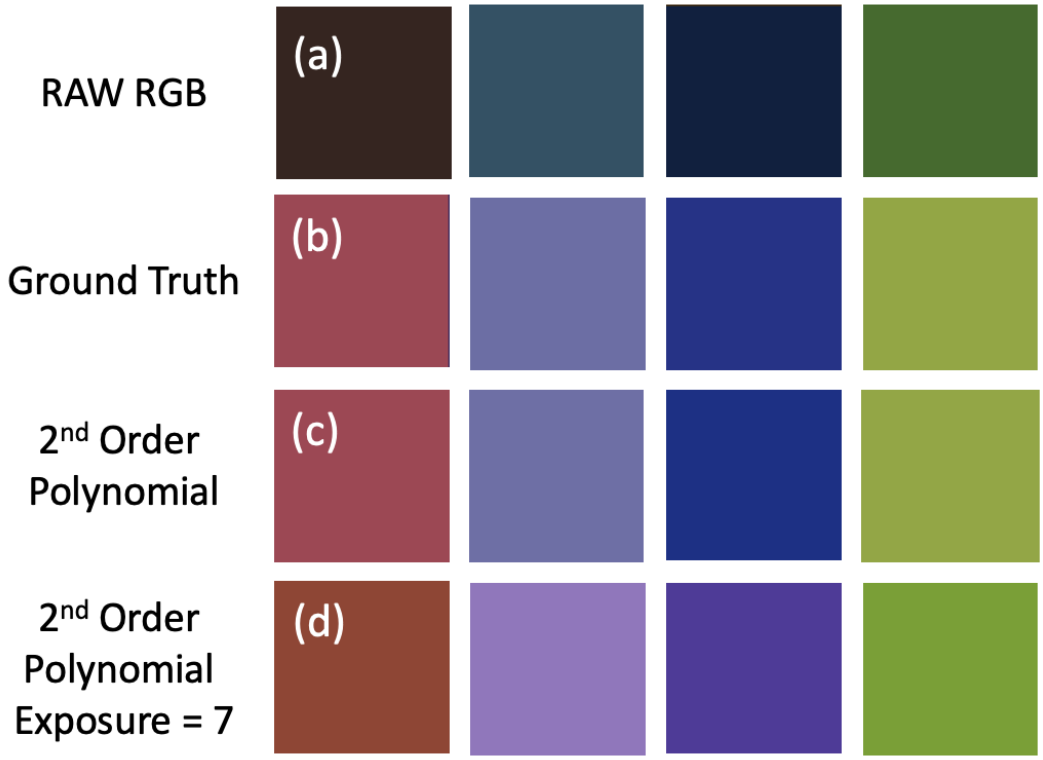


**Figure 3.3:** *Top (a), four RAW RGB colour patches which are generated with Nikon D5100 camera sensitivities. (b), the true sRGB rendering of the patches. (c), the Nikon camera image corrected with a second-order polynomial regression. Bottom (d) shows the output of polynomial colour correction, calculated for exposure = 1, applied to the camera RGBs × 7 (exposure level = 7). This simulates an extreme high-dynamic-range scenario to test the model's generalisation across signal magnitudes, assuming a theoretical sensor response capable of avoiding saturation. All three images are scaled so the brightest value (across all three colour channels) is 1.*

the colours rendered under the same exposure conditions (panel (c) is more similar to (b) than (d) is to (b)).

In the second part of this chapter, we try to solve this problem: we seek to make neural network colour correction exposure invariant. We investigate two approaches. First, we augment the training data used to define the neural network with data drawn from many different exposure levels. Second, we design a new network, which, by construction, is exposure invariant. Our new network has two components. The chromaticity component network attempts to map camera rgb chromaticity to colorimetric xyz chromaticity. In the second component, we linearly correct R, G and B to predict X + Y + Z (mean colorimetric brightness). Given this summed brightness and the target xyz

chromaticity, we can calculate XYZ. By construction, the combination of the chromaticity-correcting network and the linear brightness predictor generates XYZs in an exposure-invariant manner. Experiments demonstrate that both of our exposure-invariant networks continue to deliver better colour correction than a $3 \times 3$ linear matrix.

In prior work in this area, colour correction algorithms were assessed with respect to a single data set. For example, the 1995 reflectances (SFU set (Barnard et al., 2002)) are widely used. Often $k$ fold cross-validation is used, i.e., the reflectance set is split into $k$ similar-sized folds. Then, each fold is used in turn as a test set and the other four are used for training the algorithms, and the performance of an algorithm is averaged over the $k$ test sets. However, when a single set is used, often the statistics of all the folds are similar and so training on a subset of a given dataset can be almost as good as training on the whole set. Thus, an important contribution of this chapter is to run a cross-validation experiment where the *folds* are different datasets (with known a priori statistics in common).

In Section 3.2, we summarise colour correction and introduce the regression and Neural Net (NN) algorithms. We show how NN methods can be made exposure invariant in Section 3.3 and we report on a comprehensive set of colour correction experiments in Section 3.4. We have results in Section 3.5 and conclusion in Section 3.6.

## 3.2 Background

Let $Q_k(\lambda)$ denote the k-th camera spectral response function and $\mathbf{Q}(\lambda)$ denote the vector of these functions as in Figure 3.1. The camera response to a spectral power distribution $I(\lambda)$ illuminating the $j$-th reflectance $S_j(\lambda)$ is written as:

$$\underline{\rho} = \int_{\omega} \mathbf{Q}(\lambda) I(\lambda) S_j(\lambda) d\lambda \tag{3.2}$$

where $\omega$ denotes the visible spectrum (400 to 700 nm) and $\underline{\rho}$ denotes the vector of RGB responses. Similarly, given the XYZ colour matching $\mathbf{X}(\lambda)$, the tristimulus response $\underline{x}$ is written as:

$$\underline{x} = \int_{\omega} \mathbf{X}(\lambda) I(\lambda) S_j(\lambda) d\lambda \tag{3.3}$$

Suppose $n \times 3$ matrices $\mathbf{P}$ and $\mathbf{X}$ record (in rows) the camera responses and tristimuli of $n$ surface reflectances, respectively. To find the $3 \times 3$ matrix $\mathbf{M}$ - that is, the best linear map - in Equation (3.1), we minimise:

$$\arg\min_{\mathbf{M}} \|\mathbf{PM} - \mathbf{X}\|_F \tag{3.4}$$

where $\|.\|_F$ denotes the Frobenius norm (Horn and Johnson, 1990). We can solve for $\mathbf{M}$ in closed form using the Moore–Penrose inverse (Penrose, 1955):

$$\mathbf{M} = \left[\mathbf{P}^T\mathbf{P}\right]^{-1}\mathbf{P}^T\mathbf{X} \tag{3.5}$$

To extend the regression method, we define a basis function ${f_e}^o()$ where the subscript $e$ denotes the type of expansion - here $e{=}p$ and $e{=}r$, respectively, denote polynomial and root-polynomial expansions - and the superscript $o$ denotes the order of the expansion. As an example, if we are using the second-order root-polynomial expansion (Finlayson et al., 2015), then we write:

$${f_r}^2(\underline{\rho}) = \left[R\, G\, B\, \sqrt{RG}\, \sqrt{RB}\, \sqrt{GB}\right]^T \tag{3.6}$$

Again, we can use Equations 3.4 and 3.5 to solve for the regression matrix $\mathbf{M}$, though $\mathbf{M}$ will be non-square (and depend on the number of terms in the expansion). For our second-order root-polynomial expansion, the columns of $\mathbf{P}$ will be the six terms in the root-polynomial expansion ($\mathbf{P}$ is an $n \times 6$ matrix) and $\mathbf{M}$ will be $6 \times 3$. See, for example, (Finlayson et al., 2015) for details of higher-order expansions.

Optimising for the Frobenius norm in Equation 3.4 may be undesirable because the Euclidean differences in the XYZ colour space do not correspond to the perceived differences in colour. Instead, it is more desirable to optimise for the differences in perceptually uniform colour spaces such as CIELAB (Hunt and Pointer, 2011) or using colour difference formulas such as CIE Delta E 2000 (Sharma et al., 2005). Let us denote the magnitude of the difference vector between a mapped camera response vector and its corresponding ground truth CIELAB value as:

$$\Delta^{\mathbf{M},o,e} = \|C(\mathbf{M}^T {f_e}^o(\underline{\rho}), \underline{w}) - C(\underline{x}, \underline{w})\| \tag{3.7}$$

where C() maps input vectors according to the CIELAB function to corresponding Lab triplets and the superscripts $e$ and $o$ are as before. The parameter $\underline{w}$ denotes the XYZ tristimulus of a perfect white diffuser and is required to calculate CIELAB values. To find the best regression matrix, we seek to minimise:

$$\arg\min_{\mathbf{M}} \sum_{i=1}^{n} \Delta_i^{\mathbf{M},o,e} \tag{3.8}$$

Unfortunately, there is no closed-form solution to the optimisation described in Equation (3.1). Instead, a search-based strategy such as the Nelder–Mead simplex method (Lagarias et al., 1998) can be used to find $\mathbf{M}$ (though there is no guarantee that the global optimum result is found, (Lagarias et al., 1998) is a local minimiser).

In this chapter, we will also be interested in minimising CIE Delta E 2000 (Sharma et al., 2005). Here, it is not possible to model colour difference as the Euclidean distance between triplets (which are non-linear transforms of XYZs and regressed RGBs). For the Delta E 2000 case, we write the error $\Delta_{2000}^{\mathbf{M},o,e}$ (subscript identifies the errors as CIE Delta E 2000) in Equation (3.9) and the minimisation to solve for is again given by Equation (3.8). As before, we minimise Delta E 2000 using the search-based algorithm. In Equation (3.9),

we denote the function that calculates the CIE 2000 error as $f()$. We see this function takes three inputs: the regressed RGB ($\mathbf{M}^T {f_e}^o(\underline{\rho})$), the corresponding XYZ ($\underline{x}$) and the XYZ for the illuminant ($\underline{w}$). For details of $f()$, the reader is referred to (Sharma et al., 2005).

$$\Delta_{2000}^{\mathbf{M},o,e} = f(\mathbf{M}^T {f_e}^o(\underline{\rho}), \underline{x}, \underline{w}) \tag{3.9}$$

As an alternative to regression methods, colour correction can also be implemented as an artificial neural network. MacDonald and Mayer's (MacDonald and Mayer, 2021) recently published neural network is illustrated in Figure 3.4 and is a leading method for neural network colour correction.

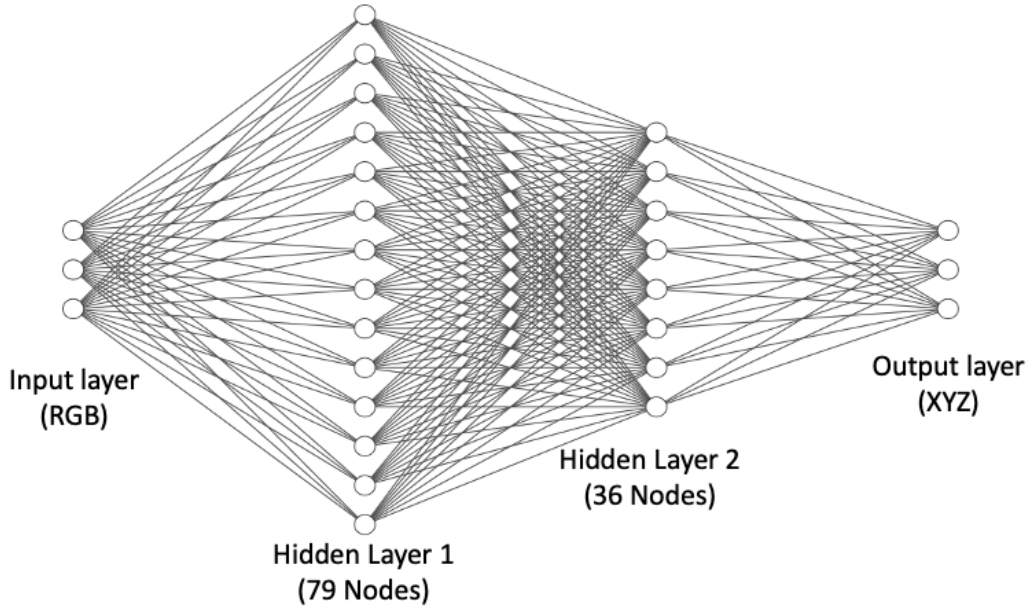


**Figure 3.4:** *MacDonald and Mayer's Neural Net (MacDonald and Mayer, 2021). Input and output layers consist of three nodes which are RGB and XYZ, respectively. In between, there are two hidden layers formed by 79 and 36 nodes.*

This Neural Net has 3189 'connections', indicating the cost of colour correction is on the order of 3189 multiplications and additions (the number of operations applied as data flows from left to right). In comparison, the complexity of the second-order root-polynomial correction has three square root operations and (when the $6 \times 3$ correction matrix is applied) 18 multiplications and 15 additions, i.e., it is 2 orders of magnitude quicker to compute.

In part, the practical utility or otherwise of the Neural Net approach will rest on the trade-off between how well it improves colour correction (say compared to the linear method) and its higher computational cost. The Neural Net is trained to minimise the Delta E 2000 errors.

## 3.3 Exposure-Invariant Neural Nets

Abstractly, we can think of a neural network as implementing a vector function $f()$ such that:

$$f(\underline{\rho}) \approx \underline{x} \tag{3.10}$$

where $\underline{\rho}$ and $\underline{x}$ denote the RAW RGB camera response vector and the XYZ tristimulus, respectively. When exposure changes - for example, if we double the quantity of light - then, physically, the RGB and XYZ responses also double in magnitude. We would like a colour correction function to be exposure invariant:

$$f(k\underline{\rho}) \approx k\underline{x} \tag{3.11}$$

where $k$ in Equation (3.11) is a positive scalar. This *homogeneity* property is actually rare in mathematical functions. It holds for linear transforms - $f(\underline{\rho}) = \mathbf{M}\underline{\rho}$ implies that $f(k\underline{\rho}) = k\mathbf{M}\underline{\rho}$ - and root-polynomials, but it is not true for polynomial expansions (Finlayson et al., 2015). A Neural Net, in order to not collapse to a simple affine transformation, uses non-linear activation functions. These non-linearities, while an essential aspect of the network, make it difficult to attain homogeneity. This homogeneity is required if colour correction is to be invariant to a changing exposure level. The MacDonald and Mayer network is found not to be exposure invariant. In fact, this is entirely to be expected; there is no reason why NNs should exhibit the homogeneity property and every reason why they should not. Significantly, the variation

in performance with exposure is a problem. In the experimental section, we show that colour correction performance drops markedly when there are large changes in exposure.

Now, let us consider alternative methods for enhancing the robustness of neural networks to exposure variations or achieving exact exposure invariance. In neural network research, if we observe poor performance for some input data, then the *trick* is to retrain the network where more of the problematic data are added to the training set. In neural network parlance, we *augment* the training data set. Here, we have the problem that a network trained for one light level delivers poor colour correction when the light levels change (e.g., when there is double the light illuminating a scene). So, to achieve better colour correction as exposure levels change, we will augment our colour correction training data - the corresponding RGBs and XYZs for a single exposure level - with corresponding RGBs and XYZs for several exposure levels. Our retrained MacDonald and Mayer Network using the exposure level augmented dataset is our first (more) exposure-invariant neural network solution to colour correction.

Perhaps a more elegant approach to solving the exposure problem is to redesign the network so it is, by construction, exactly exposure invariant. We show such an architecture in Figure 3.5. In the top network we learn - using Macdonald and Mayer's **NN** - the mapping from input $r$, $g$ and $b$ chromaticity to $x$, $y$ and $z$ chromaticity. When the camera and tristimulus response are denoted as $[R\ G\ B]^T$ and $[X\ Y\ Z]^T$, then the corresponding chromaticities are defined as $r = R/(R+G+B)$, $g = G/(R+G+B)$ and $b = B/(R+G+B)$; and $x = X/(X+Y+Z)$, $y = Y/(X+Y+Z)$ and $z = Z/(X+Y+Z)$. In the 'intensity' network (bottom of Figure 3.5) we map $R$, $G$ and $B$ to predict $X+Y+Z$ by using only a linear activation function. Multiplying the estimated $[x\ y\ z]^T$ by the estimated $X+Y+Z$ returns an estimated $[X\ Y\ Z]^T$.

In other words, when we change the scalar $k$, the input of the first network remains unchanged, as it depends on the chromaticities. Consequently, the

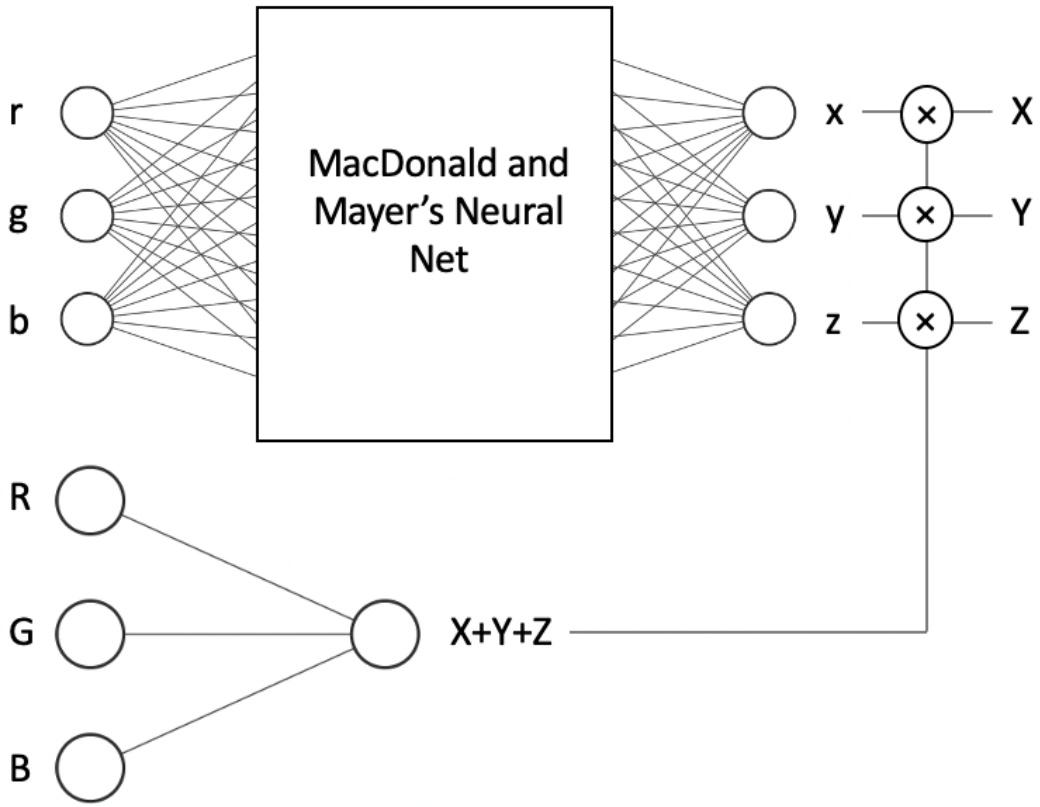


**Figure 3.5: NN-EI** *with two networks, while the top one learns chromaticities, the bottom one learns the sum of XYZs. Multiplications of these give us the XYZs.*

output also remains the same. The only elements that change are the input and output of the second (intensity) network, as it utilises the actual RGB input and calculates the sum of the actual XYZ. However, this change is linear since it involves a dot product; there are no activation functions or biases, only three multiplications. Thus, for the given input RGB $\times$ $k$, the multiplication of the output of the first network (xyz chromaticities) and the second network (intensity $\times$ $k$) yields (XYZ $\times$ $k$). As a result, the system is exposure invariant, meaning that chromaticities do not change at different illumination levels, and inputs and outputs scale linearly with the scalar $k$.

Given that the intensity network consists of only three connections with a linear activation function and no biases, the additional computational cost compared to the operation original network is quite low. It involves just one addition and six multiplication operations. Specifically, three multiplications and one addition are used for calculating the intensity, while the other three

multiplications are performed to multiply the intensity with the x, y and z chromaticities in order to obtain the corresponding XYZ values.

Informally, let us step through an example to show that the network is exposure invariant. That is, we want to show that the respective RGBs $\underline{\rho}$ and $k\underline{\rho}$ are mapped to the estimated XYZs $\underline{x}$ and $k\underline{x}$. Let us consider the RGB vector [10, 50, 40]. To make the $r$, $g$, $b$ chromaticities, we divide RGB values by the sum of RGB yielding the r, g, b chromaticities: [0.1, 0.5, 0.4]. Suppose our chromaticity network outputs [0.3, 0.4, 0.5] (the estimates of the $x$, $y$, $z$ chromaticities) and the second network (the bottom one in Figure 3.5) returns 50 as the prediction of X+Y+Z. Now, we multiply output $x$, $y$, $z$ chromaticities by 50, and we generate the XYZ output: [15, 20, 25].

Now, let us double the RGB values: [20, 100, 80]. Clearly, the chromaticities are unchanged ([0.1, 0.5, 0.4]). The output of the second network is a simple linear dot-product; thus, the output must be equal to 100 (as opposed to 50 before the exposure doubling). Finally, we multiply the estimated $x$, $y$, $z$ chromaticities, [0.3, 0.4, 0.5], by 100 and the final output is [30, 40, 50] (which is exactly double as before). This simple example demonstrates that if the exposure changes by a scalar $k$ then the output of the network also scales by $k$ and so our new network is exposure invariant.

## 3.4 Experiments

### 3.4.1 Preparation of Datasets

In our experiments, we used four spectral datasets: the Simon Fraser University (SFU) reflectance set (Barnard et al., 2002), the Ben-Gurion University Dataset (BGU) (Arad and Ben-Shahar, 2016), the Columbia University Dataset (CAVE) (Yasuma et al., 2010), and the Foster et al. Dataset (FOSTER) (Foster et al., 2006). The spectral sensitivities of the Nikon D5100 camera (Darrodi et al., 2015) and D65 viewing illuminant (Judd et al., 1964)

are also used in all experiments. All input RGB values and their corresponding target XYZ values are calculated using numerical integration, without any gamma correction.

The SFU reflectance set (Barnard et al., 2002) comprises 1995 spectral surface reflectances, including the 24 Macbeth colour checker patches, 1269 Munsell chips, 120 Dupont paint chips, 170 natural objects, and 407 additional surfaces. In Figure 3.6, in the CIE 1931 chromaticity diagram, we plot the xy chromaticities of the SFU dataset. We also show the gamut of colours achievable using Rec 709 primaries (white triangle). It is evident that the SFU reflectance set comprises a wide range of colours.

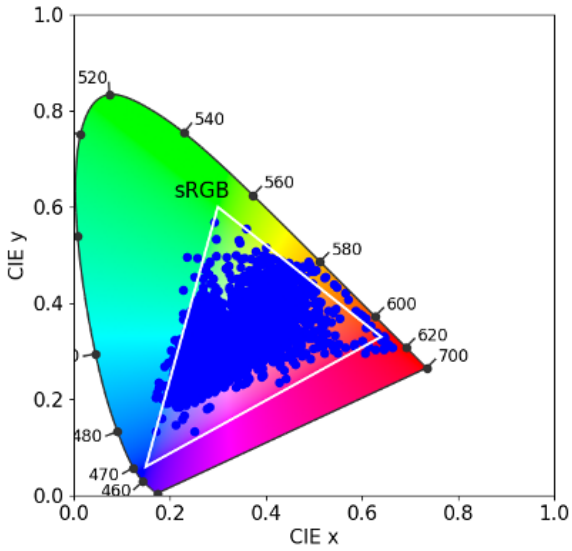


**Figure 3.6:** *Gamut of the SFU dataset on the CIE 1931 chromaticity diagram. The white triangle shows the sRGB gamut.*

The BGU Dataset comprises 201 multi-spectral outdoor images of various sizes. To make each image equally important (statistically), we resized them using bilinear resampling to 1000 $\times$ 1000 in size. In all our experiments, we used D65 as our viewing illuminant. However, the BGU images have radiance spectra and we would like the light component of each radiance spectrum to be D65. Thus, we set out to manually identify achromatic surfaces in each of the 201 scenes with the additional constraint that we judged the image to be predominantly lit by one light. We then used the corresponding spectrum (for

the achromatic surface) to be the spectrum of the prevailing light. Dividing by the prevailing light and multiplying by D65 re-renders the scene for D65 illumination.

We were only confident that 57 of the 201 scenes met the two constraints of having a clearly identifiable achromatic surface and being lit by a single prevailing light. Thus, only 57 of the BGU images re-rendered to D65 were actually used. Finally, each image was scaled so that the maximum value in the image is 1.

The CAVE Dataset comprises 32 indoor reflectance images, segregated into five distinct categories: stuff, skin and hair, food and drinks, real and fake, and paints. Each image has dimensions of 512 $\times$ 512 pixels. We found it necessary to exclude one image, entitled "watercolors", due to the presence of missing data. Consequently, we were left with a total of 8,126,464 pixels for analysis (31 $\times$ 512 $\times$ 512).

Lastly, we used the FOSTER hyperspectral image set (Foster et al., 2006) which consists of eight different images. Similar to the BGU Dataset, the various-sized images were resized to 1000 $\times$ 1000 and 8,000,000 pixels were obtained (8 $\times$ 1000 $\times$ 1000).

As the CAVE and FOSTER images contain only reflectance data, we multiply each spectral measurement in each image and per pixel by the D65 illuminant spectrum.

### 3.4.2 Algorithms

The following algorithms are investigated in this chapter:

(i) **LS**: Least-Squares Regression.

(ii) **LS-P**: denotes Least-Squares Polynomial Regression. Here, we use the second-order expansion which maps each three-element vector to a ten-element vector.

(iii) **LS-RP**: Least-Squares Root-Polynomial Regression. Again, a second-order expansion is used which for root-polynomials has six terms.

(iv) **LS-Opt**.

(v) **LS-P-Opt**.

(vi) **LS-RP-Opt**.

(vii) **NN**: MacDonald and Mayer's Neural Net (MacDonald and Mayer, 2021).

(viii) **NN-AUG**: The **NN** with an augmented training data with different exposure levels.

(ix) **NN-EI**: Here, we use two different neural networks. The first one learns to calculate chromaticities and the second one for the sum of XYZ.

**Opt** denotes optimisation, where we use the CIELAB or CIE Delta E 2000 loss values for training depending on the experiment (we make clear which is used in which experiment) with the Nelder–Mead simplex method (Lagarias et al., 1998) (see Equations (3.7) and (3.8)) to solve for the linear (iv), polynomial (v) and root-polynomial regressions (vi). To ensure robust convergence and guaranteed improvement, we initialise the Nelder–Mead algorithm using the corresponding closed-form least-squares solution (i.e., the matrix derived via standard regression prior to perceptual loss minimisation). The regressions (i) through (iii) - minimising error in the XYZ tristimuli space - are found in closed form using the Moore–Penrose inverse (Equations (3.5) and (3.6)).

As suggested in MacDonald and Meyer's original paper all the colour corrections, NNs are trained to minimise the CIE Delta E 2000 error.

Apart from the cross-dataset experiments, our colour correction algorithms are tested on the SFU dataset using a five-fold cross-validation methodology for the fixed and different exposures experiments. Here, the reflectance dataset is split into five equal-sized folds (399 reflectances per fold). Each algorithm is trained using four of the folds and then tested on the remaining

fold to generate error statistics. The process is repeated five times (so every fold is the test set exactly once). According to this methodology, the reported error statistics are averages over the five experiments.

### 3.4.3 Details about How the Neural Net Was Trained

For **NN**, we used MacDonald and Mayer's (MacDonald and Mayer, 2021) neural network, which has RGB values as input and XYZ as target, the $3 \times 79 \times 36 \times 3$ fully connected MLP architecture with two hidden layers as shown in Figure 3.4. As in the original paper, we used the Adam optimiser with a learning rate of 0.001 to train the network to minimise CIE Delta E 2000 (Sharma et al., 2005). We had to raise the number of epochs from 65 (used in the original study) to 500 for the neural network to develop a successful mapping because we were working on relatively small datasets. We also used mini-batch gradient descent with a batch size of 8. Our model used 20% of the training data for the validation set and used the early stopping method, which means that the training ends automatically if there is no improvement in validation loss after a specified number of epochs (which in our model is 100) with a call-back function. We chose the best model based on the validation loss. The **NN-AUG** and **NN-EI** were trained using the same methodology.

### 3.4.4 Details about Exposure Experiments

The colour correction performance for all algorithms was first calculated for a fixed reference exposure (exposure = 1) level. Then, we tested our models under different exposure levels to understand their performance when the exposure level changed. We use exposure values of 0.2, 0.5, 1, 2 and 5 (e.g., 0.2 and 5, respectively, meaning the amount of light was 1/5 and 5 times the reference condition of exposure 1).

In **NN-AUG**, in order to achieve successful results at different exposure levels, we augmented the training data with different exposure factors, which

are 0.1, 0.2, 0.5, 2, 5 and 10 times the original samples. Then, we tested the models with the test samples with an original exposure level.

The **NN-EI** is, by construction, exposure invariant; thus, it was only trained using the data for the reference exposure level.

### 3.4.5 Details about Cross-Dataset Experiments

In the cross-dataset section, we trained our algorithms with a single dataset and individually tested them with the remaining datasets. We repeated this process four times (i.e., every dataset was used for training once and for testing three times). Additionally, for the NN method, we allocated 20% of the training data as a validation set. The whole training set was used for the regression colour correction methods.

Even our relatively small image sets comprise millions of pixels. Training with millions of pixels is computationally expensive, especially for the NN and search-based regression methods. To mitigate training complexity, we reduced the multi-spectral images of the BGU, CAVE, and FOSTER datasets to thumbnails of size $40 \times 40 \times 31$ (using nearest neighbour downsampling). Additionally, for all our datasets (including SFU), we sampled the set of images and included an observed spectrum if and only if it was at least 5 degrees apart from any spectrum already in the set (we built the set incrementally by adding a new spectrum if it was 5 degrees away from all previously selected members).

Thumbnail creation and angular threshold filtering resulted in, respectively, 523, 354, 4188 and 4331 spectra for the SFU, BGU, CAVE and FOSTER datasets. All our colour correction algorithms were trained for these small sets of spectra. However, testing was conducted on the full-sized spectral data, which were 1995 samples for SFU, 57,000,000 for BGU ($57 \times 1000 \times 1000$), 8,126,464 for CAVE ($31 \times 512 \times 512$) and 8,000,000 for FOSTER ($8 \times 1000 \times 1000$).

## 3.5 Results

### 3.5.1 SFU Results for the Fixed Illumination

We report the CIELAB and CIE Delta E 2000 error results for our seven algorithms for the SFU dataset in Tables 3.1 and 3.2, respectively. As discussed earlier, we ran a five-fold cross-validation experiment. This means that there are five sets of colour correction - four folds used for training and the other for testing. We calculate the Mean, Max, Median and 95% (percentile) errors across five folds. The figures in Tables 3.1 and 3.2 report these error statistics averaged over the five sets. For the '-Opt' algorithms, we minimised the mean CIELAB Delta E error and then in Table 3.1, we assessed the correction performance also using CIELAB. Table 3.2 reports the same result, where Delta E 2000 was used (but again the algorithms were trained to minimise the CIELAB error.

Table 3.1: CIE LAB Delta E error statistics.

| **Methods** | **Mean** | **Max** | **Med** | **95%** |
|---|---|---|---|---|
| LS | 1.62 | 15.47 | 0.93 | 5.32 |
| LS-P | 1.29 | 11.00 | 0.78 | 4.01 |
| LS-RP | 1.19 | 13.97 | 0.70 | 3.62 |
| LS-Opt | 1.48 | 10.62 | 0.9 | 4.63 |
| LS-P-Opt | 1.17 | 8.47 | 0.77 | 3.62 |
| LS-RP-Opt | 1.10 | 7.36 | 0.72 | 3.39 |
| NN | 1.40 | 12.26 | 0.93 | 4.06 |

Table 3.2: CIE Delta E 2000 error statistics.

| **Methods** | **Mean** | **Max** | **Med** | **95%** |
|---|---|---|---|---|
| LS | 0.94 | 7.71 | 0.70 | 2.62 |
| LS-P | 0.79 | 4.63 | 0.59 | 2.18 |
| LS-RP | 0.72 | 7.13 | 0.49 | 2.14 |
| LS-Opt | 0.91 | 5.52 | 0.67 | 2.45 |
| LS-P-Opt | 0.75 | 4.25 | 0.54 | 2.08 |
| LS-RP-Opt | 0.69 | 3.81 | 0.52 | 1.96 |
| NN | 0.85 | 4.18 | 0.66 | 2.15 |

We see that, although the neural network algorithm returns significantly

better results compared with the standard Least-Squares Regression model, it delivers poorer colour correction compared with the other regression methods. The polynomial and root-polynomial deliver low errors, and both can be improved using the differences in the CIELAB space as a loss function. The -Opt variants of the regression algorithms have significantly lower maximum errors. The reader might be interested to know that if we train the neural network with the CIELAB loss function or use CIE Delta E 2000 on the classical methods, we found that these do not change the rank ordering of the results.

In Figure 3.7, we plot the CIELAB and CIE Delta E 2000 error distributions as violin plots for four of our algorithms. The tails of the violin plots are long because of the large maximum values (outliers). The width of the violin captures the probability density of the error distribution (Hintze and Nelson, 1998). Again, the distributions show that the root-polynomial regression with the CIELAB loss shows the best performance.

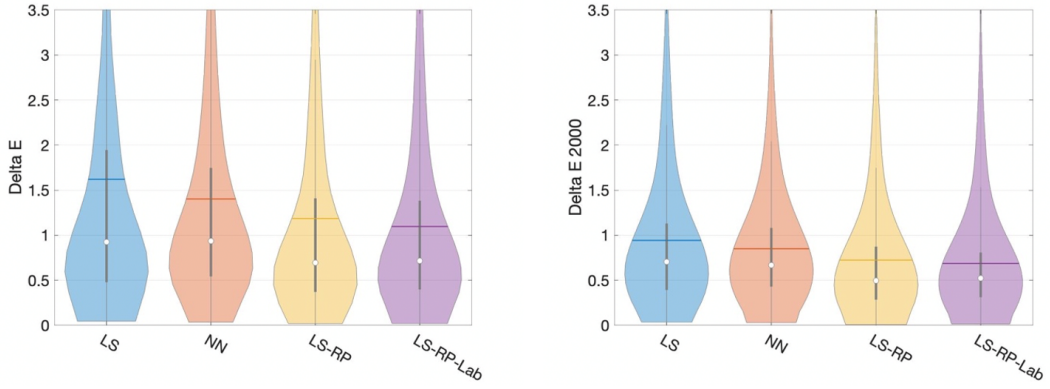


**Figure 3.7:** *A comparison of error distribution of four main methods by CIE Delta E and CIE Delta E 2000. Inside each 'violin', the white dot, the horizontal line, the vertical bar and the black line indicate, respectively, the median, mean, interquartile range and 1.5 interquartile range.*

We performed the sign test (Hordley and Finlayson, 2006) to establish if there is a statistically significant difference between the NN and LS-RP-Opt results. Both the p-values for CIELAB and CIE Delta E 2000 are less than 0.0001 in the 99% confidence level. The difference in algorithm performance is statistically significant.

### 3.5.2 Results for Different Exposure Values with Exposure Invariant Neural Nets

In Table 3.3, we report the performance results of the six methods at different illumination levels. As a reminder, in this experiment, the algorithms were trained with the original light level and tested with different intensity levels. The linear and root-polynomial regression methods are unaffected by exposure (they have the same performance across exposure changes). Equally, it is evident that the original MacDonald and Meyer **NN** performs poorly when the exposure changes.

Although the **NN-AUG** method exhibits a fair degree of exposure invariance, its performance still degrades slightly as the change in light levels is more extreme (compared to the reference condition). The **NN-EI** that was designed to be exactly exposure invariant delivers better results than the **NN-AUG** method, especially for the conditions where the exposure level is smaller than 0.5 or bigger than 5.

Table 3.3: CIE LAB Delta E error statistics at different exposure levels.

| **Methods** | **0.2** | **0.5** | **1** | **2** | **5** |
|---|---|---|---|---|---|
| LS | 1.62 | 1.62 | 1.62 | 1.62 | 1.62 |
| LS-RP | 1.19 | 1.19 | 1.19 | 1.19 | 1.19 |
| LS-RP-Opt | 1.10 | 1.10 | 1.10 | 1.10 | 1.10 |
| NN | 2.60 | 1.57 | 1.40 | 1.92 | 3.77 |
| NN-AUG | 2.25 | 1.50 | 1.30 | 1.25 | 1.38 |
| NN-EI | 1.53 | 1.53 | 1.53 | 1.53 | 1.53 |

### 3.5.3 Results for Cross-Dataset Experiments

Tables 3.4–3.7 present the test results of the seven algorithms trained on the SFU, BGU, CAVE, and FOSTER datasets, respectively (and then tested on the combination of the three remaining datasets). The regression models labelled as 'Opt' in this section were optimised using the CIE Delta E 2000 loss function, similar to the NN. All results displayed in the tables represent the av-

erage CIE Delta E 2000 errors. It is worth noting that to avoid overwhelming the article with results, we did not include tables containing CIELAB results. We used CIELAB to train the algorithms previously and here we use CIE 2000. Irrespective of the colour difference metric used to train or test the methods, the ranking of the algorithms stays the same.

Regarding Table 3.4, we observe the cross-dataset test results of models trained with SFU. All models demonstrate their best performance on the FOSTER dataset, followed by the CAVE dataset, and finally on the BGU dataset. However, regardless of the test data, LS-P consistently remains the top-performing model, while NN exhibits the poorest performance. The change in the ranking of polynomial and root-polynomial algorithms is not entirely unexpected. Here, there is a greater difference in the spectra in training and testing datasets. The previous experiments might be interpreted as that the RP method works well when training and testing datasets share similar spectral statistics.

Table 3.4: Mean Delta E 2000 error statistics (SFU used as training set).

| **Methods** | **BGU** | **CAVE** | **FOSTER** |
|---|---|---|---|
| LS | 3.49 | 4.60 | 2.49 |
| LS-P | <u>3.12</u> | <u>4.28</u> | <u>2.03</u> |
| LS-RP | 3.52 | 4.56 | 2.36 |
| LS-Opt | 3.49 | 4.69 | 2.32 |
| LS-P-Opt | 3.31 | 4.52 | 2.04 |
| LS-RP-Opt | 3.46 | 4.47 | 2.28 |
| NN | 3.93 | 4.86 | 2.51 |

In Table 3.5, we trained using the BGU dataset and tested on the other datasets. Here, both polynomial models and the NN performed poorly. The best method overall is the Linear $3 \times 3$ matrix (LS).

Table 3.6 reports the results of models which are trained on the CAVE Dataset and tested on other datasets. Here, we observe that models trained on the CAVE Dataset yield better results compared to the other tables. This might indicate that CAVE best represents the data found in the other test

Table 3.5: Mean Delta E 2000 error statistics (BGU used as training set).

| **Methods** | **SFU** | **CAVE** | **FOSTER** |
|---|---|---|---|
| LS | 3.31 | 4.45 | 2.95 |
| LS-P | 10.80 | 6.59 | 5.64 |
| LS-RP | 5.27 | 5.88 | 2.81 |
| LS-Opt | 3.81 | 4.30 | 3.37 |
| LS-P-Opt | 10.85 | 6.61 | 5.63 |
| LS-RP-Opt | 5.51 | 5.58 | 3.16 |
| NN | 17.10 | 18.75 | 21.98 |

sets. Here, the root-polynomial method performs best.

Table 3.6: Mean Delta E 2000 error statistics (CAVE used as training set).

| **Methods** | **SFU** | **BGU** | **FOSTER** |
|---|---|---|---|
| LS | 2.15 | 1.88 | 2.54 |
| LS-P | 2.43 | 1.47 | 2.57 |
| LS-RP | 1.91 | 1.58 | 2.48 |
| LS-Opt | 4.11 | 3.69 | 2.98 |
| LS-P-Opt | 3.26 | 1.74 | 3.21 |
| LS-RP-Opt | 3.75 | 3.39 | 2.83 |
| NN | 2.61 | 2.02 | 3.57 |

In Table 3.7, we examine the training results using the FOSTER dataset. The best models for SFU, BGU and CAVE were LS, LS-Opt and LS-P-Opt, respectively.

Table 3.7: Mean Delta E 2000 error statistics (FOSTER used as training set).

| **Methods** | **SFU** | **BGU** | **CAVE** |
|---|---|---|---|
| LS | 2.38 | 2.44 | 4.77 |
| LS-P | 2.58 | 2.47 | 4.20 |
| LS-RP | 3.12 | 2.40 | 4.38 |
| LS-Opt | 2.49 | 1.96 | 4.30 |
| LS-P-Opt | 2.82 | 3.01 | 3.96 |
| LS-RP-Opt | 2.91 | 1.98 | 4.18 |
| NN | 5.62 | 5.52 | 4.71 |

Tables 3.4–3.7 reveal several interesting findings. First, when the statistics of training and testing data are different, there is not a clear ranking in the algorithms in general. Second, even though the testing methodology does not contain a change in exposure, we find that the NN method delivers poor

performance, most likely due to overfitting. The algorithm that generalises to other datasets is always one of the classical regression methods and never the NN method.

## 3.6 Conclusion

Recently, it has been proposed that neural networks can be used to solve the colour correction problem (MacDonald and Mayer, 2021). Indeed, in line with previous work, we found the **NN** approach delivered a modest performance increment compared to the (almost) universally used linear correction method (at least where training and testing data contain similar spectral data). However, we also found that the **NN** approach was not exposure invariant. Specifically, a network trained for one light could actually deliver poor colour correction as the exposure changed (there was more or less light in the scene).

However, we showed that NNs could be made robust to changes in exposure through data augmentation by training the NNs with data drawn from many different light levels. In a second approach, we redesigned the neural network architecture so that, by construction, it was exactly exposure invariant. Experiments demonstrated that both exposure-invariant networks continued to outperform linear colour correction. However, a classical method - the simple exposure-invariant root-polynomial regression method - worked best overall (outperforming the **NN** by about 25%).

Finally, we carried out cross-dataset experiments to test the performance of our algorithms. This is a stringent test as the spectral statistics of the datasets used are quite different from one another. Our results showed that there is no clear ranking in the performance of regression-based methods for the cross-dataset condition. However, as for the exposure change test, we found that the NN method performed worse overall.

The general conclusion of this chapter is that - at least for now - classical colour correction regression methods outperform the tested NN algorithm.

Having addressed colour correction, the next chapter shifts the focus to the spectral recovery domain, introducing the Matrix-R framework to enforce physical constraints in spectral reconstruction and pan-sharpening.

# Chapter 4

# Matrix-R Theory: A Simple Generic Method To Improve RGB-Guided Spectral Recovery Algorithms

The content of this chapter is based on a published conference paper (Lin et al., 2023).

While the previous chapter addressed color correction problem, this chapter tackles spectral recovery topic. Here, we revisit the Matrix-R algorithm to demonstrate how enforcing fundamental physical constraints can systematically enhance the accuracy of existing spectral reconstruction and pan-sharpening methods.

RGB-Guided spectral recovery algorithms include both spectral reconstruction (SR) methods that map image RGBs to spectra and pan-sharpening (PS) methods, where an RGB image is used to guide the upsampling of a low-resolution spectral image. In this chapter, we exploit Matrix-R theory in developing a post-processing algorithm that, when applied to the outputs of any and all spectral recovery algorithms, almost always improves their spectral recovery accuracy (and never makes it worse). In Matrix-R theory, any spec-

trum can be decomposed into a component - called the fundamental metamer - in the space spanned by the spectral sensitivities and a second component - the metameric black - that is orthogonal to this subspace. In the proposed post-processing algorithm, we substitute the correct fundamental metamer - which we calculate directly from the RGB image - for the estimated (and generally incorrect) fundamental metamer that is returned by a spectral recovery algorithm. Significantly, it is proven that substituting the correct fundamental metamer always reduces the recovery error. The concept of this substitution, along with its mathematical proof, was initially introduced for spectral reconstruction (Lin, 2023). This chapter reviews and significantly extends its application. We demonstrate its utility for pan-sharpening tasks, including its adaptation for multispectral images. A key contribution is also showing how our Matrix-R post-processing algorithm can effectively leverage an additional physical constraint when the spectra in a target application are known to be accurately described by a low-dimensional linear model. In experiments, we demonstrate that the proposed Matrix-R post-processing improves the performance of a variety of spectral reconstruction and pan-sharpening algorithms.

## 4.1 Introduction

Compared to RGB cameras where there are only 3 values per pixel (Chakrabarti and Zickler, 2011), hyperspectral and multispectral cameras record more detailed spectral signatures from a scene. The additional information in a multi- or hyper-spectral capture has been shown to be important in applications ranging from medical imaging (Lv et al., 2021; Courtenay et al., 2021), remote sensing (Wang et al., 2021; Torun and Yuksel, 2021), food processing (Chen et al., 2021; Gomes et al., 2021; Pane et al., 2021) and art conservation (Picollo et al., 2020; Grillini et al., 2020). However, the higher price tag, lower spatial resolution, longer integration time and/or bulkiness of spectral imagers limit their practical use.

There are many algorithms - exploiting statistical regression and machine learning - that attempt to recover high-quality spectral images from (or with the help of) the RGB images. In Spectral Reconstruction (**SR**), hyperspectral images are recovered directly from their RGB image counterparts. Here, a ground-truth dataset of paired hyperspectral and RGB data is used to train the SR method. Example approaches include regression (pixel-based one-to-one mapping) (Heikkinen et al., 2008; Aeschbacher et al., 2017; Nguyen et al., 2014; Lin and Finlayson, 2023; Arad et al., 2022) and deep-learning-based algorithms (patch-by-patch mapping) (Arad et al., 2022; Shi et al., 2018; Li et al., 2020b).

In RGB pan-sharpening (**PS**), a low-resolution hyperspectral or multispectral image is upsampled to full resolution using a full-resolution RGB image as a guide. The term "sharpened" comes from the fact that if we naively upsampled the images (e.g., using bilinear upsampling) the spectral image would appear blurred relative to the RGB counterpart. When pan-sharpening works well it *looks like* the low-resolution spectral image has been sharpened.

There are two variants of RGB-guided pan-sharpening. When we upsample a low-resolution hyperspectral image (where finely sampled spectra are measured at every pixel), we call it *hyperspectral* pan-sharpening. Often, however, the image we wish to upsample is still a multichannel image but with more channels than a 3-channel RGB image. In this case, we call this *multispectral* pan-sharpening. While the image data is different, the algorithms themselves can often be applied to both the hyper- and multi-spectral capture scenarios, e.g., (Lanaras et al., 2015; Yokoya et al., 2011; Liu et al., 2022; Hu et al., 2022a). Together, SR and PS are examples of RGB-guided spectral recovery algorithms.

Unlike most recent works in spectral recovery, in this chapter, we take a step back and ask a fundamental question: "Given a recorded RGB response and assuming the camera sensitivities are known, are there fundamental properties that any recovered spectrum must adhere to?" In 1953, Wyszescki

(Wyszecki, 1953) first described that each radiance spectrum is composed of a fundamental component intrinsic to its RGB tristimulus response (later called the "fundamental metamer") and its "metameric black". The fundamental metamer integrates to the same given RGB and the black component integrates to zero RGB, [0, 0, 0] (which is where 'black' comes from).

Given an RGB of a spectrum and the device spectral sensitivities we can find the *actual* fundamental metamer defined to be the spectrum in the span of the spectral sensitivities of the camera sensors that projects to the given RGB. Then, we call the projection of a given estimated spectrum (estimated by an SR or PS algorithm) onto the same 3-dimensional spectral subspace spanned by the spectral sensitivities the *estimated* fundamental metamer. We are being careful in our definitions here as, generally, a spectral recovery algorithm - even though the actual RGB is known - will recover a spectrum where the estimated fundamental metamer is not equal to the actual fundamental metamer. One consequence of this result is that when the estimated spectrum - from a given input RGB - is numerically integrated with the camera sensitivities, the calculated output RGB will not be the same as the input (Lin and Finlayson, 2020b; Arad et al., 2020). This also means that the estimated spectrum - for most prior-art algorithms - *must* be the wrong answer.

As we further apply Matrix-R theory to the application of spectral recovery, we learn that a given RGB suggests its corresponding spectrum must have a unique fundamental metamer but can have different metameric blacks. This said, we would argue that the problem of spectral recovery *should be* about recovering the metameric black because for a given RGB the fundamental metamer is uniquely prescribed by the - assumed known - spectral sensitivities of the camera. Yet, curiously, the vast majority of algorithms, e.g., (Arad et al., 2022, 2020, 2018; Hu et al., 2022a; Dong et al., 2021; Zhang et al., 2020), formulate spectral recovery as minimizing a figure of merit (e.g., RMSE) for a given dataset. And, in so doing, the individual recovered spectra can have the wrong estimated fundamental metamers. In a couple of recent works, the

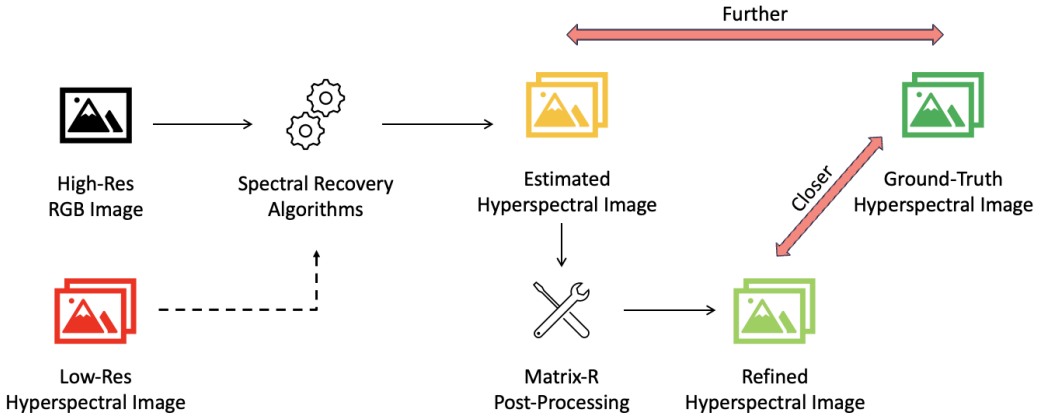


**Figure 4.1:** *We demonstrate how the Matrix-R post-processing method works. First, either RGB images alone (in the spectral reconstruction case) or RGB images combined with low-resolution hyperspectral images (in pan-sharpening) are fed into existing spectral recovery algorithms. The output images are then enhanced by the Matrix-R post-processing algorithm. These refined images consistently achieve greater accuracy and are closer to the ground-truth compared to the initial estimates.*

idea that spectral reconstruction should focus on recovering the metameric black has been investigated with promising results (Lin and Finlayson, 2020b; Stiebel et al., 2020).

A preferred approach (Lin, 2023), the Matrix-R post-processing concept, was introduced as an alternative to re-architecting and retraining existing spectral reconstruction algorithms. This chapter expands significantly on that foundation by, firstly, adapting and applying this methodology to pan-sharpening. Secondly, we demonstrate its successful generalization to multispectral images, making the benefits of Matrix-R post-processing accessible for both hyperspectral and multispectral data. Finally, we show that the method's performance can be further improved by representing the data with a low-dimensional linear model. We illustrate the method in Figure 4.1. Here, the output of the existing spectral recovery algorithms is refined by the Matrix-R post-processing step, bringing it closer to the ground-truth hyperspectral images. Low-resolution hyperspectral images are indicated with a dashed arrow, as they are only used by the pan-sharpening algorithms; otherwise, if high-resolution RGB images alone are used, it is referred to as spectral reconstruction.

In more technical terms, Figure 4.2 illustrates how the post-processing

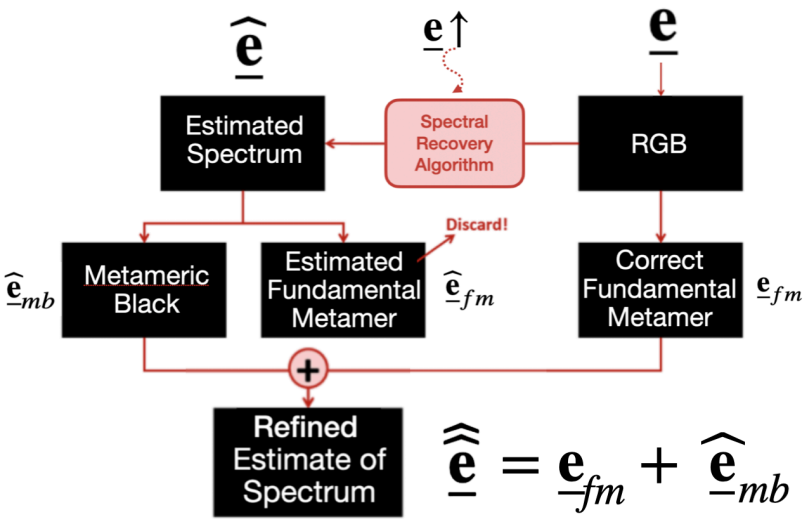


**Figure 4.2:** *A spectrum of light measured by a camera results in an RGB. In SR, an estimated spectrum is returned directly from analyzing the RGB image. In PS an upsampled (see up arrow in diagram) low-res hyperspectral image can also guide spectral estimation. We group PS and SR algorithms in the single 'Spectral Recovery Algorithm' box. The estimated spectrum is decomposed into estimated metameric black and fundamental metamer components. Combining the actual fundamental metamer - calculated directly from the RGB - with this estimated metameric black returns a refined estimate of the spectrum. Refining a spectral estimate in this way is called "Matrix-R post-processing". See text for a description of the mathematical notation.*

method is deployed. First, we denote the spectrum that forms an RGB - the ground-truth - as $\underline{\mathbf{e}}$ (a spectrum is represented by a vector of measurements made across a range of sample wavelengths). A spectral reconstruction or pan-sharpening algorithm returns an estimated spectrum $\underline{\hat{\mathbf{e}}}$ (where, optionally, pan-sharpening also takes input from an upsampled spectrum $\underline{\mathbf{e}}\uparrow$ from a low-res hyperspectral image that is available). In the post-processing method, the spectral estimate is uniquely decomposed into estimated fundamental metamer and metameric black components, respectively denoted $\underline{\hat{e}}_{fm}$ and $\underline{\hat{e}}_{mb}$. The actual fundamental metamer can be calculated directly from the RGB (assuming the spectral sensitivities are known) and is denoted $\underline{e}_{fm}$. According to the method, the refined spectral estimate is calculated as $\underline{\hat{\hat{e}}} = \underline{e}_{fm} + \underline{\hat{e}}_{mb}$. A key result of this chapter is to prove that the refined estimate must be at least as close to the actual ground truth than the original estimate made by the SR or PS algorithms and it is, empirically, often much closer. This post-processing is

generic and can be applied to all algorithms - using classical or deep learning approaches - reported in the literature.

We call the method "Matrix-R post-processing" because the algorithm illustrated in Figure 4.2 depends on a particular projector matrix $\mathbf{R}$ (Wyszecki, 1953) and the post-processing can be described in terms of simple matrix multiplications in terms of $\mathbf{R}$. The operation of matrix $\mathbf{R}$ is summarised in the next section, and the post-processing algorithm and a proof of its efficacy are presented in Section 4.3.

We go on to develop the underlying theory when additional constraints are known about the spectra in a scene. Specifically, it is well known that spectral reflectances are smooth and are well described by low-dimensional linear models (of dimension around 6 to 8 (Parkkinen et al., 1989; Chen et al., 2008)). Concomitantly, the spectra in a scene illuminated by a single dominant light the radiance spectra will have the same dimension as the reflectances (though they will span a different subspace as light spectra are often not smooth). We show how we extend our method to incorporate this linear basis constraint into the Matrix-R theory.

We empirically test our generalised post-processing algorithm on several spectral reconstruction and spectral pan-sharpening algorithms. In all cases, we reduce the recovery error. When the spectra in a scene belong to a low-dimensional linear basis, the post-processing algorithm delivers an even larger reduction in the recovery error.

We also consider the case where we attempt not to recover full spectra but rather a multispectral representation (i.e., the multispectral pan-sharpening), where given an RGB guide and an $m$ channel low-dimensional measurement of the scene (where $m > 3$), we seek to recover the full-res $m$-channel multispectral image. We show how the developed theory can also be applied in this case and include a final small experimental section as a proof of concept.

## 4.2 Background

### 4.2.1 Colour and spectral image formation

A radiance spectrum reflected/emitted from a scene is written as the continuous spectral function $E(\lambda)$. Using a hyperspectral imaging device, we can measure $E(\lambda)$ at finely-sampled wavelengths.

Assuming we sample $n$ points within the 400 to 700 nm visible range ($n \gg 3$), we get an $n$-dimensional vector of measurements $\underline{e} = [E(\lambda_1), E(\lambda_2), \cdots, E(\lambda_n)]^T$. Often $n = 31$, where 10nm sampling is used (Arad et al., 2018, 2020, 2022). Here and throughout this chapter, $^T$ denotes the transpose operator.

In contrast, either an RGB or a multispectral camera uses multiple coloured sensors with different spectral sensitivities. Let us denote the $k$th-channel spectral sensitivity as $Q_k(\lambda)$ (a function of wavelength), we write (Wandell, 1987):

$$\int_{\omega} E(\lambda) Q_k(\lambda) d\lambda = \rho_k \ . \tag{4.1}$$

Here, $\rho_k$ is the $k$th-channel camera response depending on $Q_k$. For the RGB camera, $k = 1, 2, 3$, meaning three different colour sensors are used, whereas for a multispectral camera, $m$ sensor functions will be considered (where $m > 3$). In this chapter, we consider $\omega$, the range of integration, to be the visible range.

A discrete variant of Equation (4.1) is (Wandell, 1987):

$$\mathbf{Q}^T \underline{e} = \underline{\rho} \tag{4.2}$$

where the columns of $\mathbf{Q}$ are the discretized $Q_k(\lambda)$'s ($\mathbf{Q}$ is an $n \times m$ matrix). When $m = 3$, $\underline{\rho} = [\rho_1, \rho_2, \rho_3]^T$ is an RGB vector. In the case of multispectral imaging, in this chapter, we will denote the resulting camera response vector as $\underline{c}$ instead of $\underline{\rho}$ for distinction. The vectors $\underline{\rho}$ and $\underline{c}$ will always, respectively denote 3 and $m$ dimensional response vectors (where $m > 3$).

This chapter aims to study how we could improve spectral recoveries when

the camera spectral sensitivities $\mathbf{Q}$, are "known". Therefore, $\mathbf{Q}$ is assumed to have been measured, using techniques such as a spectral-scanning monochromator (Jiang et al., 2013).

### 4.2.2 Matrix-R

The eponymous "Matrix-R" is the matrix that projects any spectra onto the column space of $\mathbf{Q}$ (the projection is in the span of $\mathbf{Q}$ and is closest in a least-squares sense). In linear algebra, this *projection matrix* is written as (Carl, 2000):

$$\mathbf{R} = \mathbf{Q}[\mathbf{Q}^T\mathbf{Q}]^{-1}\mathbf{Q}^T \tag{4.3}$$

Using the projector matrix $\mathbf{R}$, we calculate the component of a given spectrum $\underline{e}$ - the (actual) fundamental metamer $\underline{e}_{fm}$ - that lies in the column space of $\mathbf{Q}$:

$$\begin{aligned} \underline{e}_{fm} &= \mathbf{R}\underline{e} \\ &= \mathbf{Q}[\mathbf{Q}^T\mathbf{Q}]^{-1}\mathbf{Q}^T\underline{e} \\ &= \mathbf{Q}[\mathbf{Q}^T\mathbf{Q}]^{-1}\underline{\rho} \end{aligned} \tag{4.4}$$

It is clear that:

1. $\mathbf{Q}^T\underline{e}_{fm} = \mathbf{Q}^T\underline{e} = \underline{\rho}$, meaning that $\underline{e}_{fm}$ has the same RGB sensor response as $\underline{e}$.

2. $\underline{e}_{fm}$ is fixed for all spectra satisfying $\mathbf{Q}^T\underline{e} = \underline{\rho}$ (all spectra that return the same colour when observed by the camera sensitivities $\mathbf{Q}$).

3. $\underline{e}_{fm}$ can be exactly calculated given camera's spectral sensitivities $\mathbf{Q}$ and the RGB sensor response $\underline{\rho}$.

The residual - or metameric black - component of $\underline{e}$ is denoted $\underline{e}_{mb}$ and is calculated as:

$$\underline{e}_{mb} = \underline{e} - \underline{e}_{fm} \ = [\mathbf{I} - \mathbf{R}]\underline{e} \tag{4.5}$$

here $\mathbf{I}$ is the $n \times n$ identity matrix. The term metameric black is used because

the camera's response to this signal is zero:

$$\mathbf{Q}^T \underline{e}_{mb} = [0, 0, 0]^T \tag{4.6}$$

In the language of linear algebra, we say $\underline{e}_{mb}$ lies in the *null space* of the column space of $\mathbf{Q}$. And, this null space is in fact the residual $n - 3$ dimensions in the spectral space that are perpendicular to the 3-dimensional camera sensor subspace spanned by columns of $\mathbf{Q}$ (Carl, 2000; Lin and Finlayson, 2020a). The matrix $[\mathbf{I} - \mathbf{R}]$ is the projection matrix for this null space.

Unlike $\underline{e}_{fm}$ which can be calculated directly from the RGB, $\underline{e}_{mb}$ is *unbounded* by the colour image formation. Indeed, $\underline{e}_{mb}$ can be any vector in the $(n - 3)$-dimensional null space of $\mathbf{Q}$ without altering the RGB observation $\underline{\rho}$.

### 4.2.3 Spectral reconstruction

In spectral reconstruction (SR), high-resolution hyperspectral images are directly recovered from their RGB counterparts (Fig. 4.3, red box). One of the simplest SR methods is linear regression (Heikkinen et al., 2008), where a matrix transformation is found that maps the 3-dimensional RGBs to $n$-dimensional spectra. A simple extension to linear regression is to map each RGB to higher dimensional terms, e.g., using a polynomial expansion, (Connah and Hardeberg, 2005; Lin and Finlayson, 2019). These regression models are - in effect - a form of look-up-table since each RGB will always be mapped to the same output spectrum. Other more general one-to-one mappings reported in the literature include radial basis function regression (Nguyen et al., 2014), A+ sparse coding (Aeschbacher et al., 2017) and A++ sparse coding method (Lin and Finlayson, 2023). Significantly, A++ delivers better recovery performance than many of the leading deep-learning methods.

Recent SR methods are often based on machine learning and Deep Neural Networks. Leading methods include HSCNN-D (Shi et al., 2018), HSCNN-R (Shi et al., 2018) and AWAN (Li et al., 2020b). HSCNN-D and HSCNN-R

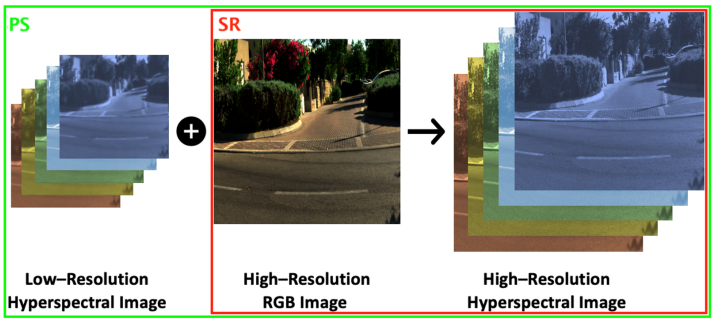


**Figure 4.3:** *The illustration of the RGB-based hyperspectral pan-sharpening (PS; green box) and spectral reconstruction (SR; red box). The images are generated from the ICVL hyperspectral image database (Arad and Ben-Shahar, 2016). Left: the demonstration of the low-resolution hyperspectral image. Center: the high-resolution RGB image. Right: the target high-resolution hyperspectral image.*

are respectively the winner and runner-up of the NTIRE 2018 competition on SR (Arad et al., 2018), where the former adopted a densely-connected convolutional network (Huang et al., 2017), and the latter is based on the deep residual network architecture (He et al., 2016). Then, the AWAN method is the winner of the 2020 edition of the NTIRE competition (Arad et al., 2020), where the non-local attention mechanism (Xia et al., 2019) is incorporated.

Several classical SR methods in the literature make simplifying assumptions as an aid to recovering spectra from RGBs. Beginning with Maloney and Wandell (Maloney and Wandell, 1986), early SR approaches, e.g., (Drew and Funt, 1992), represented spectra with a 3-dimensional linear model. With respect to this 3-D model the spectral weights - a 3-dimensional vector - were shown to be in linear relation to the recorded RGBs. It followed that spectral recovery involved simply inverting the linear relation. Morovic and Finlayson (Morovic and Finlayson, 2006) developed a Bayesian framework that is based on known spectral sensitivities and which complies with Matrix-R (in the sense that Matrix-R post-processing does not further improve the estimate). According to this method, any given input RGB, all spectra with the correct fundamental metamer are shown to form a "metamer set" that acts as a constraint for Bayesian inference. The Matrix-R post-processing was applied to spectra recovered by Zhao et al. (Zhao and Berns, 2007) to ensure the correct

fundamental metamer (without considering its optimality).

The idea that any recovered spectrum should integrate to the same (or similar) RGB has recently appeared in the deep-network-based SR literature. The AWAN (Li et al., 2020b) method incorporates a colour (RGB) difference term in its loss function, where the RGBs of ground-truth and reconstructed spectra are calculated via Equation 4.2 with the camera spectral sensitivities and compared. Despite AWAN's effort to lower the colour error, it is still not completely accurate in colour, as shown in (Arad et al., 2020). This means that the fundamental metamers of the recovered spectra by AWAN still do not match with the ones derived from the RGBs. Lin and Finlayson (Lin and Finlayson, 2020b) addressed the general problem of Matrix-R non-compliance in SR by restricting all algorithms to only predict the metameric black components of the ground-truth spectra while keeping the fundamental metamer components identical to the ones derived from the RGBs. Nonetheless, this approach requires retraining of the entire algorithms, and the performance improvement is not guaranteed (Lin and Finlayson, 2020b).

### 4.2.4 Pan-sharpening

In pan-sharpening, we wish to fuse low-resolution hyperspectral or multispectral images with a high-resolution RGB counterpart, see Figure 4.3. Here, left, a low-res hyperspectral image (or relatedly a multispectral image) is fused in some way with the full-resolution RGB image (middle) to, hopefully, produce a good estimated high-res image (right). We note that in the pan-sharpening literature (especially older algorithms), imaging systems often only have access to a *grayscale* full-res image instead of RGB. Here and throughout this chapter, we only consider the pan-sharpening problem guided by RGB input.

Interestingly, many of the prior-art methods generate their full-res outputs without formally considering whether their algorithm is accurate or not. That is because these algorithms were not trained against a benchmark ground-truth

dataset. Prominent examples of this approach include the Coupled Nonnegative Matrix Factorization (CNMF) (Yokoya et al., 2011) and the coupled spectral unmixing methods (Lanaras et al., 2015). The latter method can be seen as an improvement of the former where physical constraints are placed on the spectra that are recovered.

In more recent research, e.g., (Hu et al., 2022a, 2021), has introduced algorithms that leverage the knowledge of ground-truth data. These methods involve the use of deep neural networks (DNNs) to map RGB images combined with low-resolution spectral input to high-resolution ground-truth outputs. Once trained, these networks can be applied to unseen data for predictions. However, one significant drawback of the DNN approach is its reliance on millions of parameters, which can make the models highly complex and computationally intensive. Moreover, there is often an insufficient amount of training data available to reliably train these networks, raising concerns about generalization (Lin and Finlayson, 2023). Another challenge arises when new training data become available or when researchers switch to different camera sensitivities (Lin and Finlayson, 2022). In such cases, the entire network may need to be re-trained, which can be a time-consuming and resource-intensive process.

There are hybrid methods such as the Model-Inspired Autoencoder (MIAE) (Liu et al., 2022) which are still based on finding the prior spectral and spatial prior to solve a Non-Negative Matrix Factorization problem on per-scene basis, while formulating the prior-crafting as a deep learning problem. A key component of MIAE relevant to the research we report in this chapter is that this algorithm exploits knowledge of the camera's spectral sensitivities. As MIAE also delivers leading performance results, we will use it here as an exemplar deep learning algorithm to benchmark against.

Finally, we note that there have been methods that incorporate knowledge of the RGB camera spectral sensitivities to refine their pan-sharpening method. Most notably, Imai and Burns (Imai and Berns, 1998) directly ap-

plied Matrix-R compliance as a standalone pan-sharpening algorithm: the lower-resolution hyperspectral image is first resized (upsampled) to the same image dimension as the RGB image, and then, at each pixel, we replace the fundamental metamer component of the low-resolution spectrum by the one calculated from the RGB image. Essentially, our method extends the work of Imai and Burns. We also prove that post-processing with Matrix-R must always result in improved spectral estimation for any SR or PS algorithm. This is an important point as the Imai and Berns method - viewed as the vantage point of the performance afforded by today's most effective algorithms - delivers relatively poor performance (Matrix-R alone does not suffice). Finally, we extend the Matrix-R theory so it can be more powerfully applied when we know something about the lower-dimensional spectral subspace where scene spectra lie, which is a key innovation to obtaining the best performance.

## 4.3 Matrix-R Post-processing

### 4.3.1 Matrix-R post-processing for improving hyperspectral recovery

Let us denote $\underline{e}$ as the ground-truth spectrum at a pixel (measured by a hyperspectral imager) and $\widehat{\underline{e}}$ a spectral estimate made using a PS (pan-sharpening) or SR (spectral reconstruction) algorithm. Here, we will use the convention that the overscript $\widehat{\ }$ denotes an estimated quantity. Since the Matrix-R method (see Figure 4.2) refines an estimate to bring an estimate closer to the actual (ground-truth) spectrum, we use the *double-hat* $\widehat{\widehat{\ }}$ notation to denote the refined estimate.

*A priori*, we can write the actual spectrum of light $\underline{e}$ and its estimate $\widehat{\underline{e}}$ (from a given SR or PS algorithm) as sums of fundamental metamers and

metameric blacks (see section 5.1):

$$\begin{cases} \underline{e} = \underline{e}_{fm} + \underline{e}_{mb} \\ \widehat{\underline{e}} = \widehat{\underline{e}}_{fm} + \widehat{\underline{e}}_{mb} \end{cases} \tag{4.7}$$

The subscript $_{fm}$ (for fundamental metamer) indicates the part of a spectrum in the span of the spectral sensitivities of a camera system, and $_{mb}$ for the part of the spectrum lying in the null-space of the camera.

While we do not know the ground-truth spectrum $\underline{e}$ in practice, we can still calculate $\underline{e}_{fm}$ from the input RGB $\underline{\rho}$, as shown in Equation (4.4). Significantly, in almost all data-driven PS and SR algorithms, the estimated and actual fundamental metamers are not equal to one another: $\widehat{\underline{e}}_{fm} \neq \underline{e}_{fm}$.

In this chapter, we investigate using the Matrix-R theory (summarised in Figure 4.2), we can refine the estimate $\widehat{\underline{e}}$ to calculate $\widehat{\widehat{\underline{e}}}$ where the refined estimate is closer to the ground-truth. In mathematical terms, the refinement process is written as:

$$\begin{aligned} \widehat{\widehat{\underline{e}}} &= \widehat{\underline{e}} - \widehat{\underline{e}}_{fm} + \underline{e}_{fm} \\ &= \underline{e}_{fm} + \widehat{\underline{e}}_{mb} \end{aligned} \tag{4.8}$$

**Theorem 4.1.** *The refined output, $\widehat{\widehat{\underline{e}}}$, is always as close or closer to the ground-truth $\underline{e}$ than the initial estimate $\widehat{\underline{e}}$, i.e., $||\underline{e} - \widehat{\widehat{\underline{e}}}|| \leq ||\underline{e} - \widehat{\underline{e}}||$ (where $||\cdot||$ denotes the L-2 norm).*

*Proof.* Let us denote $\widehat{\Delta} = ||\underline{e} - \widehat{\underline{e}}||^2$ and $\widehat{\widehat{\Delta}} = ||\underline{e} - \widehat{\widehat{\underline{e}}}||^2$ . Clearly, the theorem will be proved if we prove $\widehat{\widehat{\Delta}} \leq \widehat{\Delta}$ .

First, let us consider $\widehat{\Delta}$ with respect to the fundamental metamer and

metameric black decomposition:

$$\begin{aligned}\widehat{\Delta} &= ||\underline{e} - \widehat{\underline{e}}||^2 \\ &= ||(\underline{e}_{fm} + \underline{e}_{mb}) - (\widehat{\underline{e}}_{fm} + \widehat{\underline{e}}_{mb})||^2 \\ &= ||(\underline{e}_{fm} - \widehat{\underline{e}}_{fm}) + (\underline{e}_{mb} - \widehat{\underline{e}}_{mb})||^2 \\ &= ||\underline{e}_{fm} - \widehat{\underline{e}}_{fm}||^2 + ||\underline{e}_{mb} - \widehat{\underline{e}}_{mb}||^2 \\ &\quad + 2 \cdot [\underline{e}_{fm} - \widehat{\underline{e}}_{fm}]^T [\underline{e}_{mb} - \widehat{\underline{e}}_{mb}]\end{aligned} \tag{4.9}$$

Here, the cross-term:

$$[\underline{e}_{fm} - \widehat{\underline{e}}_{fm}]^T [\underline{e}_{mb} - \widehat{\underline{e}}_{mb}] = 0 \tag{4.10}$$

Indeed, because both $\underline{e}_{fm}$ and $\widehat{\underline{e}}_{fm}$ lie in the spectral subspace spanned by columns of $\mathbf{Q}$, $[\underline{e}_{fm} - \widehat{\underline{e}}_{fm}]$ is also a vector in this subspace; on the other hand, $[\underline{e}_{mb} - \widehat{\underline{e}}_{mb}]$ is a vector lies in the null-space of $\mathbf{Q}$ (Carl, 2000). Substituting Equation (4.10) into Equation (4.9), we get:

$$\widehat{\Delta} = ||\underline{e}_{fm} - \widehat{\underline{e}}_{fm}||^2 + ||\underline{e}_{mb} - \widehat{\underline{e}}_{mb}||^2 \tag{4.11}$$

Next, let us examine $\widehat{\widehat{\Delta}}$:

$$\begin{aligned}\widehat{\widehat{\Delta}} &= ||\underline{e} - \widehat{\widehat{\underline{e}}}||^2 \\ &= ||(\underline{e}_{fm} + \underline{e}_{mb}) - (\underline{e}_{fm} + \widehat{\underline{e}}_{mb})||^2 \\ &= ||\underline{e}_{mb} - \widehat{\underline{e}}_{mb}||^2\end{aligned} \tag{4.12}$$

Following from Equations (4.11) and (4.12), it is immediate that:

$$\begin{aligned}\widehat{\widehat{\Delta}} &= ||\underline{e}_{mb} - \widehat{\underline{e}}_{mb}||^2 \\ &\leq ||\underline{e}_{fm} - \widehat{\underline{e}}_{fm}||^2 + ||\underline{e}_{mb} - \widehat{\underline{e}}_{mb}||^2 \\ &\leq \widehat{\Delta}\end{aligned} \tag{4.13}$$

□

Equation (4.13) succinctly encapsulates that the recovered spectrum post-processed using the Matrix-R method is always as close or closer to the ground truth (compared to the original recovered spectrum returned by any PS or SR algorithm).

### 4.3.2 Generalisation of Matrix-R post-processing to multispectral recovery

In Equations (4.1) and (4.2), we proposed to sample spectra (e.g., from 400 nm to 700 nm) at $n$ wavelengths. What if $\mathbf{Q}$ and $\underline{e}$ were sampled at $2n$ or $10n$ wavelengths? Would that change any of the arguments? No, it would not - so long as Equation (4.2) remains a valid physical model of how RGBs are formed. Now, suppose we think of the columns of $\mathbf{Q}$ and $\underline{e}$ not as discrete spectral measurements but as some other functions of wavelength that still satisfy Equation (4.2). Because none of our derivations depends on the physical *meaning* of the integration/inner product step entailed in Equations (4.1) and (4.2), all the methods developed so far continue to work. We can still find fundamental metamers and metameric blacks, which lie in the space spanned by $\mathbf{Q}$ or in its null space, respectively. However, these metameric concepts are no longer linked to wavelength.

Let us make this abstract idea more concrete. We denote the $m$-dimensional measurements made by a multispectral imager ($m > 3$), as $\underline{c}$. Now, we assume there is a linear relationship between $\underline{c}$ and the RGB response, $\underline{\rho}$. Our imaging model (in direct analogy to Equation (4.2)) is:

$$\mathbf{M}^T\underline{c} = \underline{\rho} \tag{4.14}$$

It follows that for the multispectral reconstruction and pan-sharpening problem we can still apply the same Matrix-R post-processing developed thus far. All that is changed is that the matrix $\mathbf{R}$ now depends on $\mathbf{M}$ (rather than $\mathbf{Q}$). An important detail is that $\mathbf{M}$ is $m \times 3$ where $3 < m \ll n$ (a point we return

to later).

Of course, the Matrix-R theory will apply if and only if Equation (4.14) is a good model of image formation. How might we find $\mathbf{M}$ in practice? Let $\mathbf{C}$ and $\mathbf{P}$ denote, respectively, $N \times m$ and $N \times 3$ matrices of corresponding multispectral and RGB sensor responses for $N$ training stimuli. We then find $\mathbf{M}$ using a regularized least-squares regression:

$$\arg\min_{\mathbf{M}} \|\mathbf{CM} - \mathbf{P}\|_F^2 + \gamma\|\mathbf{M}\|_F^2 \tag{4.15}$$

where $\|\cdot\|_F$ represents the Frobenius norm (Horn and Johnson, 1990). The user-defined $\gamma$, bounds the magnitude of $\mathbf{M}$, effectively, mitigating overfitting (Tikhonov et al., 1995; Webb, 2010). Equation (4.15) is solved in closed form (Tikhonov et al., 1995; McDonald, 2009):

$$\mathbf{M} = \left[\mathbf{C}^\top\mathbf{C} + \gamma\mathbf{I}\right]^{-1}\mathbf{C}^\top\mathbf{P} \tag{4.16}$$

here $\mathbf{I}$ is the $m \times m$ identity matrix. The $\gamma$ terms is user-defined. We have found that setting gamma to be a small fraction - say 0.01% - of the mean variance of the data: $trace(\mathbf{C}^\top\mathbf{C})/N$ , works well for our purposes.

### 4.3.3 Matrix-R post-processing with a low-dimensional spectral representation

Now, let us suppose all spectra in a target hyperspectral image lies in a lower-dimensional space. We write:

$$\underline{e} = \mathbf{B}\underline{\alpha} \tag{4.17}$$

where $\mathbf{B}$ is an $n \times b$ basis matrix ($b < n$), and $\underline{\alpha}$ is a coefficient vector with $b$ components. Here, for convenience for later derivations, we further assume the *orthonormalization* of the columns of $\mathbf{B}$, i.e., the columns of $\mathbf{B}$ are normalized to unit vector and are orthogonal to each other. This can be achieved in vari-

ous ways given any $b$-dimensional basis, e.g., using the Gram-Schmidt process (Cheney and Kincaid, 2009). To ease notation we will still write spectra as $\underline{e}$ where we implicitly assume the linear model assumption.

Next, the definition of fundamental metamer (and metameric black) can then be defined in terms of the interaction of the $\mathbf{B}$ subspace and the camera spectral sensitivities (and so we will draw attention to this fact in our notation). We point out that only the part of $\mathbf{Q}$ spanned by the basis $\mathbf{B}$ contributes to the RGB observations of any spectra written in the form of Equation (4.17). Indeed, since $\underline{e}$ lies in the column space of $\mathbf{B}$, the part of $\mathbf{Q}$ perpendicular to $\mathbf{B}$ will have no effects in the colour image formation $\mathbf{Q}^T\underline{e}$ (Equation (4.2)). Given this prior knowledge, we can now define a new data-dependent spectral sensitivity matrix:

$$\overline{\mathbf{Q}} = \mathbf{B}\mathbf{B}^T\mathbf{Q} \tag{4.18}$$

where $\mathbf{B}\mathbf{B}^T$ is the projection matrix with respect to $\mathbf{B}$ (plugging $\mathbf{B}$ into Equation (4.3) returns this projector because $\mathbf{B}$ has orthonormal columns). We can examine the equivalence of $\mathbf{Q}$ and $\overline{\mathbf{Q}}$ in colour image formation by replacing $\mathbf{Q}$ by $\overline{\mathbf{Q}}$ in Equation (4.2):

$$\overline{\mathbf{Q}}^T\underline{e} = \mathbf{Q}^T\mathbf{B}\mathbf{B}^T\underline{e} = \mathbf{Q}^T\underline{e} = \underline{\rho} \tag{4.19}$$

Here, the $\mathbf{B}\mathbf{B}^T$ projection does not alter $\underline{e}$ because $\underline{e}$ already lies in the column space of $\mathbf{B}$, as shown in Equation (4.17). Equation (4.19) ensures that the fundamental metamer of the target ground-truth spectrum $\underline{e}$ with respect to $\overline{\mathbf{Q}}$, denoted as $\underline{e}_{fm}^{\overline{\mathbf{Q}}}$, can also be derived from the RGB camera response $\underline{\rho}$ directly (following Equation (4.4) while replacing $\mathbf{Q}$ by $\overline{\mathbf{Q}}$). In analogy to Equations (4.3) and (4.4), we define a modified matrix $\overline{\mathbf{R}}$ as

$$\overline{\mathbf{R}} = \overline{\mathbf{Q}}[\overline{\mathbf{Q}}^T\overline{\mathbf{Q}}]^{-1}\overline{\mathbf{Q}}^T \tag{4.20}$$

and write the actual and estimated fundamental metamer, $\underline{e}_{fm}^{\overline{\mathbf{Q}}}$ and $\widehat{\underline{e}}_{fm}^{\overline{\mathbf{Q}}}$, re-

spectively as:

$$\begin{cases} \underline{e}_{fm}^{\overline{\mathbf{Q}}} = \overline{\mathbf{R}}\underline{e} = \overline{\mathbf{Q}}[\overline{\mathbf{Q}}^T\overline{\mathbf{Q}}]^{-1}\underline{\rho} \\ \widehat{\underline{e}}_{fm}^{\overline{\mathbf{Q}}} = \overline{\mathbf{R}}\widehat{\underline{e}} \end{cases} \tag{4.21}$$

Similarly, if the readers are interested, the metameric black in this case is found using Equation (4.5) with the replaced $\overline{\mathbf{R}}$ instead of $\mathbf{R}$.

It follows that with respect to $\overline{\mathbf{Q}}$ - analogous to Equation (4.8) - the Matrix-R post-processing for the low-dimensional case, is written as:

$$\widehat{\widehat{\underline{e}}} = \widehat{\underline{e}} - \widehat{\underline{e}}_{fm}^{\overline{\mathbf{Q}}} + \underline{e}_{fm}^{\overline{\mathbf{Q}}} \tag{4.22}$$

**Theorem 4.2.** *Assuming that spectra are in the span of an m-dimensional linear model, the refined spectral estimate* $\widehat{\widehat{\underline{e}}}$*, calculated using Equation (4.22), will always be closer to the ground truth* $\underline{e}$ *than the initial estimate* $\widehat{\underline{e}}$*, i.e.,* $||\underline{e} - \widehat{\widehat{\underline{e}}}|| \leq ||\underline{e} - \widehat{\underline{e}}||$.

We do not need to formally prove the second theorem, as the original Matrix-R theorem does not limit us to use any particular spectral sensitivity matrix $\mathbf{Q}$ for the Matrix-R decomposition. In fact, the theorem holds for any $n \times 3$ matrix that derives RGBs from spectra - which, as Equation (4.19) has shown, applies to both $\mathbf{Q}$ and $\overline{\mathbf{Q}}$.

### Determining the basis B

The basis $\mathbf{B}$ can be calculated *a priori*, e.g., based on known reflectances and a known illuminant. For the PS application, we additionally have access to a low-res hyperspectral image of the scene, and we might extract a low-dimensional image from this image. A third alternative would be to calculate the basis from the high-res spectral reconstruction returned by a given algorithm. In all three cases, given a corpus of spectral measurements, it is easy to find the best least-squares optimal basis using techniques like Characteristic Vector Analysis (Maloney and Wandell, 1986). Importantly, there is a

reasonable expectation that a low-dimensional basis will well describe spectral data. Indeed, most spectral reflectances are smooth functions of wavelength (and are often represented by 6 to 8 basis functions). The dimensionality of observed smooth reflectances under a single non-smooth illuminant does not change.

**Extension to multispectral recovery**

Rather than thinking about image formation in the wavelength domain, we can instead adopt Equation (4.14) as our image formation model (an RGB is a linear sum of the responses from a $m$-sensor imager). With respect to this imager we can again adopt a $b$-dimensional model for spectra. According to these assumptions, we write:

$$\underline{c} = \mathbf{C}\underline{\alpha} \tag{4.23}$$

Here $\mathbf{C}$ is a $m \times b$ orthonormal basis matrix of responses (where - to elicit any computational advantage - we need $b < m$). With respect to this matrix, we can derive a new data-dependent image formation matrix

$$\overline{\mathbf{M}} = \mathbf{C}\mathbf{C}^T\mathbf{M} \tag{4.24}$$

Then the arguments from sections 4.3.2 and 4.3.3 all hold: we need only substitute, respectively, $\underline{c}$ for $\underline{e}$, $\overline{\mathbf{M}}$ for $\overline{\mathbf{Q}}$, and $\mathbf{C}\mathbf{C}^T$ for $\mathbf{B}\mathbf{B}^T$.

## 4.4 Experiments

Table 4.1: List of considered SR and PS algorithms

| Spectral Reconstruction (SR) | Pan-sharpening (PS) |
|---|---|
| 1. A++ (Lin and Finlayson, 2023) | 1. Imai and Burns (Imai and Berns, 1998) |
| 2. HSCNN-R (Shi et al., 2018) | 2. CNMF (Yokoya et al., 2011) |
| 3. HSCNN-D (Shi et al., 2018) | 3. Lanaras et al. (Lanaras et al., 2015) |
| 4. AWAN (Li et al., 2020b) | 4. MIAE (Liu et al., 2022) |

### 4.4.1 Data preparation

We will use the ICVL hyperspectral image database (Arad and Ben-Shahar, 2016) for our experiments. ICVL consists of 201 hyperspectral images of size $1300 \times 1392$ (though a handful of images are slightly smaller) and 31 spectral dimensions (10-nanometer sampling of the visible spectrum 400 and 700 nanometers). The original images are encoded in 12 bits, i.e., the maximal pixel value is 4095. We re-scale the encoding range to [0, 1] by dividing 4095 from the original pixel values.

In our experiments, the RGB images are generated pixel by pixel from the ground-truth hyperspectral images via Equation (4.2). For spectral reconstruction experiments, the CIE 1964 Color Matching Functions (Commission Internationale de L'eclairage, 1964) are used as the camera spectral sensitivities to generate RGB images since this is proposed in (Arad and Ben-Shahar, 2016) and many spectral reconstruction algorithms were developed for this definition of RGB. For our hyperspectral and multispectral pan-sharpening experiments, RGB images are generated with the camera response functions of Canon 1D Mark III (shown in Figure 4.4), and multispectral images with Spectricity's 16-channel multispectral camera sensitivity functions (Spectricity). To generate the low-resolution spectral image input for PS algorithms, both hyperspectral and multispectral images are downsampled by a factor of 8 via bilinear interpolation for pan-sharpening experiments: we simulate a hyperspectral and multispectral *thumbnails* that are 1/64 the size of the original RGB image.

For both spectral reconstruction and pan-sharpening, the root-mean-squared error (RMSE) is used as the recovery error metric.

$$RMSE = \sqrt{\frac{1}{n}||\underline{e} - \underline{e}_{\text{rec}}||^2} \ , \tag{4.25}$$

where $\underline{e}$ is the ground-truth and $\underline{e}_{\text{rec}}$ is the recovered spectra. The recovered spectrum $\underline{e}_{\text{rec}}$ could be the estimate recovered by an algorithm, $\widehat{\underline{e}}$, or the refined

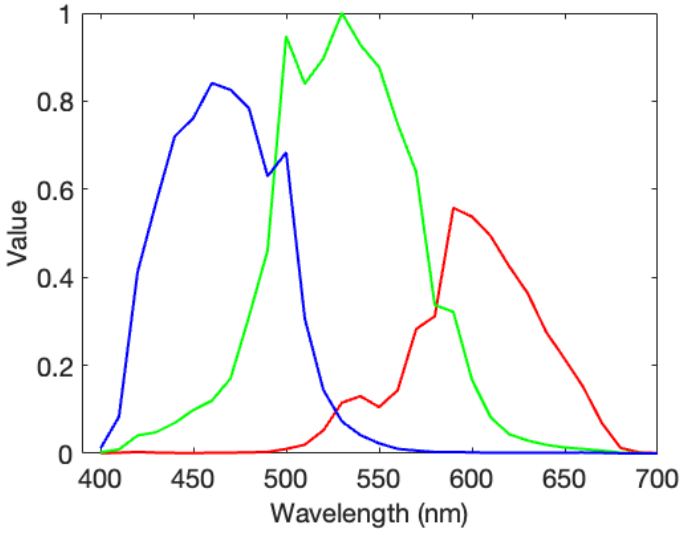


**Figure 4.4:** *Normalised sensitivity functions of Canon 1D Mark III camera.*

estimate found via Matrix-R post-processing, $\hat{\hat{\underline{e}}}$.

### 4.4.2 Spectral reconstruction results

Spectral reconstruction results are summarized in Table 4.2. We consider four algorithms: the leading regression SR method, A++ (Lin and Finlayson, 2023), and three leading deepnets: HSCNN-R, HSCNN-D, and AWAN (Shi et al., 2018; Li et al., 2020b). We applied the cross-validation approach (Lin and Finlayson, 2020c) and reported mean and 99 percentile recovery statistics. The mean recovery error of an image is the mean error of overall image pixels. Then, the mean error shown in Table 4.2 is the mean of these per image means. Similarly, the 99 percentile error is the mean of the 99 percentile recorded per image. The RMSE figures are typically small (our data is in the interval [0,1]), so all RMSE errors in Table 4.2 are multiplied by 1000 for readability. Per column in Table 4.2 (and all following result tables), boldface denotes the experimental condition yielding the best results.

The RMSE performance of the original algorithm (no post-processing) is shown in the top row. Applying Matrix-R post-processing yields the results recorded in the second row of Table 4.2. We see that the performance increment is most significant for the regression-based A++ where the mean and

Table 4.2: The RMSE ($\times 10^3$) performance of the Matrix-R method ("Matrix-R") and its lower-dimensional variants ("$m$-dim") on **spectral reconstruction** algorithms. Best results are shown in bold font and underlined.

| | **A++** | | **HSCNN-R** | | **HSCNN-D** | | **AWAN** | |
|---|---|---|---|---|---|---|---|---|
| | **Mean** | **99pt** | **Mean** | **99pt** | **Mean** | **99pt** | **Mean** | **99pt** |
| Original | 4.10 | 22.26 | 3.98 | 18.83 | 3.57 | 17.77 | **2.26** | 12.69 |
| Matrix-R | 3.93 | 21.18 | 3.96 | 18.79 | **3.48** | 17.64 | **2.26** | **12.68** |
| 3 dim | 21.24 | 84.07 | 12.35 | 51.92 | 9.51 | 40.09 | 14.53 | 66.55 |
| 4 dim | 4.05 | **19.99** | 4.35 | 18.82 | 3.78 | **17.37** | 3.06 | 14.83 |
| 5 dim | 4.03 | 20.93 | 4.09 | 18.77 | 3.62 | 17.73 | 2.49 | 13.13 |
| 6 dim | 3.93 | 21.01 | 3.97 | **18.73** | 3.50 | 17.59 | 2.34 | 12.80 |
| 7 dim | 3.92 | 21.09 | 3.95 | 18.74 | 3.49 | 17.61 | 2.30 | 12.72 |
| 8 dim | **3.91** | 21.12 | **3.95** | 18.75 | **3.48** | 17.62 | 2.28 | 12.70 |
| ⋮ | | | | | | | | |
| 21 dim | - | - | - | - | - | - | **2.26** | **12.68** |

99 percentile error of the original method are respectively 4.2 and 4.9% lower. The gains for the HSCNN networks are less but are still significant. There is a very small improvement (noticeable only in the fourth decimal place) for AWAN (which is to be expected as this network was designed to approximately recover the correct fundamental metamer).

We now adopt the linear basis assumption where, per image, the best basis of a given dimension (the "$m$"-dim) is found via a characteristic vector analysis of the original output spectral image recovered by the SR algorithms. Note this is a strong constraint as we are assuming we have access to the optimal linear basis that describes the spectra we are attempting to recover. Clearly, adopting too few basis vectors leads to an expected decrement in performance, as we see the generally ill-performed 3-dim results. However, in all cases - for all algorithms and error metrics - a linear model assumption exists that leads to better recovery performance.

For the A++ regression method, adopting Matrix-R post-processing together with a linear basis assumption results in, respectively, a 4.6 and 10.2% improvement in the mean and 99 percentile RMSE errors, which is a critical improvement that makes A++ outperform the much more complex HSCNN-R in mean performance.

### 4.4.3 Hyperspectral pan-sharpening results

Here, we have access to the full-resolution RGB image and a 1/8-resolution hyperspectral image. We wish to fuse these two images to recover a full-resolution hyperspectral image. We consider 4 algorithms. First, we only bilinearly upsample the low-resolution hyperspectral image (this is a control for our experiments). Then we benchmark against the classical algorithms: CNMF (Yokoya et al., 2011) and Lanaras et al. (Lanaras et al., 2015). Finally, we look at the performance of MIAE (Liu et al., 2022), one of the leading deep-net pan-sharpening algorithms. The results are summarized in Table 4.3.

Table 4.3: The RMSE ($\times 10^3$) performance of the Matrix-R method ("Matrix-R") and its lower-dimensional variants ("$m$-dim") on **hyperspectral pan-sharpening** algorithms. Best results are shown in bold font and underlined.

| | **Bilinearly Upsample** | | **CNMF** | | **Lanaras et al.** | | **MIAE** | |
|---|---|---|---|---|---|---|---|---|
| | **Mean** | **99pt** | **Mean** | **99pt** | **Mean** | **99pt** | **Mean** | **99pt** |
| Original | 7.94 | 62.06 | 3.98 | 14.34 | 1.96 | 10.54 | 1.37 | 6.19 |
| Matrix-R | 3.54 | 26.51 | 2.92 | 11.19 | 1.71 | 8.70 | 1.36 | 6.14 |
| 3 dim | 13.52 | 63.86 | 13.52 | 63.86 | 13.52 | 63.86 | 13.52 | 63.86 |
| 4 dim | **2.77** | **11.87** | 3.04 | 11.59 | 2.54 | 10.32 | 2.52 | 10.55 |
| 5 dim | 3.33 | 22.37 | **2.83** | 10.72 | 1.90 | **8.42** | 1.69 | 7.08 |
| 6 dim | 3.33 | 24.02 | 2.85 | 10.91 | 1.76 | 8.49 | 1.46 | 6.32 |
| ⋮ | | | | | | | | |
| 10 dim | 3.39 | 25.31 | 2.88 | 11.08 | **1.69** | 8.62 | **1.35** | **6.09** |

Applying our Matrix-R post-processing for bilinear upsampling, the mean RMSE error of 8.75 is reduced to 4.69 (close to a 50 % reduction), and with a 4-dimensional linear model, the mean error was further lowered to 3.87, surpassing the performance delivered by the original CNMF. The Matrix-R post-processing with the respectively best performing low-dimensional linear model also significantly improves the performance of CNMF and Lanaras (their mean RMSE errors were reduced by 29 and 12%, respectively). Benchmarked against MIAE, the performance increment is much more modest. The 99 percentile error improvements follow similar trends.

Note that here, and later in the multispectral PS section, the $m$-dim bases are found by applying characteristic vector analysis on each input low-dimensional hyper- and multi-spectral image, respectively.

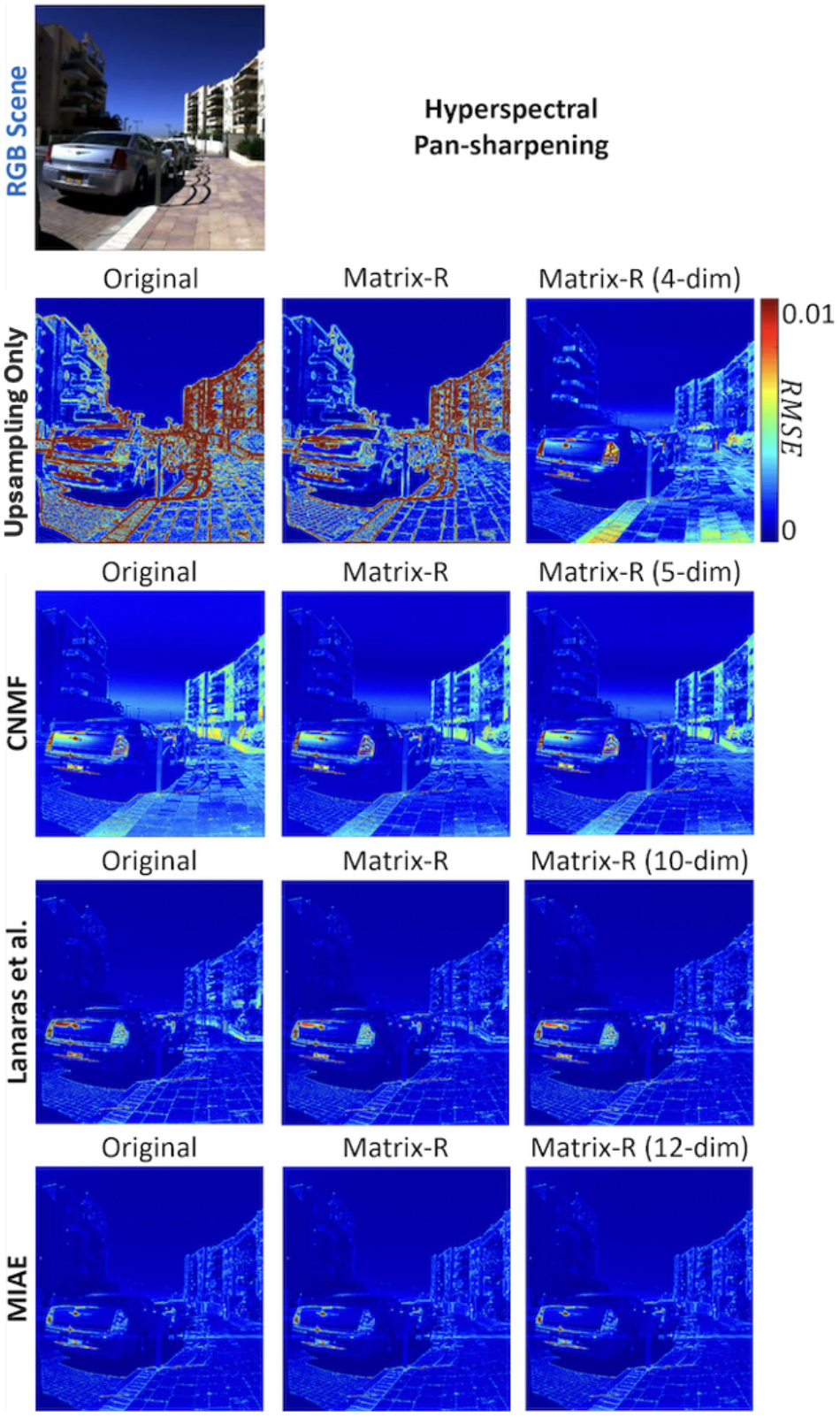


**Figure 4.5:** *The Original, Matrix-R, and Matrix-R with a lower-dimensional spectral assumption results in RMSE error heat maps for the tested hyperspectral pan-sharpening algorithms.*

In Figure 4.5, we visualise the recovery errors for four pan-sharpening algorithms and their post-processing by the Matrix-R algorithm and the Matrix-R with the linear basis constraint. The visualisations reflect the error statistics conveyed in Tables 4.3. There is a large improvement for 'upsampling only' and modest improvements for CNMF and Lanaras et al. For this image, it is hard to visually discern the improvement of post-processing for the MIAE algorithm.

### 4.4.4 Multispectral pan-sharpening results

Finally, we conduct a proof-of-concept experiment on multispectral pan-sharpening, aiming to achieve high-resolution $m$-channel multispectral images by fusing low-resolution multispectral and corresponding high-resolution RGB images. Since this setup lacks a standard method or dataset, we perform hypothetical experiments to test Matrix-R and its lower-dimensional variants for recovering multispectral images. The RGB and multispectral images are generated by integrating the spectral images from the ICVL dataset with, respectively, the Canon 1D Mark III and Spectricity's 16-channel multispectral camera sensitivity functions (Spectricity).

Pan Sharpening for this experimental scenario is discussed in section 4.3.3, and we follow that methodology here. Importantly, to use this method, we need to know how RGBs and multispectral measurements are related to each other. The challenge is that there is no, *a priori*, known direct mapping from multispectral to RGB (from $\underline{c}$ to $\underline{\rho}$ in Equation (4.14)), unlike in hyperspectral cases where the camera sensitivity functions directly mapping hyperspectral data to RGB are available. Thus, we must compute the transformation matrix $\mathbf{M}$ with regularization for multispectral pan-sharpening, i.e., Equation (4.16). To do that, we downscale the RGB images to match the pixels between low-resolution multispectral images for each scene. Then, we regress the multispectral images onto the RGBs and find individual matrix $\mathbf{M}$ per image. In solving the

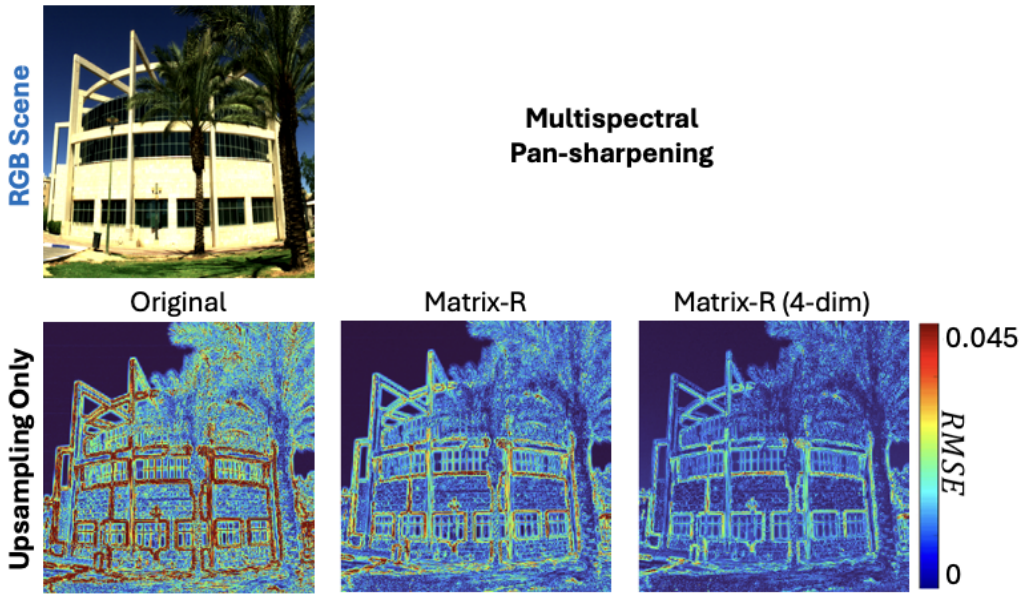


**Figure 4.6:** *The original upsampling-only, post-processing by Matrix-R, and Matrix-R with a lower-dimensional spectral assumption results in RMSE error heat maps for multispectral pan-sharpening.*

regression in Equation (4.16), there is a $\lambda$ parameter controlling the penalty term. In our experiments, $\lambda$ equals to 0.01% of the mean variance of the data: $trace(\mathbf{C}^\top\mathbf{C})/N$.

Results are reported in Table 4.4, where we benchmark our multispectral pan-sharpening approach against the error found when we only bilinearly resize the multispectral image. Clearly, Matrix-R improves the upsampling-only results, and using an $m$-dimensional linear model consistently yields better or equivalent performance compared to standalone Matrix-R when $m > 3$ (again, the $m = 3$ case points to insufficient linear model representation of spectra as in the spectral reconstruction and hyperspectral pan-sharpening results). The best performance is achieved when $m = 5$ for the mean, and $m = 4$ for the 99th-percentile RMSE. In both cases, the errors are reduced by about 75%.

The effectiveness of our method in multispectral pan-sharpening is also illustrated in Figure 4.6.

Table 4.4: The RMSE ($\times 10^3$) performance of the Matrix-R method as an **multispectral pan-sharpening** algorithm and its lower-dim variants ("$m$-dim") on resized multispectral images.

| | **Mean** | **99pt** |
|---|---|---|
| Upsampling Only | 7.09 | 52.62 |
| Matrix-R | 2.09 | 15.81 |
| 3 dim | 32.92 | 167.03 |
| 4 dim | 1.97 | **13.93** |
| 5 dim | **1.93** | 14.54 |
| 6 dim | 2.10 | 15.78 |

### 4.4.5 Discussion

The proposed post-processing Matrix-R method can be applied in a wide context: the proposed process could be used to enhance the performance of off-the-shelf "black box" algorithms where the algorithm source code is not available. Indeed, the theorem does not require knowledge of the algorithm itself. We need only the input RGB and camera sensitivity information for the Matrix-R decomposition. **Theorem 1** also gives the user *comfort*. It does no evil: it will *always* either improve the performance of any algorithm or, failing that, it will not reduce the algorithm's performance.

As for **Theorem 2**, in our experiment, we were not given a known lower-dimensional basis that will definitely represent all spectra in each scene. This means that the interchangeability of $\mathbf{Q}$ and $\overline{\mathbf{Q}}$ in Equation (4.19) may not hold, i.e., $\overline{\mathbf{Q}}$ can create colour error, and subsequently, error in calculating the ground-truth fundamental metamer $\underline{e}_{fm}^{\overline{\mathbf{Q}}}$ from the RGB input $\underline{\rho}$. Of course, we may sensibly assume that as the assumed spectral dimension $m$ increases (i.e., when $m$ approaching $n$, the original spectral dimension), we get more accurate low-dimensional model that represents spectra. And yet, according to our results, this cannot be the only factor that affects the optimal selection of $m$. Indeed, we see that the optimal $m$ for different algorithms can be very different: ranging from $m = 4$ for hyperspectral pan-sharpening with upsampling-only images, to $m = 21$ for the AWAN SR algorithm, which suggests strong dependence between the optimal $m$ and individual algorithms.

Another observation on optimal $m$ is that, for some algorithms, the optimal $m$ is different for mean and worst-case (99-percentile) results, and the latter generally suggests a smaller optimal $m$. This is understandable as a lower-dimensional linear model has the effect of bounding the outliers from exceeding what the underlying assumed basis can explain. Conversely, some originally more accurate pixels could be overgeneralized by the basis and lose accuracy. We can observe both effects in the "Upsampling Only" result in Fig. 4.5. Here, in the subplot labeled "Matrix-R (4-dim)", we see that a 4-dimensional spectral representation makes the boundaries of buildings and the sky much more accurate, while losing accuracy in areas around the sidewalk.

## 4.5 Conclusion

The Matrix-R theorem teaches that, given the RGB observation and the spectral sensitivity functions of the sensors, we can certainly calculate the *fundamental metamer* component of the ground-truth spectrum, leaving the residual *metameric black* component to be uncertain. On the other hand, hyperspectral pan-sharpening algorithms seek to super-resolve low-spatial-resolution hyperspectral images given their high-spatial-resolution RGB counterparts, and spectral reconstruction (SR) algorithms recover hyperspectral images directly from the RGBs. Yet, most of these algorithms do not guarantee the exact reproduction of the fundamental metamers.

In this chapter, we showed how the Matrix-R method can be used to always improve the performance of pan-sharpening and spectral recovery: we simply make sure that it has the correct fundamental metamer. And, we provide mathematical proof that this improvement will always happen. Furthermore, we developed the Matrix-R method where spectra are represented by a low-dimensional linear model.

Experiments on several historic and state-of-the-art PS and SR algorithms show that the proposed post-processing Matrix-R method always improved

these algorithms. In addition, the low-dimensional linear basis variant of the theorem was shown to yield the best recovery results. Finally, our exploration of multispectral pan-sharpening reaffirms the efficacy of the Matrix-R method and its lower-dimensional variant.

Building upon this theoretical foundation, the next chapter addresses the practical limitation of requiring explicitly known camera sensitivities by introducing a novel, data-driven least-squares correction framework.

# Chapter 5

# Post-processing Pan-sharpening

In the previous chapter, we introduced the physics-based Matrix-R post-processing framework. We demonstrated that for any spectral recovery task—whether spectral reconstruction or pan-sharpening—the recovered spectra must satisfy a fundamental consistency constraint: when mathematically integrated with the camera's spectral sensitivities, the result must match the actual RGB input. We proved that enforcing this constraint via the substitution of the fundamental metamer consistently reduces the reconstruction error.

However, the practical application of the method presented in Chapter 4 relies on a critical assumption: that the camera spectral sensitivity functions are precisely known. In many real-world scenarios, these sensitivities are unavailable or difficult to measure without specialized laboratory equipment. Moreover, even when measured, the sensitivity data may contain calibration errors or noise, which can introduce inaccuracies into the physics-based correction. Furthermore, while Matrix-R guarantees improvement and ensures physical consistency, it does not necessarily find the optimal mapping from the estimated spectrum to the ground truth.

In this chapter, we propose a data-driven alternative to the physics-based Matrix-R method. Instead of relying on explicit knowledge of the camera sensitivities, we introduce a Least-Squares (LS) regression-based post-processing technique. In our method, we learn a correction function from a training

dataset that maps the output of any existing pan-sharpening algorithm closer to the ground-truth hyperspectral image. We then deploy this correction function to unseen data.

Advantageously, this approach overcomes the primary limitation of the previous chapter: we do not need to know the spectral sensitivities of the camera. Furthermore, we demonstrate that this simple regression-based post-processing delivers state-of-the-art pan-sharpening performance, effectively rehabilitating simple algorithms to perform competitively with complex deep-learning methods. We also show that this method effectively extends the Matrix-R theory; specifically, re-applying the physics-based Matrix-R post-processing to our LS-corrected results yields no further improvement, indicating that our data-driven method implicitly learns and respects the physical constraints.

## 5.1 Background

We adopt the mathematical notation of the previous chapter. Recall the discrete linear image formation model:

$$\underline{\rho} = \mathbf{Q}^T \underline{e} \tag{5.1}$$

where $\underline{\rho}$ is the 3-channel RGB vector, $\underline{e}$ is the $n$-channel spectral vector (typically $n = 31$ for the visible range 400nm-700nm sampled at 10nm intervals), and $\mathbf{Q}$ is the $n \times 3$ matrix of camera spectral sensitivities.

An advantage of the matrix formulation is that we can represent the RGB response to many pixels in a single equation. Let the number of pixels in an image be $N = x \times y$, where $x$ and $y$ are the number of rows and columns. We represent the colour image spatially as $\underline{\rho}(a, b)$, where $a \in [1, x]$ and $b \in [1, y]$.

To process this algebraically, we flatten these images into matrices. Let $\mathbf{E}$ be the $n \times N$ matrix containing the scene radiance spectra for all pixels, and

$\mathbf{P}$ be the $3 \times N$ matrix of corresponding RGB responses. The image formation equation for the whole image becomes:

$$\mathbf{P} = \mathbf{Q}^T \mathbf{E} \tag{5.2}$$

In RGB pan-sharpening, we essentially attempt to recover $\mathbf{E}$ given $\mathbf{P}$ and a low-resolution hyperspectral image. Let $\underline{l}(c, d)$ denote the low-resolution spectral image with dimensions $x' \times y'$, where $c \in [1, x']$ and $d \in [1, y']$ and $x' \ll x$ and $y' \ll y$. This image is upsampled to the full resolution $x \times y$ (usually via bilinear interpolation) to produce the upsampled image $\underline{L}(a, b)$. We define the corresponding $n \times N$ matrix $\mathbf{L}$ where the columns are the flattened vectors of the upsampled image.

In the spatial domain, given the upsampled low-resolution image $\underline{L}(a, b)$ and the full-resolution RGB image $\underline{\rho}(a, b)$, a pan-sharpening algorithm $\Psi$ can be written as:

$$\Psi(\mathbf{Q}, \underline{L}(a, b), \underline{\rho}(a, b)) = \widehat{\underline{e}}(a, b) \approx \underline{e}(a, b) \tag{5.3}$$

where $\widehat{\underline{e}}(a, b)$ is the estimated high-resolution spectrum.

Since the post-processing methods we develop here do not explicitly exploit spatial correlations (they operate pixel-wise or on the flattened data), we will express the pan-sharpening operation using the matrix notation:

$$\Psi(\mathbf{Q}, \mathbf{L}, \mathbf{P}) = \widehat{\mathbf{E}} \approx \mathbf{E} \tag{5.4}$$

where $\widehat{\mathbf{E}}$ is the $n \times N$ matrix of estimated spectra.

Kawakami et al. (Kawakami et al., 2011) was the first to try and solve the RGB pan-sharpening problem by using sparsity. They recover the high-resolution hyperspectral image in two steps. First, using the low-res hyperspectral image they solve for a spectral basis (which will be used to describe the spectra in the high-res image). Second, this basis is used in concert with the RGB image to recover the high-res spectral image. Sparseness as a con-

cept is used both in the derivation of the spectral basis and when it is used to *effectively* convert RGB pixel values into spectra.

In terms of our matrix notation, with respect to Kawakami et al's method, the ground truth low-dimensional image is factored as:

$$\mathbf{L} = \mathbf{B}\mathbf{A}_\mathbf{L} \tag{5.5}$$

where $\mathbf{B}$ is $31 \times k$ spectral basis (called the *endmember* basis by Kawakami) and $\mathbf{A}_\mathbf{L}$ is a $k \times n$ coefficient matrix. Applying the constraint of sparseness - the entries in $\mathbf{A}$ should be as few as possible - allows a unique factorization. Now, it is supposed that the high-res RGB image should be a sparse combinations of the RGBs formed for each basis vector. Denoting the $k$ RGBs for the basis as $\mathbf{B}_\mathbf{Q} = \mathbf{Q}^T\mathbf{B}$ then, per pixel, Kawakami solves

$$\min_{\underline{h}} ||\underline{h}||_1 \; s.t \; ||\mathbf{B}_\mathbf{Q}\underline{h} - \underline{\rho}||_F < \epsilon \tag{5.6}$$

where $\underline{h}$ is $k \times 1$ pixel level coefficients vector. Here, minimizing the $L_1$ norm of the coefficient vector delivers sparsity and we constrain the minimization so that we approximate the actual RGB within a threshold. This $\epsilon$ threshold accounts for noise in the image and modelling process.

Later studies such as Coupled Nonnegative Matrix Factorization (CNMF) (Yokoya et al., 2011) demonstrated that solving for the spectral basis and mixing coefficients in a coupled fashion yields superior performance. Lanaras et al. (Lanaras et al., 2015) incorporated additional constraints - including that reflectances are larger than 0 and less than 1 - that further improved the estimation process.

Neural networks solutions for pan-sharpening have been developed, including solving pan-sharpening with CNNs (Yang et al., 2018; Dong et al., 2022) or using Transformers (Hu et al., 2022b; Zhang et al., 2023). The latest models employ either dual-optimization or dual-branch network structures to learn spectral and spatial features simultaneously. Transformer models offer

significant advantages, such as the ability to explore global features across the entire image, in contrast to CNNs, which primarily focus on local features (Hu et al., 2022b). The current state-of-the-art Transformer method, namely the Crossed Dual-Branch U-Net (CD-UNet) by Zhang et al., adopts a U-Net network architecture that incorporates spectral and spatial self-attention mechanisms (Zhang et al., 2023).

### 5.1.1 Matrix-R from a Least-Squares Perspective

The Matrix-R approach is predicated on the idea that since we know the camera RGB $\underline{\rho}$, we effectively must also know how the estimated spectrum $\widehat{\underline{e}}$ *should* project into the subspace spanned by the camera sensors.

From a vector space point of view, the fundamental metamer $\underline{e}_{fm}$ is the component of the spectrum lying in the column space of $\mathbf{Q}$. We can solve for the linear combination $\underline{r}$ such that $\underline{e}_{fm} = \mathbf{Q}\underline{r}$. Since $\mathbf{Q}^T\underline{e}_{fm} = \underline{\rho}$, we have $\mathbf{Q}^T\mathbf{Q}\underline{r} = \underline{\rho}$, leading to:

$$\underline{e}_{fm} = \mathbf{Q}[\mathbf{Q}^T\mathbf{Q}]^{-1}\underline{\rho} \tag{5.7}$$

This structure is identical to the solution of a linear regression problem. Recall that given a target row-vector $\underline{v}^T$, the best linear combination of the rows of a matrix $\mathbf{A}^T$ is found by:

$$\underline{c}^T = \underline{v}^T\mathbf{A}[\mathbf{A}^T\mathbf{A}]^{-1} = \underline{v}^T\mathbf{A}^+ \tag{5.8}$$

where $\mathbf{A}^+$ is the Moore-Penrose inverse. By analogy, we can define the fundamental metamer directly using the pseudo-inverse notation for $\mathbf{Q}$:

$$\underline{e}_{fm} = [\mathbf{Q}^+]\underline{\rho} \tag{5.9}$$

This formulation highlights that the fundamental metamer is simply the RGB input mapped back into spectral space via the pseudo-inverse of the sensitivities.

Similarly, the projection of an estimated radiance spectrum $\widehat{\underline{e}}$ onto the subspace spanned by the columns of $\mathbf{Q}$ is written as:

$$\widehat{\underline{e}}_{fm} = \mathbf{Q}[\mathbf{Q}^T\mathbf{Q}]^{-1}\mathbf{Q}^T\widehat{\underline{e}} = \mathcal{P}(\mathbf{Q})\widehat{\underline{e}} \tag{5.10}$$

Using our definition of $\mathbf{Q}^+$, the projector matrix $\mathcal{P}(\mathbf{Q})$ (size $n \times n$) can be written compactly as:

$$\mathcal{P}(\mathbf{Q}) = \mathbf{Q}^+\mathbf{Q}^T \tag{5.11}$$

Crucially, the projection onto the subspace *orthogonal* to $\mathbf{Q}$ (the null space) defines the metameric black. This is written as:

$$\widehat{\underline{e}}_{mb} = [\mathcal{I} - \mathbf{Q}^+\mathbf{Q}^T]\widehat{\underline{e}} = \mathcal{P}(\mathbf{Q}^\perp)\widehat{\underline{e}} \tag{5.12}$$

where $\mathcal{I}$ is the identity matrix.

The Matrix-R substitution method refines an estimate by replacing its fundamental metamer component with the actual one derived from the RGB:

$$\widehat{\widehat{\underline{e}}} = \widehat{\underline{e}} - \widehat{\underline{e}}_{fm} + \underline{e}_{fm} \tag{5.13}$$

## 5.2 Proposed Method

We begin this section by examining the structure of the Matrix-R method in more detail, as our proposed data-driven method builds upon this theoretical foundation.

Using the introduced notation, we can rewrite Equation (5.13) as:

$$\widehat{\widehat{\underline{e}}} = [\mathcal{I} - \mathbf{Q}^+\mathbf{Q}^T]\widehat{\underline{e}} + \mathbf{Q}^+\underline{\rho} \tag{5.14}$$

This equation reveals that the corrected spectrum is a linear combination of the estimated spectrum and the input RGB vector. To formalize this as a single linear operation, let us define a partitioned transformation matrix $\mathcal{M}(\mathbf{Q})$ of

size $n \times (n+3)$:

$$\mathcal{M}(\mathbf{Q}) = \left[ \begin{array}{c|c} [\mathcal{I} - \mathbf{Q}^{+}\mathbf{Q}^{T}] & \mathbf{Q}^{+} \end{array} \right] \tag{5.15}$$

and a stacking function $s(\cdot;\cdot)$ that concatenates vectors (or matrices) vertically:

$$s(\underline{\widehat{e}};\underline{\rho}) = \left[ \begin{array}{c} \underline{\widehat{e}} \\ \underline{\rho} \end{array} \right] \tag{5.16}$$

It follows that the physics-based Matrix-R post-processing can be expressed as a single matrix multiplication:

$$\underline{\widehat{\widehat{e}}} = \mathcal{M}(\mathbf{Q})s(\underline{\widehat{e}};\underline{\rho}) \tag{5.17}$$

Rearranging the equations in this way simplifies the notation, but more importantly, it highlights a structural opportunity. The partitioned matrix $\mathcal{M}(\mathbf{Q})$ is composed of a projector (rank $n-3$) and the pseudo-inverse (rank 3). However, it is strictly derived from physical principles and requires precise knowledge of $\mathbf{Q}$.

The key proposal we make in this chapter is to replace this fixed, physics-based matrix with an arbitrary regression matrix $\mathcal{M}$—dropping the dependence on $\mathbf{Q}$—that best recovers the hyperspectral image from the data itself.

In effect, we place Equation (5.17) into a regression framework. Let $\widehat{\mathbf{E}}$ denote the $n \times N$ matrix of estimated spectra and $\mathbf{P}$ denote the $3 \times N$ matrix of RGB values. The ground-truth spectral matrix is denoted by $\mathbf{E}$. We seek the optimal matrix $\mathcal{M}$ in a least-squares sense by minimizing:

$$\min_{\mathcal{M}} ||\mathbf{E} - \mathcal{M}s(\widehat{\mathbf{E}};\mathbf{P})||_F^2 \tag{5.18}$$

where $||\cdot||_F$ denotes the Frobenius norm. The closed-form solution for $\mathcal{M}$ is given by:

$$\mathcal{M} = \mathbf{E}[s(\widehat{\mathbf{E}};\mathbf{P})]^{+} \tag{5.19}$$

where + denotes the standard Moore-Penrose pseudo-inverse of the stacked input data. Once learned, the refined estimates $\widehat{\widehat{\mathbf{E}}}$ are obtained via:

$$\widehat{\widehat{\mathbf{E}}} = \mathcal{M}s(\widehat{\mathbf{E}}; \mathbf{P}) \tag{5.20}$$

The reader might wonder if, in allowing $\mathcal{M}$ to have an arbitrary structure, we lose the guarantee of physical consistency provided by the Matrix-R theorem. The following theorem asserts that this is not the case.

**Theorem 5.1.** *The estimates* $\widehat{\widehat{\mathbf{E}}} = \mathcal{M}s(\widehat{\mathbf{E}}; \mathbf{P})$*—where* $\mathcal{M}$ *is least-squares optimal—cannot be further improved by deploying the Matrix-R substitution method (Equation 5.14).*

*Proof.* Without loss of generality, we consider an augmented regression problem:

$$s(\mathbf{E}; \mathbf{P}) \approx \mathcal{N}s(\widehat{\mathbf{E}}; \mathbf{P}) \tag{5.21}$$

Let $\mathbf{I}^{row}$ denote a matrix which is all zero except for the last 3 columns which contain the $3 \times 3$ identity matrix. If $\mathcal{M}$ denotes the least-squares optimal regression matrix mapping the input stack to $\mathbf{E}$, clearly $\mathcal{N} = s(\mathcal{M}; \mathbf{I}^{row})$ is also a least-squares optimal solution for the augmented target (since the RGB part of the input stack maps perfectly to the RGB part of the target stack via identity).

In this augmented form, the Matrix-R substitution rule can be written as a linear operation on the output of $\mathcal{N}$:

$$s(\mathbf{E}; \mathbf{P}) \approx s(\mathcal{M}(\mathbf{Q}); \mathbf{I}^{row})\mathcal{N}s(\widehat{\mathbf{E}}; \mathbf{P}) \tag{5.22}$$

Now let us assume that applying the Matrix-R correction after applying the regression matrix $\mathcal{N}$ results in an improved - i.e., lower -regression error. It follows that

$$||\mathcal{N}s(\widehat{\mathbf{E}};\mathbf{P}) - s(\mathbf{E};\mathbf{P})||_F > ||s(\mathcal{M}(\mathbf{Q});\mathbf{I}^{row})\mathcal{N}s(\widehat{\mathbf{E}};\mathbf{P}) - s(\mathbf{E};\mathbf{P})||_F \quad (5.23)$$

But this cannot be the case because $\mathcal{N}$ is already, by assumption, the least-squares optimal solution. This contradicts our assumption, and our theorem is proven.

□

**Regularization:** In practice, simply calculating the pseudo-inverse can lead to overfitting or instability. To address this, we employ Tikhonov regularization. Let $\mathbf{S} = s(\widehat{\mathbf{E}}_{train};\mathbf{P}_{train})$ be the $(n+3) \times N_{train}$ matrix of stacked training inputs. The regularized solution for $\mathcal{M}$ is:

$$\mathcal{M} = \mathbf{E}_{train}\mathbf{S}^T[\mathbf{S}\mathbf{S}^T + \gamma\mathcal{I}]^{-1} \quad (5.24)$$

where $\mathcal{I}$ is the $(n+3) \times (n+3)$ identity matrix and $\gamma$ is a user-defined regularization parameter. In our experiments, we set $\gamma$ to be a small fraction (e.g., 0.01%) of the mean variance of the data, calculated as $trace(\mathbf{S}\mathbf{S}^T)/N_{train}$.

**Direct Regression (Identity Case):** Interestingly, we also consider the case where the initial pan-sharpening algorithm $\Psi$ is simply the identity transform of the upsampled low-resolution image, i.e., $\Psi_{identity} = \underline{L}$. In this scenario, we regress $\mathcal{M}$ based solely on the upsampled low-resolution image and the full-resolution RGB:

$$\mathcal{M} = \mathbf{E}_{train}[s(\mathbf{L}_{train};\mathbf{P}_{train})]^+ \quad (5.25)$$

As we will see in the experimental section, treating pan-sharpening as a direct regression from these basic inputs works surprisingly well, often outperforming traditional model-based methods while bypassing the need for complex, iterative optimization or known camera sensitivities.


**Figure 5.1:** *Example RGB images from the ICVL (Arad and Ben-Shahar, 2016), CAVE (Yasuma et al., 2010), and Harvard (Chakrabarti and Zickler, 2011) Datasets. These images were generated numerically with gamma correction applied.*

## 5.3 Experiments

### 5.3.1 Datasets

In our study, we utilized three distinct hyperspectral image datasets as demonstrated in Figure 5.1: the Ben-Gurion University Dataset (ICVL) (Arad and Ben-Shahar, 2016), the Columbia University Dataset (CAVE) (Yasuma et al., 2010), and the Harvard University Dataset (Harvard) (Chakrabarti and Zickler, 2011).

The Harvard Dataset (Chakrabarti and Zickler, 2011) encompasses 50 hyperspectral images capturing a mix of indoor and outdoor scenes, each measuring 1040 x 1392 pixels. These images span a spectral range of 420nm to 720nm at 10nm intervals. Given our RGB camera sensitivities cover the 400-700nm range, we aligned our focus to the overlapping 29 spectral channels from 420-700nm. Additionally, this dataset provides masks to address artifacts from moving objects or dust because of high acquisition times, which we applied pre-experimentation to ensure data integrity. All images were in pixel-wise correspondence. For the supervised methods - CD-UNet and standalone LS - we used the first 30 images for training and the other 20 images for testing as in (Zhang et al., 2023). Although the rest of the methods are unsupervised and don't need separate training and testing sets, we run these algorithms only on the test images to calculate the RMSE for a fair comparison.

The CAVE Dataset (Yasuma et al., 2010) features 32 indoor images categorized into five themes: stuff, skin and hair, food and drinks, real and fake, and paints, each with a resolution of 512 × 512 pixels covering the same spectral range between 400nm - 700nm with a 10nm interval. Each spectral band is stored as a 16-bit grayscale image. Similar to the Harvard dataset, we used roughly the first 60 percent of the data for training and the last 40 percent for testing - 20 images to train and 12 images to test - for CD-UNet and LS algorithms. The other algorithms' error metrics are also calculated on these 12 test images for consistency.

The ICVL Dataset (Arad and Ben-Shahar, 2016) contains 201 outdoor hyperspectral images encoded in 12 bits with resolutions typically at 1392 × 1300 pixels spanning 31 spectral bands. These bands are evenly distributed over 10nm intervals from 400 to 700 nanometers, capturing a broad spectrum of visible light. The ICVL dataset is the largest used in our experiments. Since training the CD-UNet is computationally expensive, we used only the center portion of each image in the dataset. Specifically, we cropped 512 × 512 pixel images starting from an offset point of (500, 500). The dataset was then split into training and testing sets, with the first 150 images used for training and the remainder for testing. All error calculations were performed on these test images.

To assess the performance of various pan-sharpening algorithms, we constructed pairs of high-resolution RGB and low-resolution hyperspectral images (serving as inputs) derived from the original hyperspectral data (serving as the ground truth). The input RGB images were produced using a numerical integration method with the spectral sensitivity functions of Canon 1D Mark III camera, excluding gamma correction. Meanwhile, the low-resolution hyperspectral inputs were obtained by downscaling the original images by a factor of 8 through bilinear interpolation, effectively mimicking a hyperspectral thumbnail at 1/64 the size of the ground truth.

### 5.3.2 Algorithms

The following algorithms are investigated in this chapter:

(i) **MR**: The original Matrix-R solution (Imai and Berns, 1998).

(ii) **LS**: The proposed least-squares solution on augmented data combining hyperspectral and RGB pixels.

(iii) **CNMF**: The Coupled Non-negative Matrix Factorization (Yokoya et al., 2011).

(iv) **CNMF - MR**: CNMF enhanced with Matrix-R post-correction (Lin et al., 2023).

(v) **CNMF - LS**: CNMF enhanced with proposed LS post-correction.

(vi) **Lanaras**: The Coupled Spectral Unmixing by Lanaras et al. (Lanaras et al., 2015).

(vii) **Lanaras - MR**: Lanaras enhanced with Matrix-R post-correction (Lin et al., 2023).

(viii) **Lanaras - LS**: Lanaras enhanced with proposed LS post-correction.

(ix) **CD-UNet**: The Crossed Dual-Branch U-Net (Zhang et al., 2023).

(x) **CD-UNet - MR**: The CD-UNet enhanced with Matrix-R post-correction.

(xi) **CD-UNet - LS**: The CD-UNet method enhanced with proposed LS post-correction.

For Matrix-R, CNMF, and Lanaras algorithms, the RGB camera sensitivities, high-resolution RGB images, and low-resolution hyperspectral images are utilized as inputs. The calculations are conducted on an image-by-image basis on the test datasets.

For standalone LS and CD-UNet methods, we used separate training and testing sets as they are supervised learning algorithms. For each dataset, we

take the first 60 percent of the images as training and the rest as testing. In these data-driven approaches, we relied solely on high-resolution RGB images and low-resolution hyperspectral images as inputs, as camera sensitivities are not required. Here, we also need to mention that our experiments slightly differ from the one in (Zhang et al., 2023) where CD-UNet is introduced as they use different camera sensitivity functions to simulate RGB images and also a different scaling strategy as a Gaussian filter with a scale factor of 4 - as opposed to our experiments where we conducted bilinear upsampling with the scale factor of 8. Additionally, they did not mention whether they discarded the pixels using the masks provided for each image in the Harvard dataset.

For (iv), (vii), and (x), the Matrix-R was applied as a post-processing step on hyperspectral images already pan-sharpened by the CNMF, Lanaras, and CD-UNet methods, respectively. Similarly, for (v), (viii), and (xi), our proposed LS solution was applied as a post-correction step on previously sharpened images by the original methods. In the case of CNMF, it was observed that the original algorithm occasionally produces significant outliers, which adversely impact the subsequent least-squares post-correction process. To mitigate this issue, the pixel values sharpened by CNMF that fall outside the [0,1] range are clipped prior to the application of post-processing.

For the methods with the LS post-processing, we only had the pan-sharpened test images to work on such as 12 test images for the CAVE dataset. Due to the limited data to calculate the LS matrix and errors on those images, we implemented a 5-fold cross-validation technique to assess the method's effectiveness. For each fold, 80% of the test dataset was used to calculate the projection matrix with regularization, and 20% of the test data was utilized to compute the error. This process was repeated five times to ensure testing covered all samples in the testing set, leading to the calculation of the mean error metrics from the 5-fold cross-validation.

We also recognize that different algorithms may use different normalization methods on images, resulting in varying average brightness levels in the

output images. To ensure a fair comparison, we scaled each output image so that it has the same average pixel value as the corresponding ground truth image before the error calculations.

## 5.4 Results

Results from the experiments are shown in Table 5.1 for the ICVL, CAVE, and Harvard Datasets. This research utilizes RMSE as the metric for evaluation. Figure 5.2 provides a visual illustration of how the spectral error is calculated for a single pixel. We calculated RMSE for every pixel as:

$$RMSE = \sqrt{\frac{1}{n}||\underline{\widehat{E}} - \underline{E}||^2} \tag{5.26}$$

where $\underline{\widehat{E}}$ and $\underline{E}$ the sharpened and ground-truth hyperspectral pixel vectors, respectively. Then we noted the average and the 99th percentile (99pt) RMSE values for each image within the datasets. The displayed statistics in the tables are the averages of these per-image metrics across all test images for each respective dataset. The original error refers to the average RMSE between the input low-resolution hyperspectral images and the ground-truth high-resolution hyperspectral images. We highlighted results in Table 5.1 with four groups. First, standalone Matrix-R-based methods - the original Matrix-R and the proposed standalone LS. Then, we have CNMF, Lanaras, and CD-UNet groups, where we examine the original algorithms and their Matrix-R and LS post-processed versions, respectively.

In Table 5.1, we first present the mean RMSE ($\times 10^4$) metrics for the Harvard dataset. The results show the average mean and 99th percentile RMSE values across 20 test images. The Matrix-R solution alone significantly reduces the original RMSE by approximately 45%. On the other hand, optimization methods deliver superior results compared to Matrix-R alone; in particular, Lanaras achieves slightly better scores than CNMF. However, the

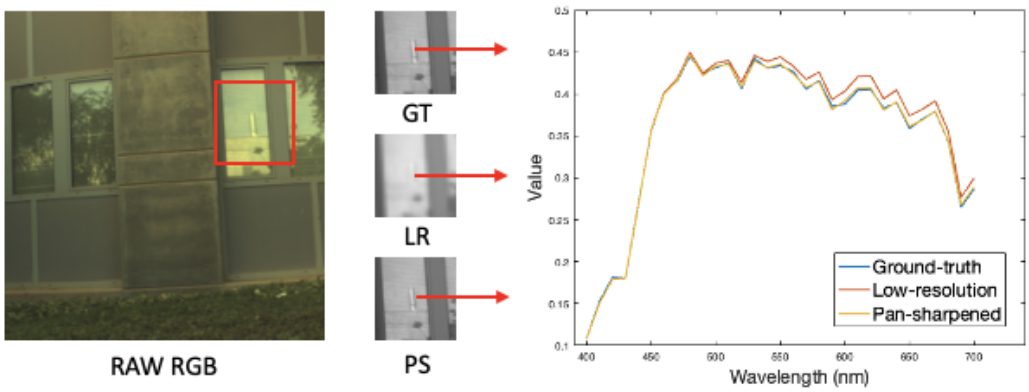


**Figure 5.2:** *Illustration of the spectral accuracy improvement and RMSE calculation for a selected pixel from the ICVL dataset (Arad and Ben-Shahar, 2016). (Left) The source RAW RGB image. (Middle) Zoomed-in patches showing a white object on the window frame for the Ground-Truth (GT), the bilinearly upsampled Low-Resolution (LR) input, and the Pan-Sharpened (PS) result. (Right) The spectral plots for the single pixel indicated by the arrows. The LR spectrum deviates noticeably from the GT (RMSE* $= 9.4 \times 10^{-3}$*), while the LS-sharpened spectrum is a much closer match (RMSE* $= 2.0 \times 10^{-3}$*), demonstrating a significant reduction in spectral error.*

current state-of-the-art neural network method, CD-UNet, outperforms the traditional algorithms by a substantial margin. In all groups, it is evident that post-processing steps consistently improve the performance of the methods. Here, our proposed LS post-processing step demonstrates better performance than the previously introduced Matrix-R post-processing across all methods. Overall, CD-UNet with the LS post-correction yields the best score for the Harvard dataset. Nevertheless, it is important to note that the standalone LS method surprisingly has the second-best score, just after the CD-UNet-based methods. Given that it is the simplest and fastest algorithm in the table, it shows promise for real-time use, in contrast to cumbersome deep learning solutions.

The RMSE ($\times 10^3$) results for the CAVE dataset are calculated using 12 test images. The ranking of the main algorithms is similar to the ICVL results: CD-UNet outperforms the traditional optimization algorithms, but all methods still show room for improvement with post-correction, as they do not structurally guarantee the correct fundamental metamer. However, this time, LS post-processing does not achieve better scores than Matrix-R post-

Table 5.1: Average RMSE Metrics for the Harvard, CAVE, and ICVL Datasets

| | **Harvard** | | **CAVE** | | **ICVL** | |
|---|---|---|---|---|---|---|
| | **Mean** | **99pt** | **Mean** | **99pt** | **Mean** | **99pt** |
| Original Error | 7.21 | 53.65 | 1.87 | 15.55 | 22.58 | 142.70 |
| **Standalone Matrix-R Based** | | | | | | |
| MR | 4.02 | 28.71 | 0.99 | 7.36 | 10.50 | 66.87 |
| LS | 1.60 | 6.66 | 0.46 | 2.66 | 2.95 | 12.61 |
| **CNMF** | | | | | | |
| CNMF | 2.00 | 8.59 | 0.90 | 4.73 | 9.15 | 41.94 |
| CNMF - MR | 1.80 | 7.43 | 0.83 | 4.38 | 5.11 | 23.39 |
| CNMF - LS | 1.62 | 6.21 | 0.89 | 4.29 | 3.07 | 12.22 |
| **Lanaras** | | | | | | |
| Lanaras | 1.91 | 8.50 | 0.78 | 5.17 | 7.07 | 36.53 |
| Lanaras - MR | 1.79 | 7.64 | 0.71 | 4.52 | 5.23 | 25.73 |
| Lanaras - LS | 1.71 | 6.83 | 0.75 | 4.27 | 3.35 | 13.76 |
| **CD-UNet** | | | | | | |
| CD-UNet | 1.56 | 5.54 | 0.41 | 2.58 | 2.13 | 6.56 |
| CD-UNet - MR | 1.47 | 5.38 | 0.38 | 2.51 | 1.82 | 6.03 |
| CD-UNet - LS | 1.46 | 5.27 | 0.40 | 2.60 | 1.71 | 5.73 |

processing. This is understandable, as LS is a data-driven technique, and here we have only 12 test images, each with a size of 512x512 pixels. This means we use 10 images for training and 2 for testing in each iteration of 5-fold cross-validation when performing post-correction with LS. Clearly, this is insufficient to generalize the least-squares matrix, especially for CAVE images as they are very different than each other. (For curious readers, we note that if we use the entire set for both training and testing, LS post-processing always delivers the best score, as expected from the equations. However, for a fair comparison, we only report scores where training and testing are done on different sets.) On the other hand, the standalone LS method, where we use the first 20 images for training and the remaining 12 images for testing, similar to CD-UNet, again achieves the second-best scores, which are slightly worse than those of the

original CD-UNet model.

For the ICVL results, we used the average RMSE ($\times 10^3$) metrics for 51 cropped test images. Once again, we confirm that CD-UNet delivers superior performance compared to the traditional algorithms, including Matrix-R, CNMF, and Lanaras. Similar to the Harvard results, the proposed LS post-processing method produces less error than Matrix-R post-processing, while the standalone LS method remains competitive with CD-UNet.

In Figure 5.3, the visualization of the results of different methods is shown for an example image from the Harvard dataset. The "Original Error" represents the RMSE values between the ground truth and the input hyperspectral images, with the error values averaged across all channels. As can be seen, high errors typically occur at edges, shadows, or saturated pixels. In the second row, the performance of the original Matrix-R and the proposed LS methods are displayed. Although both methods are based on Matrix-R, their performance differences are evident. In the subsequent rows, error maps for other algorithms are shown, including their MR and LS post-corrected versions. MR provides relatively small but consistent improvements to the original algorithms when applied as post-processing. In contrast, LS shows visibly better overall improvement (as the error image becomes darker) but may introduce the risk of error increases in certain local areas (e.g., reflections on some windows). This is expected, as data-driven techniques are sensitive to the training and testing data. Additionally, the least-squares method aims to minimize the average error, regardless of the maximum error, although the applied regularization helps mitigate this to some extent. It should be noted that this single example does not reflect the performance or ranking of the models across the entire dataset or other datasets, as these may vary from image to image. The figure is provided merely to illustrate the pan-sharpening error and the effectiveness of post-processing methods.

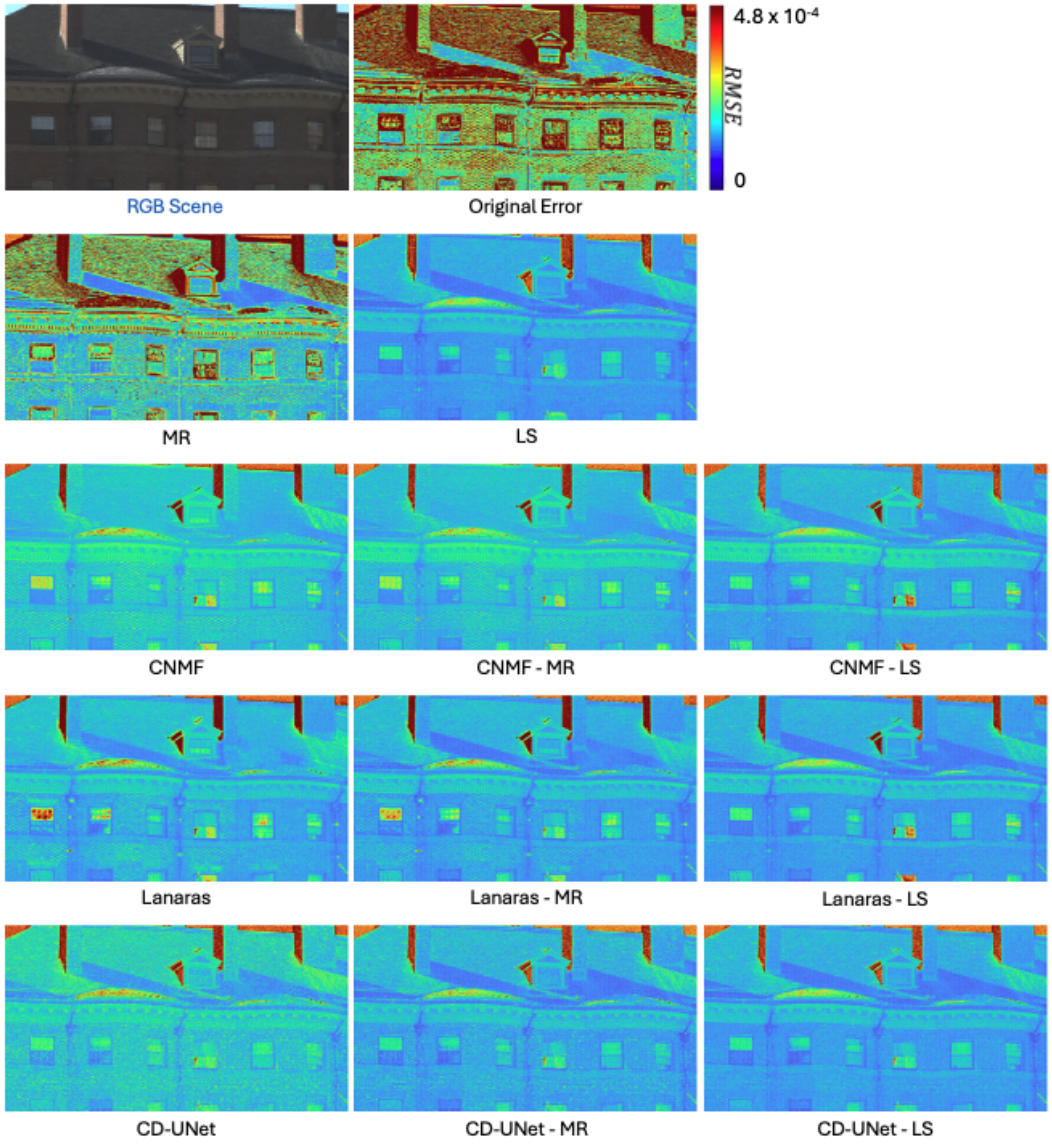


**Figure 5.3:** *The RMSE error heat maps of the algorithms for the tested RGB pan-sharpening on an example image from the Harvard Dataset.*

## 5.5 Discussion

While deep-net models have demonstrated state-of-the-art performance in spectral super-resolution tasks, a natural question arises: If a simple post-processing step, such as least-squares regression, consistently improves results, why is this correction not already learned by the network itself during training?

We believe there are several factors at play. First, no neural network — including state-of-the-art Transformers — guarantees the globally optimal solution. Their performance is inherently tied to the characteristics of the training data, optimisation dynamics, loss function design, and inductive biases encoded in the architecture. These models are excellent at approximating complex mappings, but they remain fundamentally limited by their learning capacity, the non-convexity of the optimisation landscape, and generalisation gaps.

In contrast, the Matrix-R post-processing step is derived from a mathematically grounded formulation that injects known truths into the system. Specifically, when we know that certain parts of the output (the fundamental metamer) are accurate, we can constrain the final prediction accordingly, as shown in the previous chapter. So, it is not surprising that Matrix-R consistently improves performance by enforcing this partial ground-truth in a principled.

Furthermore, in this chapter, we go beyond the fixed Matrix-R formulation and introduce an optimisation-based refinement that adapts post-correction in a more flexible way. Conceptually, this can be seen as appending an additional lightweight module to the network — one that operates deterministically and ensures consistency with known constraints. Although one could imagine training the network to learn this mapping end-to-end, this is rarely effective in practice. The reason lies in the complexity and depth of modern architectures: information — including the RGB guidance signal — tends to diffuse or attenuate across layers. By the time the signal reaches the output, its influence

is significantly diminished.

In contrast, the post-processing step re-introduces the original RGB data at full strength, aligned with the output hyperspectral estimate. This acts as a final, targeted correction — directly linking the high-confidence RGB information to the uncertain spectral reconstruction. In essence, the post-processing step behaves like a shortcut connection that bypasses the deep network and applies a clean, effective transformation rooted in known physics. Thus, we argue that our proposed LS post-processing is not merely a redundant correction but a principled and efficient enhancement — one that ensures that available ground-truth information is fully leveraged.

## 5.6 Conclusion

In this chapter, we embarked on an exploration of RGB-based hyperspectral pan-sharpening, a crucial technique for enhancing the spatial resolution of hyperspectral images by leveraging high-resolution RGB images. Our investigation spanned a wide array of established algorithms, from the foundational Matrix-R approach to notable optimization algorithms and current state-of-the-art transformer-based neural networks.

At the heart of our contribution is the introduction of a novel, data-driven pan-sharpening method that stands out by not requiring camera sensitivity functions - a standard requirement in many traditional approaches. It also ensures the output image already has the correct fundamental metamer eliminating the Matrix-R post-processing step that has been recommended recently for any pan-sharpening algorithms. The proposed method, which applies a least-squares solution to augmented data from both hyperspectral and RGB images, simplifies the pan-sharpening process while ensuring high accuracy and efficiency. Demonstrated across the ICVL, CAVE, and Harvard Datasets, our proposed method significantly outperforms traditional algorithms, underscoring the potential of data-driven strategies in advancing the field of hyper-

spectral pan-sharpening.

Having established this data-driven spectral recovery method, the next chapter shifts our focus back to the color correction problem. Building on the regression techniques developed here, Chapter 6 investigates a dual-camera solution that integrates multispectral data with RGB images to achieve superior colour accuracy

# Chapter 6

# Colour Correction with Multispectral Capture

While the previous chapter demonstrated the power of least-squares regression for stacked RGB and multispectral pixels, this chapter applies that same mathematical foundation to the color correction problem. Specifically, we investigate a dual-camera system that integrates multispectral data with standard RGB images to achieve superior accuracy.

RGB cameras exhibit spectral sensitivity functions that differ from those of the human visual system (HVS), leading to discrepancies between how scenes are captured by cameras and perceived by humans. To address this, the RAW pixel values from RGB cameras are typically transformed into a standard colour space, such as CIE XYZ or sRGB—a process known as colour correction. Traditionally, this transformation relies solely on RGB images. However, the increasing availability of multispectral cameras, which offer more detailed spectral sampling of a scene, opens up new possibilities. Despite their spectral advantages, multispectral cameras are typically limited by low spatial resolution. To mitigate this, their corresponding high-resolution RGB images are often used to guide a fusion process known as RGB pan-sharpening.

In this chapter, we adopt the standard pan-sharpening setup but, instead of reconstructing a sharpened multispectral image, we propose a novel colour

correction method to directly estimate a high-quality, three-channel CIE XYZ image. Our approach first upsamples the low-resolution multispectral image and stacks it with the aligned high-resolution RGB image. We then learn a regression from this combined data to the target CIE XYZ image. Experimental results show that our method significantly outperforms conventional colour correction techniques that rely solely on RGB inputs, demonstrating the value of integrating multispectral information even at low resolution.

## 6.1 Introduction

Colour correction algorithms are essential tools designed to accurately convert camera-specific RAW RGB responses into standardised, device-independent colour spaces such as CIE XYZ (Smith and Guild, 1931). Despite advances, achieving precise colour correction remains challenging due to the fundamental mismatch between camera spectral sensitivity functions and the standard XYZ colour matching functions (Finlayson and Morovic, 2000; Hong et al., 2001). This mismatch leads to a fundamental ambiguity known as *metamerism*, where spectrally distinct objects result in identical RGB triplets but differ in human perception. Since the camera cannot distinguish these 'metamers', standard correction pipelines inevitably introduce errors. Consequently, simple linear transformations, although computationally efficient, often fail to achieve high-precision colour reproduction (Finlayson et al., 2015).

Various regression methods, including linear, polynomial, and root-polynomial approaches, have been developed to enhance colour correction performance (Finlayson and Morovic, 2000; Hong et al., 2001; Finlayson et al., 2015). Polynomial regression demonstrates improved accuracy over basic linear methods by capturing nonlinear relationships between camera responses and target XYZ values (Hong et al., 2001); however, it is sensitive to exposure variations as we discussed in Section 3.3. In contrast, linear and root-polynomial regressions are inherently exposure-invariant (Finlayson et al., 2015).

Multispectral imaging presents a promising avenue to overcome the limitations of traditional RGB imaging by offering additional spectral bands beyond RGB. By increasing the dimensionality of the capture, multispectral data helps to resolve the aforementioned metameric ambiguities, allowing for the differentiation of spectral signatures that are indistinguishable to a trichromatic sensor. This enriched spectral information is not only essential for different applications (Habib et al., 2020; Cosentino et al., 2015) but also has shown significant promise in various imaging enhancement tasks, including illumination estimation, spectral reconstruction, and white balance correction (Khan et al., 2017; Li et al., 2021; Lapray et al., 2014). However, explicit applications of multispectral imaging specifically for direct camera RAW-to-XYZ colour correction remain relatively underexplored.

In Chapter 4, we explored multispectral imaging within the pan-sharpening context, addressing the fusion of low-resolution multispectral images with high-resolution RGB imagery to enhance spatial resolution. In Chapter 5, we proposed a Least-Squares (LS) regression-based post-processing technique (we also showed the direct usage of it) for low-resolution hyperspectral images, demonstrating mathematically that this approach consistently improves pan-sharpening accuracy regardless of the initial algorithm used such as (Kawakami et al., 2011; Lanaras et al., 2015). The simplicity and efficacy of this method highlighted the potential of optimised regression-based strategies.

Building on these insights, this chapter introduces a novel multispectral capture (MSC)-enhanced colour correction method. Unlike the main theme in the previous chapter, here we employ the regression framework directly for colour correction rather than as a post-processing step. Specifically, we adapt the mathematical strategy from pan-sharpening (Chapter 5), replacing the multispectral regression target with CIE XYZ tristimuli as illustrated in Figure 6.1. It is important to note that obtaining dense, pixel-wise ground-truth XYZ targets for training is non-trivial in practice. In a laboratory setting, XYZ values can be computed from hyperspectral images via numerical integration

with the colour matching functions, but acquiring such hyperspectral ground truth already requires specialised equipment. In more practical scenarios, the regression matrix must be derived from a limited set of known colour samples - for example, the 24 patches of a Macbeth ColorChecker chart - whose XYZ values can be measured with a spectrophotometer or calculated from their known reflectances. This reliance on sparse calibration data is a well-understood constraint of regression-based colour correction and applies equally to both the classical RGB-only methods and the proposed fusion approach.

A key motivation for incorporating multispectral data is to mitigate the problem of metamerism. Metamerism arises because the mapping from spectra to RGB is many-to-one: physically distinct spectra can produce identical camera RGB responses, since the three broad-band sensor channels discard most of the spectral information (as formalised by the Matrix-R framework in Chapter 4 and 5, the entire metameric black component is lost). Consequently, a colour correction method that operates solely on three-channel RGB input cannot distinguish between metameric stimuli and must map them to the same XYZ output - even when the correct XYZ values differ. By augmenting the RGB data with additional multispectral channels, the input dimensionality increases and the space of indistinguishable metamers shrinks. In the ideal case where the combined sensor set spans enough of the spectral space, two stimuli that are metameric under the RGB camera alone will produce distinct responses in the multispectral channels, allowing the regression to map them to their respective, correct XYZ values. This direct approach, therefore, leverages both the spatial detail from RGB images and the richer spectral sampling from MSC captures, significantly enhancing perceptual colour accuracy.

The remainder of this chapter provides a detailed mathematical formulation of our MSC-enhanced colour correction method, experimental validation, and benchmarking against classical methods including linear, polynomial, and root-polynomial regressions, demonstrating significant improvements in colour accuracy across diverse illumination conditions.

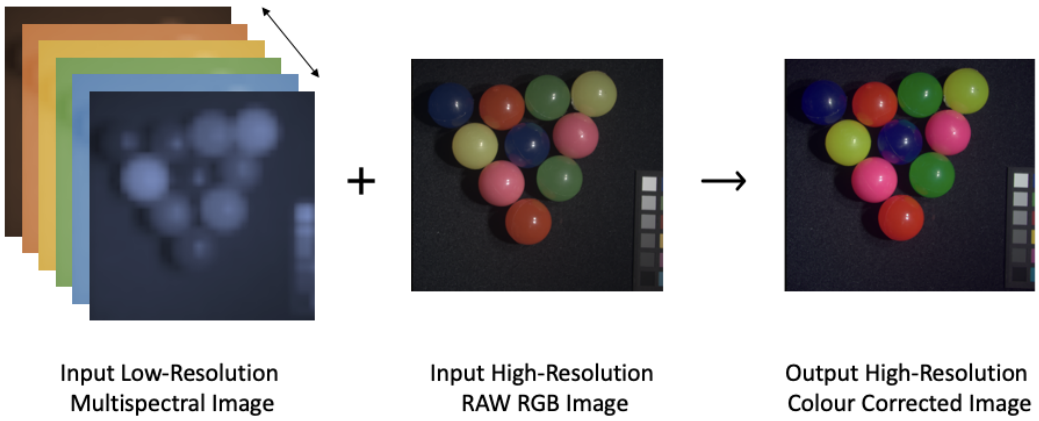


**Figure 6.1:** *The demonstration of the proposed colour correction method with multispectral image. Left: input low-resolution multispectral image. Middle: corresponding input full-res RAW RGB image. Right: the full-res colour corrected image. Images are generated from the CAVE dataset (Yasuma et al., 2010).*

## 6.2 Background

### 6.2.1 Colour Correction Revisited

We begin by restating the colour correction problem using notation established in previous chapters. Let $Q_R(\lambda)$, $Q_G(\lambda)$, and $Q_B(\lambda)$ represent the spectral sensitivity functions (SSFs) of the three channels (Red, Green, Blue) of a camera. We denote these collectively in vector form as $\mathbf{Q}(\lambda) = [Q_R(\lambda), Q_G(\lambda), Q_B(\lambda)]^T$.

The camera's response vector $\underline{\rho} = [R, G, B]^T$ for a given surface with spectral reflectance $S(\lambda)$ under an illuminant with spectral power distribution $I(\lambda)$ is calculated by integrating the product of these functions over the visible spectrum visible spectrum (typically approximated as 400–700 nm):

$$\rho_j = \int Q_j(\lambda) I(\lambda) S(\lambda) d\lambda, \quad \text{for } j \in \{R, G, B\} \tag{6.1}$$

Here, $\rho_j$ is the scalar response for the $j$-th channel, and $\lambda$ denotes the wavelength. The resulting vector $\underline{\rho}$ encapsulates the camera's recorded RGB values for that surface.

Similarly, the CIE XYZ tristimulus response vector $\underline{x} = [X, Y, Z]^T$ for the same surface and illuminant is defined using the CIE XYZ colour matching

functions (CMFs), denoted by the vector $\mathbf{X}(\lambda) = [\bar{x}(\lambda), \bar{y}(\lambda), \bar{z}(\lambda)]^T$:

$$x_k = \int \bar{x}_k(\lambda) I(\lambda) S(\lambda) d\lambda, \quad \text{for } k \in \{X, Y, Z\} \tag{6.2}$$

where $x_k$ is the scalar value for the $k$-th tristimulus component (using $\bar{x}_X = \bar{x}$, $\bar{x}_Y = \bar{y}$, $\bar{x}_Z = \bar{z}$).

In practice, we often work with measurements from $n$ different surfaces or samples. Let $\mathbf{P}$ be an $n \times 3$ matrix where each row $i$ contains the camera's RGB response vector $\underline{\rho}_i^T$ for the $i$-th sample. Similarly, let $\mathbf{X}$ be the corresponding $n \times 3$ matrix where each row $i$ contains the target XYZ tristimulus vector $\underline{x}_i^T$.

Classical linear colour correction aims to find a $3 \times 3$ transformation matrix $\mathbf{M}$ that maps the camera RGB responses to the corresponding XYZ tristimulus values. This is typically achieved by minimising the sum of squared errors between the transformed responses and the target values, often expressed using the Frobenius norm:

$$\mathbf{M}_{\text{opt}} = \arg\min_{\mathbf{M}} \|\mathbf{PM} - \mathbf{X}\|_F \tag{6.3}$$

The goal is to find the matrix $\mathbf{M}$ that best transforms the $n \times 3$ matrix of camera responses $\mathbf{P}$ into the $n \times 3$ matrix of target tristimulus values $\mathbf{X}$. The optimal solution $\mathbf{M}_{\text{opt}}$ under the least-squares criterion has a closed-form solution involving the Moore-Penrose pseudo-inverse of $\mathbf{P}$ (denoted $\mathbf{P}^+$):

$$\mathbf{M}_{\text{opt}} = (\mathbf{P}^T\mathbf{P})^{-1}\mathbf{P}^T\mathbf{X} = \mathbf{P}^+\mathbf{X} \tag{6.4}$$

Here, $\mathbf{P}^T$ denotes the transpose of $\mathbf{P}$.

Polynomial and root-polynomial regression methods generalise this linear approach. They first expand the input RGB response vector $\underline{\rho}$ into a higher-dimensional feature vector containing nonlinear terms. For instance, a common second-order root-polynomial expansion transforms $\underline{\rho} = [R, G, B]^T$ into a $k = 6$

dimensional vector $f_r^2(\underline{\rho})$:

$$f_r^2(\underline{\rho}) = [R, G, B, \sqrt{RG}, \sqrt{RB}, \sqrt{GB}]^T \tag{6.5}$$

This expansion is applied row-wise to the matrix $\mathbf{P}$, resulting in an expanded feature matrix $\mathbf{P}_{\text{exp}} = f_r^2(\mathbf{P})$ of size $n \times k$. The regression then finds a $k \times 3$ matrix $\mathbf{M}_{\text{exp}}$:

$$\mathbf{M}_{\text{exp}} = \arg\min_{\mathbf{M}} \|\mathbf{P}_{\text{exp}}\mathbf{M} - \mathbf{X}\|_F \tag{6.6}$$

The optimal solution is again found using the pseudo-inverse: $\mathbf{M}_{\text{exp}} = \mathbf{P}_{\text{exp}}^{+}\mathbf{X}$.

### 6.2.2 Multispectral Capture and Pan-Sharpening Revisited

Multispectral imaging (MSI) involves capturing image data across multiple spectral bands. Let a multispectral camera have $d$ spectral channels, with corresponding spectral sensitivities $Q_i(\lambda)$ for $i = 1, \ldots, d$. The captured image, often called a multispectral cube, contains rich spectral detail for each pixel, which is highly beneficial for applications requiring precise colour information or material analysis. However, due to physical constraints (e.g., sensor design, filter wheels, data bandwidth), multispectral cameras often exhibit lower spatial resolution compared to standard RGB cameras designed for consumer photography.

Pan-sharpening techniques aim to bridge this gap. They fuse a low-resolution multispectral image (denoted $\mathbf{L}$, with spatial dimensions $m \times n$ and $d$ spectral channels) with a co-registered high-resolution guidance image (denoted $\mathbf{H}$, often panchromatic or RGB, with spatial dimensions $k_1 m \times k_2 n$, where $k_1, k_2 > 1$ are the upscaling factors) to produce an estimated high-resolution multispectral image $\widehat{\mathbf{C}}$. Formally, given $\mathbf{L}$ and $\mathbf{H}$, a pan-sharpening algorithm $\Psi$ computes:

$$\widehat{\mathbf{C}} = \Psi(\mathbf{H}, \mathbf{L}) \tag{6.7}$$

The goal is for the estimate $\widehat{\mathbf{C}}$ (dimensions $k_1m \times k_2n \times d$) to be as close as possible to the (usually only available in simulations) ground-truth high-resolution multispectral image $\mathbf{C}$.

In Chapter 5, a regression-based post-processing step for pan-sharpening outputs was developed. This involved first taking the initial estimate $\widehat{\mathbf{C}}$, which is calculated by different pan-sharpening algorithms (or sometimes only up-sampling the original low-resolution MSC $\mathbf{L}$ by bilinear interpolation for the direct approach). This estimate was then combined with the high-resolution guidance image $\mathbf{H}$ (let's assume $\mathbf{H}$ is the $k_1m \times k_2n \times 3$ RGB image, denoted $\mathbf{P}$ for consistency with the previous section). The combination typically involves stacking the channels or vectorizing the pixel data. Let $\mathbf{s}(\widehat{\mathbf{C}}; \mathbf{P^T})$ represent the $(d+3) \times N$ matrix resulting from vectorizing and stacking the pixel data ($N = k_1m \times k_2n$ is the total number of high-resolution pixels):

$$s(\widehat{\mathbf{C}}; \mathbf{P^T}) = \begin{bmatrix} \widehat{\mathbf{C}} \\ \mathbf{P^T} \end{bmatrix} \tag{6.8}$$

A linear regression model $\mathbf{M}$ (of size $d \times (d+3)$) was then learned to map these stacked features $\mathbf{s}(\widehat{\mathbf{C}}; \mathbf{P^T})$ to the ground-truth high-resolution MSC data $\mathbf{C}$ (of size $d \times N$):

$$\arg\min_{\mathbf{M}} \|\mathbf{C} - \mathbf{M}\mathbf{s}(\widehat{\mathbf{C}}; \mathbf{P^T})\|_F \tag{6.9}$$

The optimal regression matrix for pan-sharpening is given by:

$$\mathbf{M}_{\text{ps}} = \mathbf{C}[\mathbf{s}(\widehat{\mathbf{C}}; \mathbf{P^T})]^+ \tag{6.10}$$

### 6.2.3 Proposed Method: MSC-Enhanced Colour Correction

Our proposed method adapts the regression framework developed for pan-sharpening post-processing and applies it explicitly to the task of colour correction, aiming to map directly from sensor data (RGB and MSC) to CIE XYZ

tristimulus values.

First, the available low-resolution multispectral image $\mathbf{L}$ (size $m \times n \times d$) is upsampled using bilinear interpolation to match the spatial resolution of the high-resolution RGB image $\mathbf{P}$ (size $k_1 m \times k_2 n \times 3$). Let this upsampled MSC image be $\widehat{\mathbf{C}}$ (for the consistency with the previous equations - as it is pan-sharpened).

Next, we again form the stacked feature matrix $\mathbf{s}(\widehat{\mathbf{C}}; \mathbf{P^T})$ (size $(d+3) \times N$, where $N = k_1 m \times k_2 n$) by vectorizing and combining the pixel data from the upsampled MSC and the high-resolution RGB image, analogous to Equation 6.8.

The core idea is to perform a linear regression directly from this combined feature set to the target CIE XYZ tristimulus values $\mathbf{X}$ (size $N \times 3$, where each row corresponds to the XYZ vector for a pixel). We seek the optimal $3 \times (d+3)$ regression matrix $\mathbf{M}$ that minimizes the error:

$$\arg\min_{\mathbf{M}} \|\mathbf{X^T} - \mathbf{M}\mathbf{s}(\widehat{\mathbf{C}}; \mathbf{P^T})\|_F \tag{6.11}$$

Here, $\mathbf{X}$ contains the target XYZ values (e.g., measured or calculated ground truth for each pixel location), $\widehat{\mathbf{C}}$ is the bilinearly interpolated MSC image at high resolution, and $\mathbf{P}$ contains the corresponding high-resolution RGB responses. The optimal regression matrix $\mathbf{M}_{\text{cc}}$ that maps the combined MSC and RGB features to XYZ is computed via the pseudo-inverse:

$$\mathbf{M}_{\text{cc}} = \mathbf{X^T}[\mathbf{s}(\widehat{\mathbf{C}}; \mathbf{P^T})]^+ \tag{6.12}$$

Crucially, this method leverages the spectral richness inherent in the (upsampled) MSC data $\widehat{\mathbf{C}}$. Since multispectral sensitivities can be designed to align more closely with the CIE colour matching functions or span the spectral space more effectively than typical RGB sensors, incorporating this information directly into the colour correction regression is hypothesised to improve accuracy and robustness, particularly against variations in illumination or for

spectrally complex surfaces where standard RGB correction might struggle.

## 6.3 Experiments

### 6.3.1 Dataset

In our experiments, we employed the publicly available CAVE multispectral image dataset (Yasuma et al., 2010), widely used in spectral imaging research. The CAVE dataset comprises 32 images, each captured at high spatial resolution (512 × 512 pixels) across 31 spectral bands spanning the visible spectrum (400–700 nm). Each scene in the dataset includes diverse spectral characteristics, encompassing a wide range of textures, colours, and materials. To simulate realistic illumination conditions, we tested our method under four different illuminants. The input RGB images were produced using a numerical integration method with the spectral sensitivity functions of Canon 1D Mark III camera. Similarly, multispectral images with Spectricity's 16-channel multispectral camera sensitivity functions (Spectricity). To make the multispectral images low-resolution, they are downsampled by a factor of 8 via bilinear interpolation, meaning they are 1/64 the size thumbnails of the ground truth.

### 6.3.2 Illuminants

The illuminants chosen for testing represent typical lighting scenarios encountered in real-world settings. Specifically, we used:

- **D65**: A standard daylight illuminant, closely resembling average midday sunlight, widely used as a reference illuminant in colour science.
- **LED V1**: A violet enhanced LED-based illuminant with a spectrum strongly peaking at longer wavelengths.
- **F2**: A cool white fluorescent illuminant, typically found in office environments.

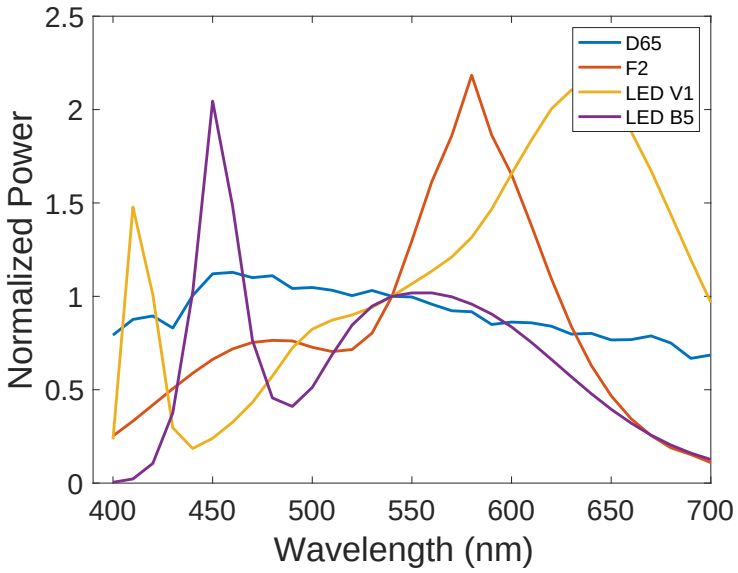


**Figure 6.2:** *Spectral Power Distributions of D65, F2, LED V1, and LED B5 illuminants.*

- **LED B5**: Another LED-based illuminant strongly peaking at shorter wavelengths, representing cool-white LED lighting conditions.

Figure 6.2 shows the spectral power distributions of these illuminants, highlighting their distinct spectral characteristics.

### 6.3.3 Algorithms

The following algorithms were evaluated and benchmarked in our experiments:

(i) **LS-Linear-RGB**: Linear Least-Squares Regression using only RGB input.

(ii) **LS-Poly-RGB**: Second-Order Polynomial Least-Squares Regression using only RGB input.

(iii) **LS-RP-RGB**: Second-Order Root-Polynomial Least-Squares Regression using only RGB input.

(iv) **LS-Linear-MSC**: Linear Least-Squares Regression using only MSC input (upsampled to RGB resolution).

(v) **LS-Linear-Fusion**: Proposed Linear Least-Squares Regression combining RGB and MSC inputs.

Methods (i), (ii), and (iii) are the traditional colour correction algorithms that use solely RGB inputs and try to estimate the corresponding XYZ target. While (ii) takes the second-order polynomial expansion of RGB, which is $[R\ G\ B\ RG\ RB\ GB\ R^2\ G^2\ B^2]$, (ii) uses the second-order root-polynomial expansion of RGB as inputs: $[R\ G\ B\ \sqrt{RG}\ \sqrt{RB}\ \sqrt{GB}\ ]$.

Method (iv) uses only upsampled MSC images as inputs. It's here to test if MSC data alone is helpful for colour correction. (v), on the other hand, is the proposed method in this chapter, which takes the fused RGB and upsampled MSC data as inputs and combines the benefits of both spatially and spectrally rich input to estimate XYZ pixels. Here for (iv) and (v), we only included simple least-squares approaches as we observed that the polynomial expansions are not bringing any significant improvement to the MSC data.

For each method, we applied a 5-fold cross-validation technique. So, for each fold, 80% of the test dataset was used to calculate the least-squares matrix with regularization (the same regularization approach defined in Chapters 4 and 5), and 20% of the test data was utilized to compute the error. This process was repeated five times to ensure testing covered all samples in the testing set, leading to the calculation of the mean error metrics from the 5-fold cross-validation.

Finally, both ground-truth and estimated XYZ values are converted to the CIELAB colour space (Hunt and Pointer, 2011) to calculate the error metrics and compare the methods.

## 6.4 Results

This research utilizes CIE Delta E as the error metric for evaluation, which measures perceptual colour differences. Thus, both ground-truth and estimated XYZ values are first converted to the CIELAB colour space (Hunt and

Pointer, 2011), and then the Delta E is calculated. We report the mean, median, and 99th percentile (99pt) Delta E values for each image, averaging these statistics across the dataset.

Table 6.1 summarises experimental results in terms of Delta E for all algorithms across four different illuminants. Lower Delta E values indicate superior perceptual accuracy.

Table 6.1: Average Delta E Metrics for the CAVE Dataset under four different illuminants

| | **D65** | | | **LED V1** | | | **F2** | | | **LED B5** | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | **Mean** | **Med** | **99pt** | **Mean** | **Med** | **99pt** | **Mean** | **Med** | **99pt** | **Mean** | **Med** | **99pt** |
| LS-Linear-RGB | 2.41 | 2.36 | 5.50 | 2.60 | 1.89 | 9.96 | 3.40 | 3.15 | 8.66 | 3.01 | 2.91 | 6.99 |
| LS-Poly-RGB | 2.17 | 2.10 | 5.25 | 2.48 | 1.96 | 9.31 | 3.16 | 2.92 | 8.63 | 2.78 | 2.66 | 6.64 |
| LS-RP-RGB | 2.15 | 2.13 | 4.67 | 1.65 | 1.31 | 6.30 | 2.80 | 2.69 | 6.81 | 2.97 | 2.87 | 6.77 |
| LS-Linear-MSC | 8.49 | 5.88 | 37.09 | 8.44 | 5.81 | 37.06 | 8.44 | 5.82 | 37.10 | 8.41 | 5.80 | 37.03 |
| LS-Linear-Fus. | 0.62 | 0.37 | 4.03 | 1.04 | 0.53 | 7.83 | 1.07 | 0.57 | 7.65 | 0.78 | 0.46 | 5.30 |

These results clearly demonstrate the superior performance of the LS-Linear-Fusion algorithm across all tested illuminants, consistently achieving significantly lower Delta E values compared to traditional regression methods using only RGB inputs. Among the classical RGB-only methods, LS-RP-RGB generally delivers the best performance, except under the LED B5 illuminant, where LS-P-RGB slightly outperforms the other two RGB-only methods. On the other hand, LS-Linear-MSC yields consistently poor results regardless of illuminant type, attributed to the limited spatial detail in the input MSC data. Although this finding discourages the direct use of upsampled MSC for colour correction, a dramatic performance improvement is observed when fusing MSC with RGB using LS-Linear-Fusion. Compared to the second-best method, LS-RP-RGB, LS-Linear-Fusion reduces the mean error by approximately 25-65%, depending on the illuminant. However, the improvement in the 99th percentile

error is less significant; indeed, under the LED V1 and F2 illuminants, the 99th percentile error for LS-Linear-Fusion is worse than that of LS-RP-RGB, likely due to errors introduced during the upsampling of the MSC data.

Figure 6.3 and 6.4 visually compare Delta E heat maps for representative scenes under different methods. LS-Linear-Fusion exhibits substantially lower and more uniform error distributions compared to other classical approaches, reaffirming the advantages of our proposed fusion approach.

This improved accuracy can be attributed to the complementary nature of RGB and MSC data, where RGB provides detailed spatial resolution and MSC contributes richer spectral content, leading to enhanced colour correction.

## 6.5 Conclusion

In this chapter, we introduced a novel MSC-enhanced colour correction approach leveraging the combined strengths of high-resolution RGB and low-resolution multispectral captures. Our comprehensive experiments on the CAVE dataset demonstrated significant accuracy improvements over classical regression approaches, validating the effectiveness and robustness of our proposed method across diverse illuminants conditions. The simplicity and powerful performance of our regression-based fusion method provide compelling evidence for the adoption of MSC-enhanced strategies in future colour correction research and applications.

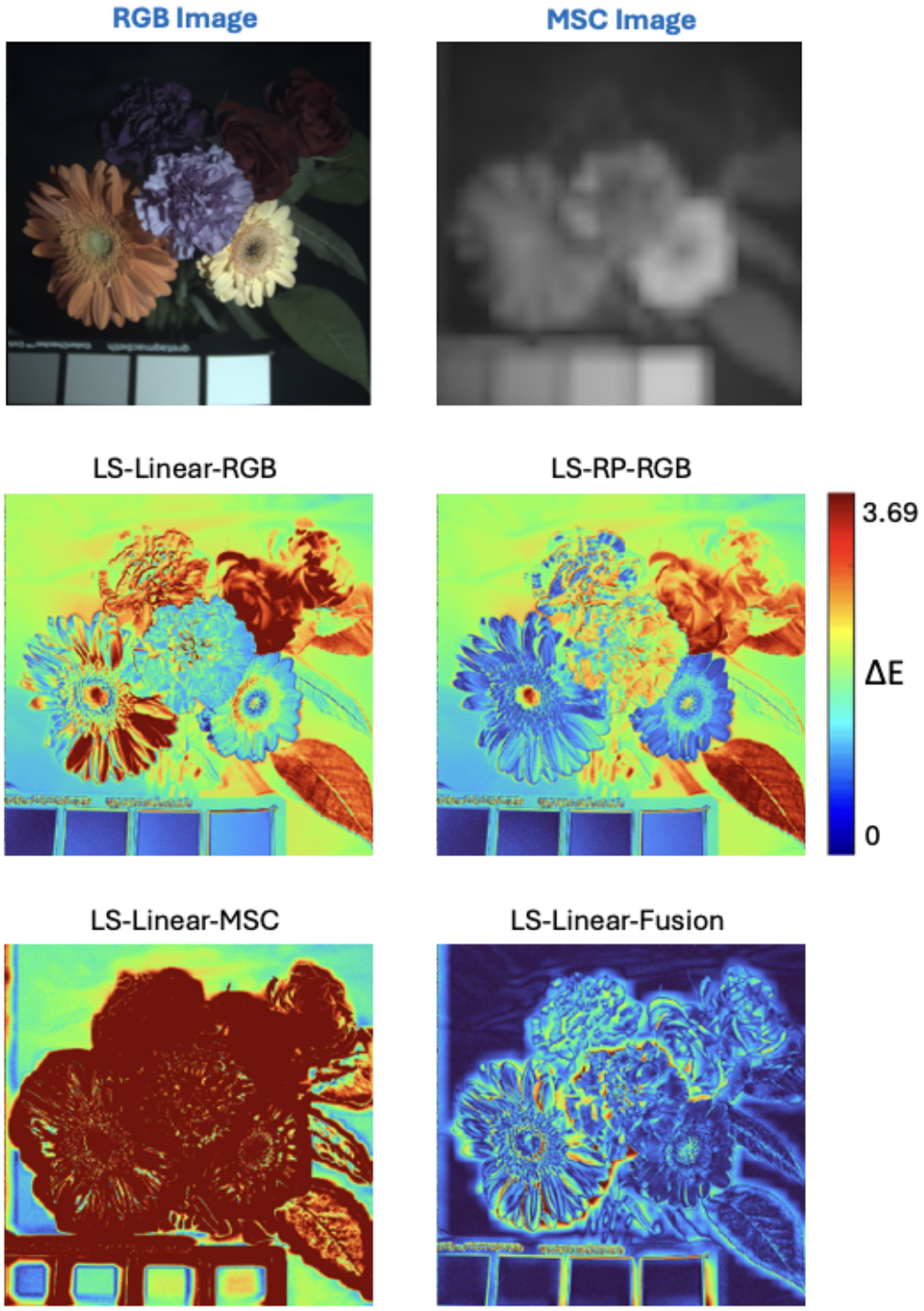


**Figure 6.3:** *The Delta E heat maps of the algorithms on an example image "Flowers" from the CAVE Dataset. For the representation of the MSC Image, we displayed the tenth channel of the upsampled multispectral data.*

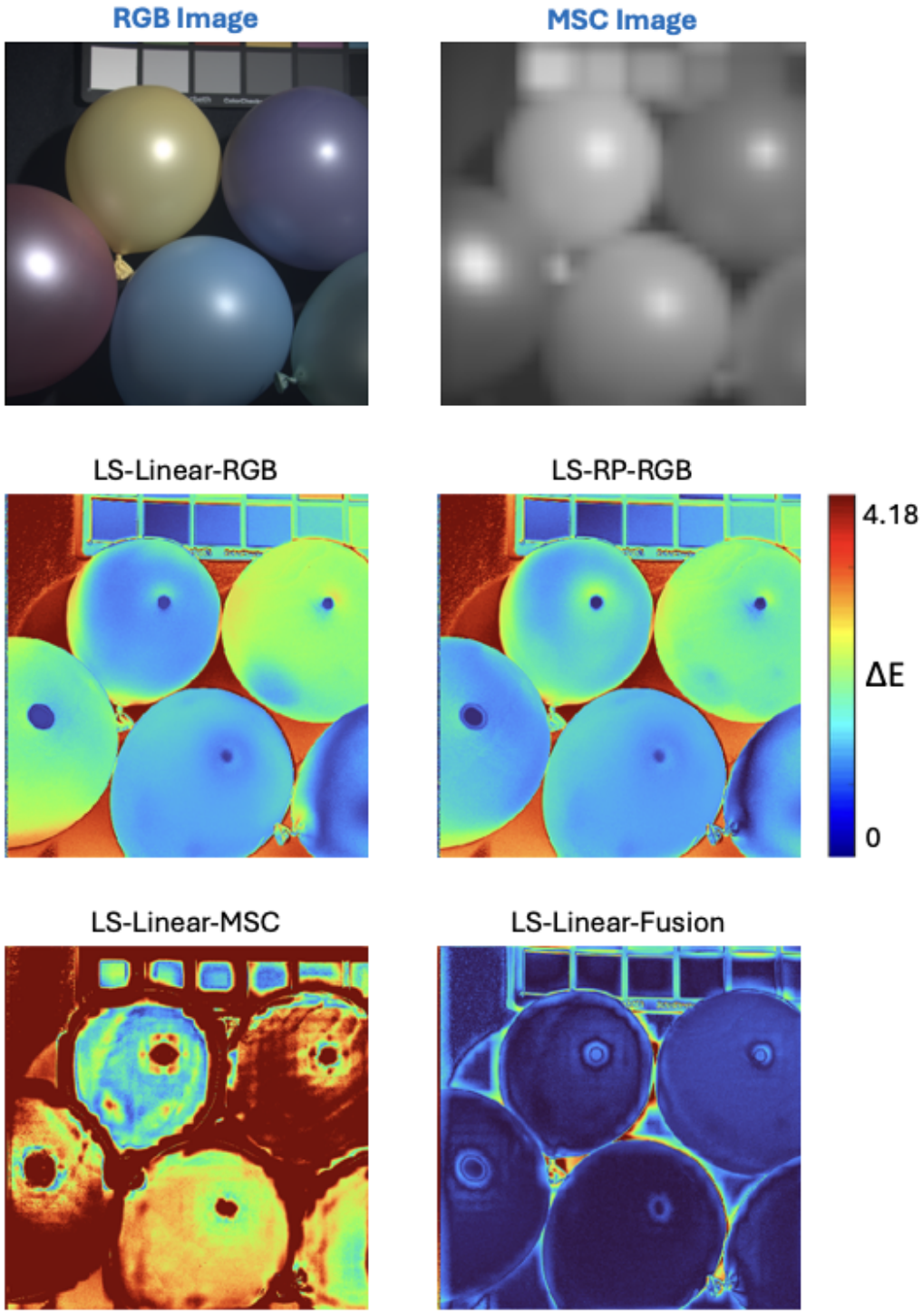


**Figure 6.4:** *The Delta E heat maps of the algorithms on an example image "Balloons" from the CAVE Dataset. For the representation of the MSC Image, we displayed the tenth channel of the upsampled multispectral data.*

# Chapter 7

# Conclusions, Limitations, and Future Work

This thesis investigates the enhancement of digital colour imaging, addressing persistent challenges in both colour correction and spectral recovery. The research presented has aimed to improve the fidelity of colour reproduction and the accuracy of spectral information retrieval by developing novel algorithms, often by integrating physical principles and innovatively combining information from different imaging modalities.

The journey began with a critical re-evaluation of colour correction techniques for standard RGB cameras. We systematically compared classical regression methods against emerging neural network solutions, revealing performance nuances and highlighting the significant challenge posed by exposure variations. This led to the development of novel exposure-invariant neural network architectures, specifically a data-augmentation strategy and a new network design that inherently maintains colour fidelity across different light levels by separating chromaticity and intensity processing. Our findings underscore the importance of robustness in practical colour correction pipelines and offer effective solutions to achieve it.

Subsequently, the focus shifted to the domain of spectral recovery, encompassing both spectral reconstruction and pan-sharpening. Recognising

that many existing algorithms, despite their sophistication, could produce physically inconsistent results, we investigated the Matrix-R post-processing framework. This universally applicable method, grounded in the theory of fundamental metamers and metameric blacks, provably refines the output of any spectral recovery algorithm by ensuring consistency with known sensor responses, without needing to retrain the original algorithm. We further extended this framework to incorporate low-dimensional linear models for spectra and to handle multispectral image data, enhancing its versatility.

Building upon the insights from physically constrained post-processing, we then developed a novel data-driven alternative: a Least-Squares (LS) regression-based post-processing technique for pan-sharpening. This method learns the optimal correction from training data and was shown to inherently respect the Matrix-R physical constraints. A key advantage of the LS approach is its independence from prior knowledge of camera spectral sensitivities, a significant practical benefit.

Finally, this thesis explored a novel way to improve colour correction by leveraging the richer information available from multispectral captures, even when these are of low spatial resolution. We successfully adapted the LS regression framework to fuse upsampled low-resolution multispectral data with corresponding high-resolution RGB imagery, directly regressing towards accurate CIE XYZ tristimulus values. This dual-camera inspired approach demonstrated a significant enhancement in colour correction accuracy compared to methods relying solely on RGB inputs, highlighting the synergistic potential of combining different sensor modalities.

## 7.1 Limitations

While the proposed methods demonstrate significant theoretical and algorithmic advancements, several practical limitations must be acknowledged regarding their real-world deployment.

The multispectral-enhanced colour correction and pan-sharpening methods discussed in this thesis rely on the assumption of precise spatial alignment between the RGB and spectral (multispectral or hyperspectral) images. In a simulation environment, this alignment is exact. However, in a practical hardware setup involving two distinct sensors (RGB and spectral), physical separation is unavoidable. This baseline distance creates a difference in viewpoint, introducing parallax errors where objects at different depths shift relative to one another, making perfect pixel-to-pixel registration impossible via simple homography. Furthermore, the two sensors may possess different fields of view, optical distortions, and exposure triggers, complicating the synchronisation and alignment process essential for the algorithms to function correctly.

A significant portion of the evaluation in this work relies on RGB and XYZ values that are mathematically synthesised from hyperspectral datasets. In practice, obtaining actual, physically measured pixel-wise XYZ ground truth for complex scenes is infeasible; typically, only sparse measurements from colour charts are available. Consequently, to train and test models on large-scale datasets, we rely on simulated image formation. While this allows for dense pixel-wise evaluation, it assumes idealised sensor models and may overlook specific hardware characteristics such as real-world noise profiles.

This thesis is primarily a theoretical and computational study, focusing on the mathematical derivation and validation of physics-based constraints and regression models. Although our proposed solutions are designed to be computationally efficient and fast, the transition to physical hardware introduces specific engineering challenges not covered here. These include the precise synchronisation of dual-camera streams and the integration of these algorithms into the firmware of embedded imaging systems.

## 7.2 Future Work

Building on the contributions and limitations identified above, several avenues for future research emerge to bridge the gap between theoretical colourimetry and practical imaging systems.

The most immediate next step is the transition from simulation to physical prototyping. Future work should involve constructing a real dual-camera rig - combining a high-resolution RGB sensor with a low-resolution spectral (multispectral or hyperspectral) camera - to test the proposed fusion algorithms for both colour correction and pan-sharpening in dynamic, real-world scenes. Since these applications rely on precise pixel alignment, this setup would necessitate the development of robust, depth-aware registration algorithms capable of handling parallax and occlusions in real-time, ensuring that the theoretical benefits of spectral fusion translate to practical image quality improvements.

In conclusion, this work provides advanced tools and deeper insights for improving the accuracy, robustness, and practical utility of digital imaging systems. The emphasis throughout has been on creating solutions that are not only effective but also understand and respect the underlying physical and perceptual principles of colour and light. These advancements pave the way for future research in creating even more faithful and versatile digital representations of our visual world.

# Bibliography


Paolo Addesso, Rocco Restaino, and Gemine Vivone. An improved version of the generalized laplacian pyramid algorithm for pansharpening. *Remote Sensing*, 13(17):3386, 2021.

D Adkins, VS Cherkassky, and ES Olson. Color mapping using neural networks. In *Color and Imaging Conference*, volume 1, pages 45–48. Society of Imaging Science and Technology, 1993.

Jonas Aeschbacher, Jiqing Wu, and Radu Timofte. In defense of shallow learned spectral reconstruction from rgb images. In *Proceedings of the IEEE International Conference on Computer Vision Workshops*, pages 471–479, 2017.

Mahmoud Afifi and Michael S Brown. Deep white-balance editing. In *Proceedings of the IEEE/CVF Conference on computer vision and pattern recognition*, pages 1397–1406, 2020.

Mahmoud Afifi, Marcus A Brubaker, and Michael S Brown. Auto white-balance correction for mixed-illuminant scenes. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 1210–1219, 2022.

Aitor Alvarez-Gila, Joost Van De Weijer, and Estibaliz Garrote. Adversarial networks for spatial context-aware spectral image reconstruction from rgb. In *Proceedings of the IEEE international conference on computer vision workshops*, pages 480–490, 2017.

Matthew Anderson, Ricardo Motta, Srinivasan Chandrasekar, and Michael Stokes. Proposal for a standard default color space for the internet—srgb. In *Color and imaging conference*, volume 4, pages 238–245. Society of Imaging Science and Technology, 1996.

B. Arad and O. Ben-Shahar. Sparse recovery of hyperspectral signal from natural RGB images. In *Proceedings of the European Conference on Computer Vision*, pages 19–34. Springer, 2016.

B. Arad, O. Ben-Shahar, R. Timofte, et al. NTIRE 2018 challenge on spectral reconstruction from RGB images. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 929–938, 2018.

B. Arad, R. Timofte, O. Ben-Shahar, Y.-T. Lin, G.D. Finlayson, et al. NTIRE 2020 challenge on spectral reconstruction from an RGB image. In *Proceed-*

*ings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 446–447, 2020.

Boaz Arad, Radu Timofte, Rony Yahel, Nimrod Morag, Amir Bernat, Yuanhao Cai, Jing Lin, Zudi Lin, Haoqian Wang, Yulun Zhang, et al. Ntire 2022 spectral recovery challenge and data set. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 863–881, 2022.

Kobus Barnard, Lindsay Martin, Brian Funt, and Adam Coath. A data set for color research. *Color Research & Application*, 27(3):147–151, 2002.

Branislav D Batinić, Miloš S Arbanas, Jovan S Bajić, Sandra R Dedijer, Vladimir M Rajs, Nikola M Laković, and Nenad R Kulundžić. Using machine learning for improvement of reflected spectrum estimations of colorimetric probe. *IEEE Transactions on Instrumentation and Measurement*, 70:1–7, 2020.

Bryce E. Bayer. Color imaging array, 1976. US Patent 3,971,065.

Willard S Boyle and George E Smith. Charge coupled semiconductor devices. *Bell System Technical Journal*, 49(4):587–593, 1970.

Gershon Buchsbaum. A spatial processor model for object colour perception. *Journal of the Franklin institute*, 310(1):1–26, 1980.

David R Bull and Fan Zhang. Digital picture formats and representations. *Communicating pictures*, pages 99–132, 2014.

Marco Buzzelli and Simone Bianco. A convolutional framework for color constancy. *IEEE Transactions on Neural Networks and Learning Systems*, 2024.

Yuanhao Cai, Jing Lin, Zudi Lin, Haoqian Wang, Yulun Zhang, Hanspeter Pfister, Radu Timofte, and Luc Van Gool. Mst++: Multi-stage spectral-wise transformer for efficient spectral reconstruction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 745–755, 2022.

Vlad C Cardei and Brian Funt. Committee-based color constancy. In *Color and Imaging Conference*, volume 7, pages 311–313. Society of Imaging Science and Technology, 1999.

D.M. Carl. *Matrix analysis and applied linear algebra*. Society for Industrial and Applied Mathematics, 2000.

A. Chakrabarti and T. Zickler. Statistics of real-world hyperspectral images. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 193–200, 2011.

Pats Chavez, Stuart C Sides, Jeffrey A Anderson, et al. Comparison of three different methods to merge multiresolution and multispectral data- landsat tm and spot panchromatic. *Photogrammetric Engineering and remote sensing*, 57(3):295–303, 1991.

Qiao Chen, Lijie Wang, and Stephen Westland. A perceptual study of linear models of spectral reflectance. In *Procedings of the IEEE Conference on Computer Science and Software Engineering*, pages 309–312, 2008.

Z. Chen, J. Wang, T. Wang, Z. Song, Y. Li, Y. Huang, L. Wang, and J. Jin. Automated in-field leaf-level hyperspectral imaging of corn plants using a cartesian robotic platform. *Computers and Electronics in Agriculture*, 183: 105996, 2021.

W. Cheney and D. Kincaid. *Linear Algebra: Theory and Applications*. The Australian Mathematical Society, 2009.

Jozef B Cohen and William E Kappauf. Metameric color stimuli, fundamental metamers, and wyszecki's metameric blacks. *The American journal of psychology*, pages 537–564, 1982.

Commission Internationale de L'eclairage. *CIE Proceedings, Vienna Session, Committee Report E-1.4. 1*, 1964.

David R Connah and Jon Y Hardeberg. Spectral recovery using polynomial models. In *Color Imaging X: Processing, Hardcopy, and Applications*, volume 5667, pages 65–75. SPIE, 2005.

Antonino Cosentino, C Heritage, S Open, and AS Antonio. Multispectral imaging and the art expert. *Spectroscopy Europe*, 27(2):6–9, 2015.

L.A. Courtenay, D. González-Aguilera, S. Lagüela, S. Del Pozo, C. Ruiz-Mendez, I. Barbero-García, C. Román-Curto, J. Cañueto, C. Santos-Durán, M.E. Cardeñoso-Álvarez, and M Roncero-Riesco. Hyperspectral imaging and robust statistics in non-melanoma skin cancer analysis. *Biomedical Optics Express*, 12(8):5107–5127, 2021.

Maryam Mohammadzadeh Darrodi, Graham Finlayson, Teresa Goodman, and Michal Mackiewicz. Reference data set for camera spectral sensitivity estimation. *Journal of the Optical Society of America a*, 32(3):381–391, 2015.

Russell L De Valois, Israel Abramov, and Gerald H Jacobs. Analysis of response patterns of lgn cells. *Journal of the optical Society of America*, 56 (7):966–977, 1966.

Weisheng Dong, Chen Zhou, Fangfang Wu, Jinjian Wu, Guangming Shi, and Xin Li. Model-guided deep hyperspectral image super-resolution. *IEEE Transactions on Image Processing*, 30:5754–5768, 2021.

Wenqian Dong, Tongzhen Zhang, Jiahui Qu, Yunsong Li, and Haoming Xia. A spatial–spectral dual-optimization model-driven deep network for hyperspectral and multispectral image fusion. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–16, 2022.

John E Dowling. *The retina: an approachable part of the brain*. Harvard University Press, 1987.

Mark S Drew and Brian V Funt. Natural metamers. *CVGIP: Image Understanding*, 56(2):139–151, 1992.

Eric Dubois. Frequency-domain methods for demosaicking of bayer-sampled color images. *IEEE Signal Processing Letters*, 12(12):847–850, 2005.

Abbas El Gamal and Helmy Eltoukhy. Cmos image sensors. *IEEE Circuits and Devices Magazine*, 21(3):6–20, 2005.

Noha El-Yamany. Robust defect pixel detection and correction for bayer imaging systems. *Electronic Imaging*, 29:46–51, 2017.

Mark D Fairchild. *Color appearance models*. John Wiley & Sons, 2013.

Mark D Fairchild. Von kries 2020: Evolution of degree of chromatic adaptation. In *Color and Imaging Conference*, volume 28, pages 252–257. Society for Imaging Science and Technology, 2020.

Hugh S Fairman, Michael H Brill, and Henry Hemmendinger. How the cie 1931 color-matching functions were derived from wright-guild data. *Color Research & Application*, 22(1):11–23, 1997.

Nawar Fdhal, Matthew Kyan, Dimitri Androutsos, and Abhay Sharma. Color space transformation from rgb to cielab using neural networks. In *Advances in Multimedia Information Processing-PCM 2009: 10th Pacific Rim Conference on Multimedia, Bangkok, Thailand, December 15-18, 2009 Proceedings 10*, pages 1011–1017. Springer, 2009.

Graham D. Finlayson and Peter Morovic. Metamer constrained colour correction. *Journal of Imaging Science and Technology*, 44(4):295–300, 2000.

Graham D Finlayson and Elisabetta Trezzi. Shades of gray and colour constancy. In *Color and imaging conference*, volume 12, pages 37–41. Society of Imaging Science and Technology, 2004.

Graham D Finlayson, Michal Mackiewicz, and Anya Hurlbert. Color correction using root-polynomial regression. *IEEE Transactions on Image Processing*, 24(5):1460–1470, 2015.

David H Foster, Kinjiro Amano, Sérgio MC Nascimento, and Michael J Foster. Frequency of metamerism in natural scenes. *Josa a*, 23(10):2359–2372, 2006.

Xueyang Fu and Xiangyong Cao. Underwater image enhancement with global–local networks and compressed-histogram equalization. *Signal Processing: Image Communication*, 86:115892, 2020.

Ying Fu, Tao Zhang, Yinqiang Zheng, Debing Zhang, and Hua Huang. Hyperspectral image super-resolution with optimized rgb guidance. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 11661–11670, 2019.

Michael E Gehm, Renu John, David J Brady, Rebecca M Willett, and Timothy J Schulz. Single-shot compressive spectral imaging with a dual-disperser architecture. *Optics express*, 15(21):14013–14027, 2007.

Arjan Gijsenij, Theo Gevers, and Joost Van De Weijer. Computational color constancy: Survey and experiments. *IEEE transactions on image processing*, 20(9):2475–2489, 2011.

Alexander FH Goetz, Gregg Vane, Jerry E Solomon, and Barrett N Rock. Imaging spectrometry for earth remote sensing. *science*, 228(4704):1147–1153, 1985.

V. Gomes, A. Mendes-Ferreira, and P. Melo-Pinto. Application of hyperspectral imaging and deep learning for robust prediction of sugar and ph levels in wine grape berries. *Sensors*, 21(10):3459, 2021.

F. Grillini, J. Thomas, and S. George. Mixing models in close-range spectral imaging for pigment mapping in cultural heritage. In *Proceedings of the International Colour Association (AIC) Conference*, pages 372–376, 2020.

John Guild. The colorimetric properties of the spectrum. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, 230(681-693):149–187, 1931.

Tanzima Habib, Phil Green, and Peter Nussbaum. Spectral reproduction: drivers, use cases and workflow. *Electronic Imaging*, 32:1–6, 2020.

K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 770–778, 2016.

Glenn E Healey and Raghava Kondepudy. Radiometric ccd camera calibration and noise estimation. *IEEE Transactions on pattern analysis and machine intelligence*, 16(3):267–276, 2002.

V. Heikkinen, R. Lenz, T. Jetsu, J. Parkkinen, M. Hauta-Kasari, and T. Jääskeläinen. Evaluation and unification of some methods for estimating reflectance spectra from RGB images. *Journal of the Optical Society of America A*, 25(10):2444–2458, 2008.

Ewald Hering. Outlines of a theory of the light sense. 1964.

Jerry L Hintze and Ray D Nelson. Violin plots: a box plot-density trace synergism. *The American Statistician*, 52(2):181–184, 1998.

Guowei Hong, M Ronnier Luo, and Peter A Rhodes. A study of digital camera colorimetric characterization based on polynomial modeling. *Color Research & Application*, 26(1):76–84, 2001.

Steven Hordley, Graham Finalyson, and Peter Morovic. A multi-spectral image database and its application to image rendering across illumination. In *Third International Conference on Image and Graphics (ICIG'04)*, pages 394–397. IEEE, 2004.

Steven D Hordley and Graham D Finlayson. Reevaluation of color constancy algorithm performance. *Journal of the Optical Society of America A*, 23(5): 1008–1020, 2006.

Roger A Horn and Charles R Johnson. Norms for vectors and matrices. *Matrix analysis*, pages 313–386, 1990.

Jin-Fan Hu, Ting-Zhu Huang, Liang-Jian Deng, Tai-Xiang Jiang, Gemine Vivone, and Jocelyn Chanussot. Hyperspectral image super-resolution via deep spatiospectral attention convolutional neural networks. *IEEE Transactions on Neural Networks and Learning Systems*, 33(12):7251–7265, 2021.

Jin-Fan Hu, Ting-Zhu Huang, Liang-Jian Deng, Hong-Xia Dou, Danfeng Hong, and Gemine Vivone. Fusformer: A transformer-based fusion network for hyperspectral image super-resolution. *IEEE Geoscience and Remote Sensing Letters*, 19:1–5, 2022a.

Jin-Fan Hu, Ting-Zhu Huang, Liang-Jian Deng, Hong-Xia Dou, Danfeng Hong, and Gemine Vivone. Fusformer: A transformer-based fusion network for hyperspectral image super-resolution. *IEEE Geoscience and Remote Sensing Letters*, 19:1–5, 2022b.

Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q Weinberger. Densely connected convolutional networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4700–4708, 2017.

Robert William Gainer Hunt and Michael R Pointer. *Measuring colour*. John Wiley & Sons, 2011.

Leo M Hurvich and Dorothea Jameson. An opponent-process theory of color vision. *Psychological review*, 64(6p1):384, 1957.

Jong Pil Hwang, Sungmin Kim, and Chang Kyu Park. Development of a color matching algorithm for digital transfer textile printing using an artificial neural network and multiple regression. *Textile Research Journal*, 85(10): 1076–1082, 2015.

F.H. Imai and R.S. Berns. High-resolution multi-spectral image archives: a hybrid approach. In *Color and imaging conference*, volume 1998, pages 224–227. Society for Imaging Science and Technology, 1998.

J. Jiang, D. Liu, J. Gu, and S. Süsstrunk. What is the space of spectral sensitivity functions for digital color cameras? In *Proceedings of the IEEE Workshop on Applications of Computer Vision*, pages 168–179, 2013.

Deane B Judd, David L MacAdam, Günter Wyszecki, HW Budde, HR Condit, ST Henderson, and JL Simonds. Spectral distribution of typical daylight as a function of correlated color temperature. *Journal of the Optical Society of America*, 54(8):1031–1040, 1964.

Henry R Kang and Peter G Anderson. Neural network applications to the color scanner and printer calibrations. *Journal of Electronic Imaging*, 1(2): 125–135, 1992.

Wen-Chung Kao, Sheng-Hong Wang, Lien-Yang Chen, and Sheng-Yuan Lin. Design considerations of color image processing pipeline for digital cameras. *IEEE Transactions on Consumer Electronics*, 52(4):1144–1152, 2007.

Ryusuke Kawakami, Yasuhiro Mukaigawa, Takeo Kanade, Norimichi Tsumura, and Katsushi Ikeuchi. High-resolution hyperspectral imaging via matrix factorization. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2329–2336, 2011.

Hafeez Anwar Khan, Jean-Baptiste Thomas, Jon Yngve Hardeberg, and Olivier Laligant. Illuminant estimation in multispectral imaging. *Journal of the Optical Society of America A*, 34(7):1085–1098, 2017.

Filippos Kokkinos and Stamatios Lefkimmiatis. Iterative joint image demosaicking and denoising using a residual denoising network. *IEEE Transactions on Image Processing*, 28(8):4177–4188, 2019.

Abdullah Kucuk, Graham D Finlayson, Rafal Mantiuk, and Maliha Ashraf. Performance comparison of classical methods and neural networks for colour correction. *Journal of Imaging*, 9(10):214, 2023.

Jeffrey C Lagarias, James A Reeds, Margaret H Wright, and Paul E Wright. Convergence properties of the nelder–mead simplex method in low dimensions. *SIAM Journal on optimization*, 9(1):112–147, 1998.

Trevor D Lamb. Why rods and cones? *Eye*, 30(2):179–185, 2016.

Charis Lanaras, Emmanuel Baltsavias, and Konrad Schindler. Hyperspectral super-resolution by coupled spectral unmixing. In *Proceedings of the IEEE international conference on computer vision*, pages 3586–3594, 2015.

Pierre-Jean Lapray, Xiuping Wang, Jean-Baptiste Thomas, and Pierre Gouton. Multispectral filter arrays: recent advances and practical implementation. *Sensors*, 14(11):21626–21659, 2014.

Chongyi Li, Saeed Anwar, and Fatih Porikli. Underwater scene prior inspired deep underwater image and video enhancement. *Pattern recognition*, 98: 107038, 2020a.

J. Li, C. Wu, R. Song, Y. Li, and F. Liu. Adaptive weighted attention network with camera spectral sensitivity prior for spectral reconstruction from RGB images. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 462–463, 2020b.

Xin Li, Bahadir Gunturk, and Lei Zhang. Image demosaicing: A systematic survey. In *Visual Communications and Image Processing 2008*, volume 6822, pages 489–503. SPIE, 2008.

Yuheng Li, Qiang Fu, and Wolfgang Heidrich. Multispectral illumination estimation using deep unrolling network. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 2326–2335, 2021.

Y.-T. Lin and G.D. Finlayson. Exposure invariance in spectral reconstruction from RGB images. In *Proceedings of the Color and Imaging Conference*, pages 284–289, 2019.

Y.-T. Lin and G.D. Finlayson. Physically plausible spectral reconstruction from RGB images. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 532–533, 2020a.

Y.-T. Lin and G.D. Finlayson. Physically plausible spectral reconstruction. *Sensors*, 20(21):6399, 2020b.

Y.-T. Lin and G.D. Finlayson. Reconstructing spectra from rgb images by relative error least-squares regression. In *Proceedings of Color and Imaging Conference*, pages 264–269, 2020c.

Yi-Tun Lin. *A Practical Study on Recovering Spectra from RGB Images*. PhD thesis, University of East Anglia, 2023.

Yi-Tun Lin and Graham D Finlayson. Evaluating the performance of different cameras for spectral reconstruction. In *Proceedings of Color and Imaging Conference*, pages 213–218, 2022.

Yi-Tun Lin and Graham D Finlayson. A rehabilitation of pixel-based spectral reconstruction from rgb images. *Sensors*, 23(8):4155, 2023.

Yi-Tun Lin, Graham D Finlayson, and Abdullah Kucuk. An optimality property of matrix-r theorem, its extension, and the application to hyperspectral pan-sharpening. In *Color and Imaging Conference*, volume 31, pages 144–149. Society for Imaging Science and Technology, 2023.

Jianjun Liu, Zebin Wu, Liang Xiao, and Xiao-Jun Wu. Model inspired autoencoder for unsupervised hyperspectral image super-resolution. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–12, 2022.

R. Lukac and K.N. Plataniotis. Color filter arrays: design and performance analysis. *IEEE Transactions on Consumer Electronics*, 2005.

M. Lv, T. Chen, Y. Yang, T. Tu, N. Zhang, W. Li, and W. Li. Membranous nephropathy classification using microscopic hyperspectral imaging and tensor patch-based discriminative linear regression. *Biomedical Optics Express*, 12(5):2968–2978, 2021.

Lindsay MacDonald and Katarina Mayer. Camera colour correction using neural networks. In *London Imaging Meeting*, volume 2, pages 54–57. Society for Imaging Science and Technology, 2021.

Laurence T Maloney. Evaluation of linear models of surface spectral reflectance with small numbers of parameters. *Journal of the Optical Society of America A*, 3(10):1673–1683, 1986.

L.T. Maloney and B.A. Wandell. Color constancy: a method for recovering surface spectral reflectance. *Journal of the Optical Society of America A*, 3 (1):29–33, 1986.

Henrique S Malvar, Li-wei He, and Ross Cutler. High-quality linear interpolation for demosaicing of bayer-patterned color images. In *2004 IEEE international conference on acoustics, speech, and signal processing*, volume 3, pages iii–485. IEEE, 2004.

Alamin Mansouri, Franck S Marzani, Jon Yngve Hardeberg, and Pierre Gouton. Optical calibration of a multispectral imaging system based on interference filters. *Optical Engineering*, 44(2):027004–027004, 2005.

James Clerk Maxwell. Xviii.—experiments on colour, as perceived by the eye, with remarks on colour-blindness. *Earth and Environmental Science Transactions of the Royal Society of Edinburgh*, 21(2):275–298, 1857.

Calvin S McCamy, Harold Marcus, James G Davidson, et al. A color-rendition chart. *J. App. Photog. Eng*, 2(3):95–99, 1976.

Gary C McDonald. Ridge regression. *Wiley Interdisciplinary Reviews: Computational Statistics*, 1(1):93–100, 2009.

Yusukex Monno, Sunao Kikuchi, Masayuki Tanaka, and Masatoshi Okutomi. A practical one-shot multispectral imaging system using a single image sensor. *IEEE Transactions on Image Processing*, 24(10):3048–3059, 2015.

P. Morovic and G.D. Finlayson. Metamer-set-based approach to estimating surface reflectance from camera RGB. *Journal of the Optical Society of America A*, 23(8):1814–1822, 2006.

Rang MH Nguyen, Dilip K Prasad, and Michael S Brown. Training-based spectral reconstruction from a single rgb image. In *Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part VII 13*, pages 186–201. Springer, 2014.

Jorge Nunez, Xavier Otazu, Octavi Fors, Albert Prades, Vicenc Pala, and Roman Arbiol. Multiresolution-based image fusion with additive wavelet decomposition. *IEEE Transactions on Geoscience and Remote sensing*, 37 (3):1204–1211, 1999.

James M Palmer. Radiometry and photometry: Units and conversions. *Handbook of Optics*, 3:7–1, 2010.

C. Pane, G. Manganiello, N. Nicastro, T. Cardi, and F. Carotenuto. Powdery mildew caused by erysiphe cruciferarum on wild rocket (diplotaxis tenuifolia): Hyperspectral imaging and machine learning modeling for non-destructive disease detection. *Agriculture*, 11(4):337, 2021.

Jussi PS Parkkinen, Jarmo Hallikainen, and Timo Jaaskelainen. Characteristic spectra of munsell colors. *Journal of the Optical society of America A*, 6(2): 318–322, 1989.

Ruben Pastilha and Anya Hurlbert. Seeing and sensing temporal variations in natural daylight. *Progress in Brain Research*, 273(1):275–301, 2022.

Roger Penrose. A generalized inverse for matrices. In *Mathematical proceedings of the Cambridge philosophical society*, volume 51, pages 406–413. Cambridge University Press, 1955.

M. Picollo, C. Cucci, A. Casini, and L. Stefani. Hyper-spectral imaging technique in the cultural heritage field: New possible scenarios. *Sensors*, 20(10): 2843, 2020.

Gaurav Sharma. Color fundamentals for digital imaging. In *Digital color imaging handbook*, pages 1–114. CRC press, 2017.

Gaurav Sharma, Wencheng Wu, and Edul N Dalal. The ciede2000 color-difference formula: Implementation notes, supplementary test data, and mathematical observations. 30(1):21–30, 2005.

Zhiwei Shi, Chen Chen, Zhihui Xiong, Ding Liu, and Feng Wu. HSCNN+: Advanced cnn-based hyperspectral recovery from rgb images. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 952–960, 2018.

Gursharn Singh, Anand Mittal, et al. Various image enhancement techniques-a critical review. *International Journal of Innovation and Scientific Research*, 10(2):267–274, 2014.

Thomas Smith and John Guild. The cie colorimetric standards and their use. *Transactions of the optical society*, 33(3):73, 1931.

Vivianne C Smith and Joel Pokorny. Spectral sensitivity of the foveal cone photopigments between 400 and 500 nm. *Vision research*, 15(2):161–171, 1975.

Spectricity. S1 multispectral image sensor and camera module. https://spectricity.com/product/. Accessed: 18.01.2025.

Tarek Stiebel, Philipp Seltsam, and Dorit Merhof. Enhancing deep spectral super-resolution from rgb images by enforcing the metameric constraint. In *Proceedings of VISIGRAPP (4: VISAPP)*, pages 57–66, 2020.

Walter Stanley Stiles and J MNPL Burch. Npl colour-matching investigation: final report (1958). *Optica Acta: International Journal of Optics*, 6(1):1–26, 1959.

Andrew Stockman, Donald IA MacLeod, and Nancy E Johnson. Spectral sensitivities of the human cones. *Journal of the Optical Society of America A*, 10(12):2491–2521, 1993.

Andrew Stockman, David H Brainard, et al. Color vision mechanisms. *OSA handbook of optics*, 3:11–1, 2010.

A.N. Tikhonov, A.V. Goncharsky, V.V. Stepanov, and A.G. Yagola. *Numerical Methods for the Solution of Ill-posed Problems*. Springer, 1995.

Carlo Tomasi and Roberto Manduchi. Bilateral filtering for gray and color images. In *Sixth international conference on computer vision (IEEE Cat. No. 98CH36271)*, pages 839–846. IEEE, 1998.

O. Torun and S.E. Yuksel. Unsupervised segmentation of lidar fused hyperspectral imagery using pointwise mutual information. *International Journal of Remote Sensing*, 42(17):6461–6476, 2021.

Te-Ming Tu, Shun-Chi Su, Hsuen-Chyun Shyu, and Ping S Huang. A new look at ihs-like image fusion methods. *Information fusion*, 2(3):177–186, 2001.

Gemine Vivone, Luciano Alparone, Jocelyn Chanussot, Mauro Dalla Mura, Andrea Garzelli, Giorgio A Licciardi, Rocco Restaino, and Lucien Wald. A critical comparison among pansharpening algorithms. *IEEE Transactions on Geoscience and Remote Sensing*, 53(5):2565–2586, 2014.

Marinos Vlachos and Dimitrios Skarlatos. An extensive literature review on underwater image colour correction. *Sensors*, 21(17):5690, 2021.

Michael J Vrhel, Ron Gershon, and Lawrence S Iwan. Measurement and analysis of object reflectance spectra. *Color Research & Application*, 19(1):4–9, 1994.

B.A. Wandell. The synthesis and analysis of color images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, (1):2–13, 1987.

W. Wang, L. Ma, M. Chen, and Q. Du. Joint correlation alignment-based graph neural network for domain adaptation of multitemporal hyperspectral remote sensing images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 14:3170–3184, 2021.

Geoffrey I. Webb. *Overfitting*, pages 744–744. Springer US, Boston, MA, 2010. ISBN 978-0-387-30164-8.

Qi Wei, Nicolas Dobigeon, and Jean-Yves Tourneret. Bayesian fusion of multiband images. *IEEE Journal of Selected Topics in Signal Processing*, 9(6): 1117–1127, 2015.

Claude Weisbuch. Historical perspective on the physics of artificial lighting. *Comptes Rendus Physique*, 19(3):89–112, 2018.

William David Wright. A re-determination of the trichromatic coefficients of the spectral colours. *Transactions of the Optical Society*, 30(4):141, 1929.

G. Wyszecki. Valenzmetrische untersuchung des zusammenhanges zwischen normaler und anomaler trichromasie. 1953.

Günther Wyszecki and Walter Stanley Stiles. *Color science: concepts and methods, quantitative data and formulae*. John wiley & sons, 2000.

Bryan Ning Xia, Yuan Gong, Yizhe Zhang, and Christian Poellabauer. Second-order non-local attention networks for person re-identification. In *Proceedings of the IEEE/CVF Conference on Computer Vision*, pages 3760–3769, 2019.

Jingxiang Yang, Yong-Qiang Zhao, and Jonathan Cheung-Wai Chan. Hyperspectral and multispectral image fusion via deep two-branches convolutional neural network. *Remote Sensing*, 10(5):800, 2018.

Fumihito Yasuma, Tomoo Mitsunaga, Daisuke Iso, and Shree K Nayar. Generalized assorted pixel camera: postcapture control of resolution, dynamic range, and spectrum. *IEEE transactions on image processing*, 19(9):2241–2253, 2010.

Naoto Yokoya, Takehisa Yairi, and Akira Iwasaki. Coupled non-negative matrix factorization (cnmf) for hyperspectral and multispectral data fusion: Application to pasture classification. In *2011 IEEE international geoscience and remote sensing symposium*, pages 1779–1782. IEEE, 2011.

Thomas Young. Ii. the bakerian lecture. on the theory of light and colours. *Philosophical transactions of the Royal Society of London*, (92):12–48, 1802.

Jingang Zhang, Runmu Su, Qiang Fu, Wenqi Ren, Felix Heide, and Yunfeng Nie. A survey on computational spectral reconstruction methods from rgb to hyperspectral imaging. *Scientific reports*, 12(1):11905, 2022.

Jingyi Zhang, Jianjun Liu, Jinlong Yang, and Zebin Wu. Crossed dual-branch u-net for hyperspectral image super-resolution. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2023.

Lei Zhang, Jiangtao Nie, Wei Wei, Yong Li, and Yanning Zhang. Deep blind hyperspectral image super-resolution. *IEEE Transactions on Neural Networks and Learning Systems*, 32(6):2388–2400, 2020.

Yonghui Zhao and Roy S Berns. Image-based spectral reflectance reconstruction using the matrix r method. *Color Research & Application*, 32(5):343–351, 2007.

Yuanjie Zheng, Stephen Lin, Chandra Kambhamettu, Jingyi Yu, and Sing Bing Kang. Single-image vignetting correction. *IEEE transactions on pattern analysis and machine intelligence*, 31(12):2243–2256, 2008.