ugr Universidad
de Granada

Design considerations for line-scan multi-spectral imaging systems

Application in reflectance and color measurements

by

Timo Eckhard
Dipl.-Ing.(FH), University of Applied Science Ulm
M.Sc., Erasmus Mundus Master CIMET

A dissertation
submitted in partial fulfillment of the requirements for the

Doctor of Philosophy

at the

Universidad de Granada
(University of Granada)

Doctoral Programme: Física y Ciencias del Espacio

31 May 2015

Design considerations for line-scan multi-spectral imaging systems

Timo Eckhard

Abstract

Spectral imaging systems have been used for spectral measurements for several decades, mainly for scientific purposes, and were usually linked to costly applications. Thanks to research in imaging science and machine learning and the technological advancement in recent years, spectral imaging became feasible for various industrial and even consumer applications.

This dissertation deals with line-scan multi-spectral imaging systems for spectral reflectance and color measurements. The most important aspects under consideration are optimization of spectral properties of the optical components, image registration and estimation of surface spectral reflectances. We focus on a particular system design, in which multiple color filtered RGB images with distinct spectral content are acquired at the same time. These images correspond to different viewpoints of the scanning scene due to the mechanical arrangement of camera sensor and optics.

By optimizing the system's optical component spectral properties, the amount of spectral information acquired can be increased and the spectral reflectance estimation can be improved. We propose a filter selection framework and demonstrate that optimization for various line-scan system configurations results in an improvement of spectral and color measurement performance. Multi-channel image registration is required to account for viewpoint differences and other sources of image channel misalignment. We develop a calibration scheme for planar scanning objects and propose scene-adaptive registration for non-planar scanning objects. For our 12-channel laboratory imaging system, subpixel accuracy is achieved. Based on the registered multi-channel image data, spectral reflectance estimation can be performed. Physical and empirical estimation methods are considered, and we propose a logarithmic kernel function for kernel ridge regression. We experimentally compare performance of various estimation methods for simulated and measured camera response data and consider different noise levels and number of spectral channels. Empirical estimation performance is influenced by model training. We compare various training sample selection approaches and propose an application dependent selection scheme. Further, adaptive training methods from related literature are unified conceptually and evaluated systematically.

We show that the aforementioned aspects of line-scan multi-spectral imaging system design are critical for spectral and color measurement, and that application specific design is often beneficial to improve system performance.

Consideraciones para el diseño de sistemas de imagen multiespectral de escáner en línea.

Resumen

Los sistemas de imagen espectral se han utilizado para medidas espectrales desde hace varias décadas, en especial con propósitos científicos, ligados habitualmente a aplicaciones de alto coste. Gracias a los recientes avances de la investigación en los campos del tratamiento de imágenes y el aprendizaje automático, la captura de imágenes espectrales ha podido popularizarse y extenderse para su uso en varias aplicaciones industriales e incluso de consumo.

Esta tesis trata sobre sistemas de imagen multiespectral de tipo escáner en línea para medidas de reflectancia espectral y del color. Los aspectos más relevantes que se consideran son la optimización de las propiedades espectrales de los componentes ópticos, el registro de imágenes y la estimación espectral de las reflectancias de los objetos escaneados. Nos centramos en el diseño de un sistema en particular, en el cual se adquieren simultáneamente múltiples imágenes con diferente contenido espectral, correspondientes a diferentes filtros situados delante de varias lentes que forman imagen sobre un sensor RGB en línea. Estas imágenes se corresponden con diferentes perspectivas de la escena escaneada, que se originan debido a la disposición de las diferentes lentes a lo largo del sensor de imagen.

Mediante la optimización de las propiedades espectrales de los componentes ópticos del sistema, puede incrementarse la cantidad de información espectral que se adquiere, y también mejorarse la calidad de la estimación espectral de las reflectancias. Proponemos un sistema específico de selección de filtros de color, y demostramos que el proceso de optimización para varias configuraciones del escáner en línea resulta en una mejora de las prestaciones de medida espectral y del color. El registro de las imágenes multi-canal es necesario para descontar las diferencias producidas por las distintas perspectivas de las lentes, y también otras fuentes de desalineación presentes en nuestro sistema. Hemos desarrollado un método de calibrado para objetos planos, y proponemos también un método de registro adaptativo para objetos con volumen. Para nuestro sistema, hemos conseguido un registro con precisión por debajo de un pixel, en condiciones de laboratorio. Utilizando las imágenes multi-canal ya registradas, podemos realizar la estimación espectral. Consideramos dos categorías principales de algoritmos de estimación: los basados en modelos físicos de adquisición de imágenes, y los puramente empíricos. Proponemos además una función logarítmica de kernel para la estimación espectral basada en métodos de kernel con regularización. Comparamos mediante varios experimentos las prestaciones de los diferentes algoritmos de estimación espectral para respuestas de sensores simuladas y reales, considerando diferentes niveles de ruido en las simulaciones, y diferente número de canales en nuestro sistema de captura. Las prestaciones de los algoritmos de estimación empíricos dependen del conjunto de muestras que se utilice como entrenamiento. Comparamos varios sistemas de selección de muestras de entrenamiento, y proponemos un sistema propio de selección dependiente de la aplicación específica para la que vaya a utilizarse la estimación. Además, se propone una unificación conceptual de los algoritmos de entrenamiento adaptativos encontrados en la literatura, que se evalúan también de forma sistemática.

En este estudio, demostramos que los aspectros mencionados anteriormente resultan críticos para la medida espectral y del color en sistemas multiespectrales de escáner en línea, y que diseñar el sistema teniendo en cuentala aplicación específica para la cual se utilizará es a menudo beneficioso para mejorar las prestaciones del mismo.

Author declaration

El doctorando Timo Eckhard y los directores de la tesis Eva Maria Valero Benito y Javier Hernández Andrés garantizamos, al firmar esta tesis doctoral, que el trabajo ha sido realizado por el doctorando bajo la dirección de los directores de la tesis y hasta donde nuestro conocimiento alcanza, en la realización del trabajo, se han respetado los derechos de otros autores a ser citados, cuando se han utilizado sus resultados o publicaciones.

Granada, 31 May 2015

Directores de la Tesis

Fdo.: ______________________________

Fdo.: ______________________________

Doctorando

Fdo.: ______________________________

For Jia

“What we build is bigger than the sum of two.”

Kings of Convenience

Acknowledgement

I would like to express gratitude to my supervisors Prof. Eva Maria Valero Benito and Prof. Javier Hernández Andrés, for their guidance, advice and support during my PhD studies at the Color Imaging Lab of the University of Granada. I believe it is also Eva and Javier who aroused my interest in the spectral imaging domain.

The open and friendly working atmosphere in the Optics Department led to my involvement in various other research projects alongside my own dissertation. I would like to thank the members of the Color Imaging Lab for the opportunities I had to contribute, to share and to build up something new. Apart from the aforementioned, I would like to name Prof. Juan Luis Nieves Gómez, Prof. Rafael Huertas Roa, Prof. Javier Romero and Miguel Ángel Martínez Domingo.

The research project related to my dissertation was funded by the company Chromasens GmbH from Konstanz in Germany. I would like to express my gratitude towards CEO Markus Schnitzlein, who initiated the fruitful collaboration between Chromasens and the Color Imaging Lab. I also would like to thank Maximilian Klammer from Chromasens for sharing his technical expertise and for his thoughtful suggestions for my work.

Many thanks to Prof. Roy S. Berns, who accepted me as a visiting scholar in his research group at the Munsell Color Science Laboratory of the Rochester Institute of Technology.

I owe my deepest gratitude to my parents Ingrid and Martin, and my brother Steffen. Whenever I leave my comfort zone, I know undoubtedly that I have them to fall back on in time of need. The most important person in my life is my wife Jia. I thank her for her dedication, support and patience.

Contents

“种一棵树最好的时间是十年前，其次是现在。”

From Chinese:
The best time to plant a tree was ten years ago.
The second best time is now.

Chinese proverb

1

Introduction

Contents

1.1 Motivation

Conventional visual imaging systems quantify the spatial distribution of radiances reflected from an image scene. The radiance sensed by an image sensor is typically in the visible range of the electromagnetic spectrum of light. In the acquisition process, image formation can be seen as the projection of a 3-dimensional image scene on a 2-dimensional image sensor. Historically, the interest in imaging was motivated by the desire to preserve *what we see.* Acquisition of the first permanent photograph on pewter coated with bitumen in 1826 by Joseph Nicéphore Niépce and subsequent development of photographic film pioneered by George Eastman in the late 19th century [1], the invention of the CCD sensor at AT&T Bell Labs by Willard Boyle and George E. Smith in 1969 and the invention of the first digital still camera by Steven Sasson in 1975 are just some milestones on this quest. In parallel and often in conjunction with each other were contributions by researchers from various disciplines that helped developing models of the physical processes involved in imaging. Further, research on the human visual system helped us understand that *what we see* is more than a spatially and spectrally quantized version of scene radiance sensed by our eyes.

At this point in time, we know the physical relations that govern light-matter interaction very well. We have a clear understanding that the spectral radiance reflected from an image scene is a multi-dimensional function of the spectral power distribution of the incident light, the illumination and observation angles and physical properties of scene objects. So this function actually contains more information than *what we see.* Obviously there exists a great interest in its measurement, as knowing this function would provide information required to solve many tasks that are done visually (by human observer or machine vision). However, accurate measurement is tedious and time consuming. Fortunately, many visual tasks only require a fraction of that information, corresponding to specific material properties. From a technological point of view it is already feasible to sense some of them, for instance *spectral reflectance factor*, which describes the relative amount of light reflected from the surface for a certain illumination and observation geometry. It can be measured point-wise with a spectroradiometer. With optical imaging technologies, spectral reflectance factor can be measured spatially, but there exist limitations with respect to sampling and measurement time.

Research in imaging science has led to the development of various approaches for so-called *spectral imaging*, which also can be used to obtain reflectance factor measurements. This work is related to multi-spectral imaging systems, which sense multiple distinct portions of the spectrum of reflected scene radiance instead of the entire spectrum of reflected light. By incorporating prior knowledge of the physical process of image acquisition or by using machine learning techniques, it is under certain conditions possible to recover the spectral reflectance factor data from such multi-spectral images.

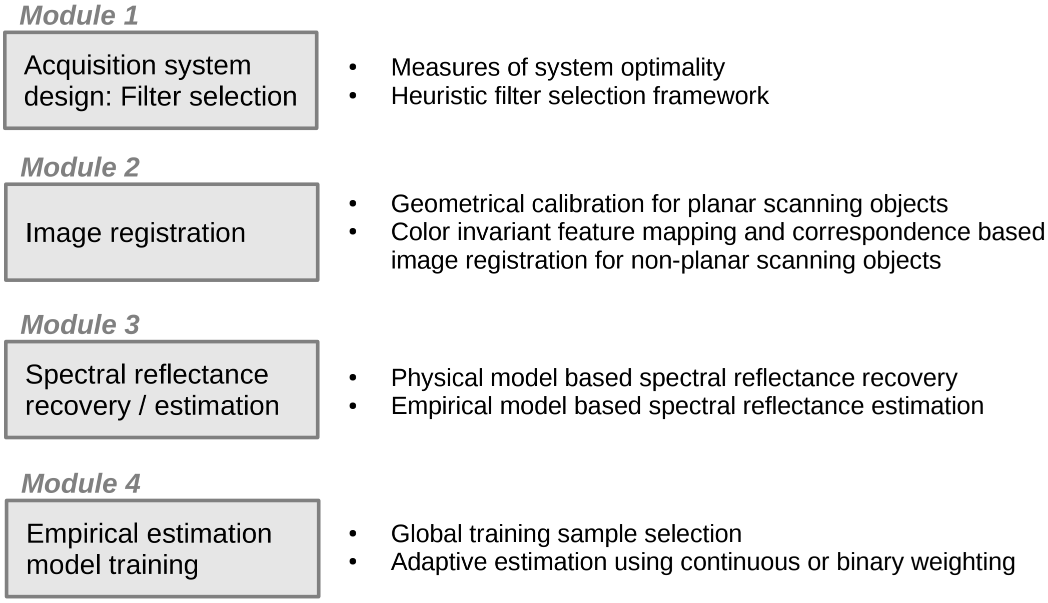
We focus on line-scan multi-spectral imaging, which means that an image of the scene is acquired line by line while translating either camera or scene object with respect to the other. The scanning principle applies especially well (but not exclusively) to applications with moving objects, for instance web-based industrial inspection. The specific imaging principle considered involves simultaneous acquisition of multiple color filtered RGB images with distinct spectral content.

In the design phase of a multi-spectral imaging system, a successful system implementation requires the consideration of various aspects related to system component design and data processing. Fundamental for the development of a specific imaging application is a sound understanding of the physical processes involved in the image formation. Based on this, mathematical models can be build, system optimality quantified and optimization performed accordingly. In this dissertation, we focus on these aspects specifically for line-scan multi-spectral imaging systems.

1.2 Structure of the work

This dissertation is divided into seven chapters that envelope the four modules of multi-spectral image system design considered in this work. These modules are optimization of spectral properties of the optical components, line-scan image registration, spectral reflectance recovery and estimation and model training in empirical estimation. We illustrate these modules in Figure 1.1.

Figure 1.1: Modules of the dissertation.



In Chapter 2, we provide necessary background information for the following chapters. We start by introducing physical models for light reflection and color perception, which are the basis of color and spectral measurements. After that, we introduce different spectral imaging technologies, the digital camera working principle and finally the line-scan multi-spectral imaging principle considered in this work. This chapter also contains a description of the experimental framework considered for system evaluation, including definitions of error functions to measure performance and model validation techniques. At last, we introduce several system configurations considered in this dissertation, as well as the datasets that are used in experiments.

Chapter 3 is related to the first module of this dissertation, namely acquisition system design. We consider system design as a multivariate optimization problem and focus on spectral properties of optical system components and system optimality criteria. After that, we direct our focus to filter selection for line-scan multi-spectral camera (LSMSC) systems, for which we propose a specific framework incorporating heuristics to reduce the computational burden of the selection process. Theoretical considerations of filter selection are verified in a set of experimental case studies, in which we exemplify LSMSC filter selection for various system configurations and application cases. For the case of a 12-channel LSMSC, we had our optimal filters manufactured and evaluated the performance in practice with our laboratory prototype LSMSC system.

Chapter 4 is devoted to line-scan multi-spectral image registration, which is a necessary data processing step of LSMSC systems and the second module of this dissertation. We distinguish two types of registration procedures. The first applies to planar scanning objects and can be considered as a geometrical calibration of the system. We analyze type and extend of image misalignment for our 12-channel laboratory LSMSC empirically and propose a polynomial and a B-spline based model to correct channel mis-alignment. We embed these models in a general framework for geometrical calibration. A calibration target is designed and semi-automatic displacement model fitting and image registration performance are evaluated. We confirm the validity of the models and the practical applicability of our geometrical calibration framework by experiment.

The second registration procedure can be considered as an extension to the first, and is required when non-planar scanning objects are considered. For this type of scanning objects, geometrical calibration can not ensure that a physical object location of the scanning scene maps to the same location in all image channels. Hence, a scene dependent registration is required. In essence, our proposed solution to this problem contains a color invariant feature mapping and pixel-wise correspondence estimation via block-matching. We evaluate each process individually by experiments and illustrate the functionality of our approach.

Chapter 5 deals with the recovery/estimation of spectral reflectance factor from multi-spectral image data. This is the third module of the dissertation. We introduce various approaches based on a physical model of the acquisition process or machine learning based techniques that use empirical data. We discuss their properties and demonstrate some specifically important features experimentally. The experimental section also contains the comparison of estimation performance for systems with different number of spectral image channels and various levels of simulated camera response noise.

Chapter 6 is related to the training process in empirical estimation models, which is the last module of this dissertation. We consider two aspects, namely training sample selection and adaptive estimation. We present a general framework for adaptive estimation schemes and show that most approaches found in related literature can be conceptually unified. We perform several case studies to evaluate training sample selection and adaptive estimation for various system configurations.

The last chapter of this dissertation contains an overall summary and conclusions of this dissertation. In this chapter, we also highlight the contributions corresponding to each module of this work.

1.3 List of publications

The following list contains first author publications related to this dissertation. A full version of the articles can be found in Appendix B.

[P1] Eckhard, T., Valero, E. M., Hernández-Andrés, J., & Heikkinen, V. (2014). Evaluating logarithmic kernel for spectral reflectance estimation—effects on model parametrization, training set size, and number of sensor spectral channels. J. Opt. Soc. Am. A, 31(3), 541-549.[1]

[P2] Eckhard, T., Klammer, M., Valero, E. M., & Hernández-Andrés, J. (2014). Improved Spectral Density Measurement from Estimated Reflectance Data with Kernel Ridge Regression. Image and Signal Processing, Springer, 2014, 79-86.[2]

[P3] Eckhard, T., Valero, E. M., Hernández-Andrés, J., & Schnitzlein, M. (2014). Adaptive global training set selection for spectral estimation of printed inks using reflectance modeling. Applied Optics, 53(4), 709-719.[3]

---

[1] Journal of the Optical Society of America A impact factor 2013: 1.448

[2] Presented at the *16th Int. Symp. on Multispectral Colour Science* in June 2014, Cherbourg (France)

[3] Applied Optics impact factor 2013: 1.649

2

Background

Contents

2.1 Preface

In this chapter we introduce the concepts and information that form the theoretical foundation for later parts of this work. At first, a physical model of the reflection of light from a surface is presented and its measurement is discussed. These radiometric concepts are related to colorimetry and the measurement of color, which is considered afterwards. We then introduce spectral imaging, the digital camera working principle and the line-scan multi-spectral principle related to the imaging system considered in this work. At last, we define a framework for system performance evaluation and introduce the datasets and camera system configurations used in experimental parts of this dissertation.

Notation
Vectors are denoted lowercase bold letters and defined as $\mathbf{x} = [x_1, x_2, ...x_n]^T \in \mathbb{R}^{n\times 1}$ for an n-dimensional vector. Matrices are denoted uppercase letters, for instance $X = [\mathbf{x}_1, \mathbf{x}_2, ...\mathbf{x}_m] \in \mathbb{R}^{n\times m}$ is a matrix composed of m column vectors of dimension n. Further, let $\mathcal{X} = C(X)$ be the column space of matrix $X \in \mathbb{R}^{n\times m}$, which is the set of all possible linear combinations of its column vectors, ie. the linear span of the column vectors. Consequently, $\mathcal{X}$ is a subspace of $\mathbb{R}^m$. More conventions for notation, variable naming and graphical illustrations are listed in Appendix A.1.

2.2 Light reflection and color perception

2.2.1 Physical model of light reflection

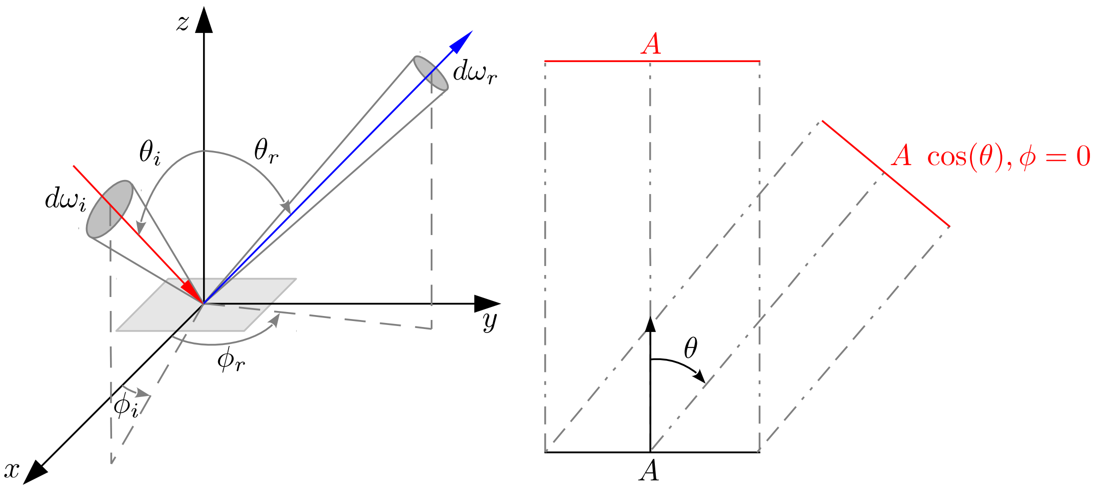
The optical part of electromagnetic radiation from a source covers the length of electromagnetic waves from 10 nm to 1000 μm, or corresponding frequencies of wave oscillation of 3×10^{11} Hertz (Hz) to 3×10^{16} Hz. This portion includes the ultraviolet, visible and infrared region. Frequency and wavelength of light propagating through a medium are related by the constant speed of light in vacuum c and the index of refraction of the medium n, such that $c = \lambda f n$. The frequency of light is denoted f and has the unit Hz, and λ denotes wavelength in unit nm. The index of refraction is the ratio of the speed of light in vacuum to the speed of light in the medium denoted v, ie. $n = \frac{c}{v}$. For air, $n = 1.0003$ and in general, n must be greater than or equal to 1 [2]. In the field of spectral science and colorimetry, electromagnetic radiation is usually described by wavelength and we therefore adapt to this convention.

In this work, we are specifically interested in the visible portion of the optical part of electromagnetic radiation. This portion is confined in a rather narrow spectral range from approximately $360-400$ nm to $760-830$ nm [3]. While there is no unique definition for this range, we consider 380 nm to 780 nm as the visible. This is the range of electromagnetic radiation in which the average human observer is sensitive.

Fundamental units that describe light are radiant energy Q with unit Joule (J) and radiant flux Φ, which is energy per unit time ($\frac{dQ}{dt}$) and is measured in Watts (W). The quantity $\Phi(\lambda, \theta, \phi)$ describes the directional spectral radiant flux at wavelength λ in direction (θ, ϕ). Variables θ and ϕ are polar and azimuth angle in a spherical coordinate system and defined with respect to the surface normal (see Figure 2.1a). A light source can be described radiometrically by the directional quantity spectral radiance, denoted as $L(\lambda, \theta, \phi)$ with units W/m^{2}sr, which is power per unit of projected area and unit of solid angle [4]:

$$L(\lambda, \theta, \phi) = \frac{d^3\Phi(\lambda, \theta, \phi)}{\cos(\theta, \phi)\ dA\ d\omega\ d\lambda}, \tag{2.1}$$

The term $\cos(\theta, \phi)$ in the denominator accounts for the apparent size change of the radiating source in direction (θ, ϕ). This effect is illustrated in Figure 2.1b. When considering the acquisition of digital images with a camera, we are interested in the radiant flux arriving at a certain pixel location of the camera sensor. Therefore, radiant flux and spectral radiance are also functions of the position in space. For this work, we remain neglecting the position vector from the equations for simplicity.



(a) **(b)**

Figure 2.1: Illustration of the spherical coordinate system used to describe directional radiometric quantities (a) and apparent size change of surface A as a function of the cosine of the viewing angle with respect to the surface normal (b). Illustrations adopted from [4, p.57] and [4, p.53] respectively.

For radiant flux that is incident on a surface, reflection, transmission and absorption can occur. Other effect of light-matter interactions are refraction and scattering. Figure 2.2 illustrates three light rays incident on an inhomogeneous material, for instance paper. The first reflection at the interface is called interface reflection (indicated in orange). If a ray is reflected on the rough surface material several times, multiple reflection occurs (indicated in green). This type of reflection is also called interface reflection. The refracted rays indicated in red, magenta and blue illustrate subsurface scattering by material particles, which can either lead to absorption of the ray (red ray) or re-emission of the ray (blue ray). The rays re-emitted from sub-surface scattering through the same interface from where they penetrated the material are termed body reflection. Rays penetrating through the material are called transmitted (magenta ray).

In this work, we are specifically interested in the reflection of light from a surface and therefore do not consider the part of light absorbed by, or transmitted through, the surface. There exist several models that describe this process geometrically in terms of radiometric quantities. We consider the *bidirectional spectral reflectance distribution function* (BSRDF), which is defined as

$$f_r(\theta_i, \phi_i, \theta_r, \phi_r, \lambda) = \frac{dL_r(\theta_i, \phi_i, \theta_r, \phi_r, \lambda)}{L_i(\theta_i, \phi_i, \lambda)\cos(\theta_i, \phi_i) d\omega_i}. \tag{2.2}$$

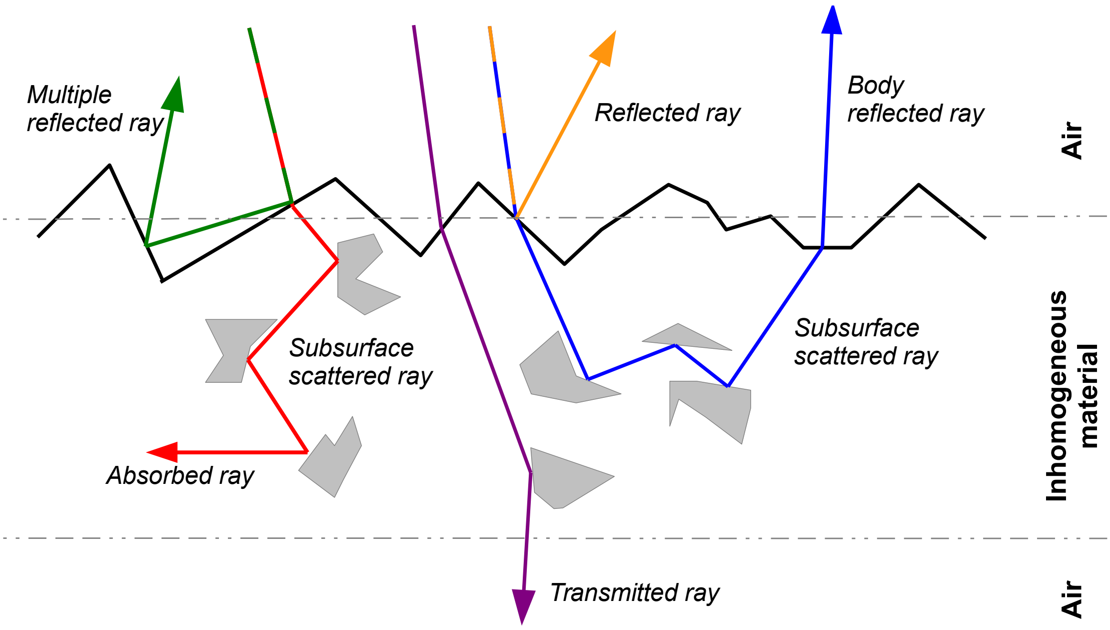


Figure 2.2: Different components of light matter interaction.

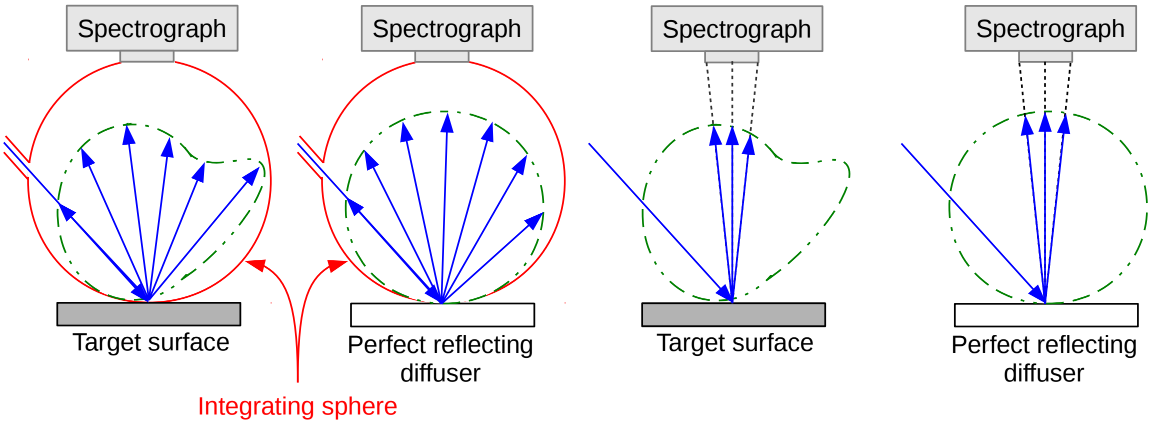
The numerator denotes the differential spectral radiance at wavelength λ reflected in direction (θ_r, ϕ_r). The denominator denotes the irradiance on the surface, which is the part of the incident spectral radiance at wavelength λ projected onto the surface in direction of travel (θ_i, ϕ_i) with differential solid angle $d\omega_i$.

A geometric interpretation of the distribution function is given in Figure 2.1a. Lee *et al.* [5, pp.148] discuss several limitations to this model, for instance that light polarization or fluorescence are not included. In this work neither of these effects are considered.

2.2.2 Spectral reflectance measurement

In the previous subsection, the BSRDF model of spectral light distribution as a function of geometrical variables was introduced. Once this function is obtained, we can predict the light reflected off a surface at any location on the hemisphere, given the distribution of incident light. Measuring the BSRDF of a surface requires measuring the reflected radiance at any combination of observation angle (θ_r, ϕ_r) and incident light angle (θ_i, ϕ_i) over the surface hemisphere. This is tedious work and is seldom done in practice. Instead, we can fix some parameters of the BSRDF to obtain a defined measurement for a light reflection measure that is much more simple to obtain.

One such measurement condition is to sum up the light reflected in all directions over the hemisphere, and the quantity we obtain by that is referred to as *directional-hemispherical reflectance* (or often only *reflectance*) $r_{dh}(\lambda)$, which is a function of the wavelength of incident light. The values of $r_{dh}(\lambda)$ are constrained to values between 0 and 1. Due to the conservation law of energy, no more light than all incident light at a given wavelength can be reflected (unless fluorescence is considered). We define the directional-hemispherical reflectance for a target surface as the integral of the reflected radiance $dL_{r,T}$ over the whole hemisphere leaving the surface (ie. 2π steradiant solid angle observation),



(a) Directional-hemispherical reflectance **(b)** Reflectance factor

Figure 2.3: Illustration of *directional-hemispherical reflectance* and *reflectance factor* measurement principle. Illustration adopted from [3].

divided by the incident flux per unit area:

$$r_{dh,T}(\lambda) = \frac{\int_{\omega_r} dL_{r,T}(\theta_i, \phi_i, \theta_r, \phi_r, \lambda)\cos(\theta_r, \phi_r)d\omega_r}{L_i(\theta_i, \phi_i, \lambda)\cos(\theta_i, \phi_i)d\omega_i}. \tag{2.3}$$

The integral quantity in the numerator can be obtained by measuring the exitant radiance over the hemisphere with an integrating sphere. Obtaining the incident radiant flux is usually achieved by indirect measurement. Let $r_{dh,W}(\lambda)$ be the directional-hemispherical reflectance of a white reference surface, such that

$$r_{dh,W}(\lambda) = \frac{\int_{\omega_r} dL_{r,W}(\theta_i, \phi_i, \theta_r, \phi_r, \lambda)\cos(\theta_r, \phi_r)d\omega_r}{L_i(\theta_i, \phi_i, \lambda)\cos(\theta_i, \phi_i)d\omega_i}. \tag{2.4}$$

Solving Equation 2.4 for $L_i(\theta_i, \phi_i, \lambda)\cos(\theta_i, \phi_i)d\omega_i$ and substitution in Equation 2.3 results in

$$r_{dh,T}(\lambda) = \frac{\int_{\omega_r} dL_{r,T}(\theta_i, \phi_i, \theta_r, \phi_r, \lambda)\cos(\theta_r, \phi_r)d\omega_r}{\int_{\omega_r} dL_{r,W}(\theta_i, \phi_i, \theta_r, \phi_r, \lambda)\cos(\theta_r, \phi_r)d\omega_r} r_{dh,W}(\lambda), \tag{2.5}$$

which now contains two integral quantities ($\int_{\omega_r} dL_{r,T}(\cdot)d\omega_r$ and $\int_{\omega_r} dL_{r,W}(\cdot)d\omega_r$) that can be obtained by measurement with an integrating sphere and a reference white surface with known directional-hemispherical reflectance.

The measurement of the directional-hemispherical reflectance does not consider the angularity of the light distribution. Another measurement condition is using a finite but limited solid angle of observation corresponding to the measurement cone of a detector with fixed aperture. This quantity is called *reflectance factor* [5, 6] and the relation of *directional-hemispherical reflectance* and *reflectance factor* is illustrated in Figure 2.3. From a technological point of view, devices measuring reflectance factor are more simple because they do not require an integrating sphere.

The angle of incident flux, solid angle as well as angle of observation of reflected radiant flux from a surface with respect to its normal are not explicitly defined for the term reflectance factor. For the same sample surface, its measurement under different geometrical settings will therefore lead to different results. The need for standardization is apparent and has motivated the work of several international standardization organizations. In this dissertation, we follow the geometrical settings for reflectance factor and colorimetric measurements defined in ISO 13655:2009(E)[1]. The norm defines four measurement conditions for reflectance factor:

M0 requires the spectral power distribution (SPD) of the illumination source to closely match CIE[2] illuminant A (see Figure 2.4a).

M1 requires the SPD of the illumination source to closely match CIE illuminant D50 (see Figure 2.4a). The UV-content for this illuminant is well defined and because of this, it can be ensured that target surfaces that exhibit fluorescent effects undergo a defined excitation of the latter.

M2 requires the SPD of the illumination source to be provided at least in the range from 420 nm to 700 nm, and additionally that there exists no substantial radiation power in the spectral range below 400 nm. This illumination condition approximates that of museums and is often referred to as UV-cut.

M3 requires the same SPD as M2 but with additional restrictions on light polarization to reduce for instance gloss effects.

The M1 norm is a widely used standard in the graphics art and printing domain. Considering a daylight-like illumination condition in color measurements of for instance printed products is a realistic setting approximating a likely observer viewing condition at the consumer side. The defined UV content of CIE illuminant D50 is important especially when printing on paper substrate is considered, because many paper substrates contain optical brightening agents. These are chemical compounds that absorb light in the ultraviolet region (typically 300-400 nm) that is re-emitted in the blue region (typically 400-500 nm) [7]. The visual effect on the substrate is the reduction of yellowish appearance because of the additional power in the blue part of the spectrum, resulting in an overall whiter appearance.

The M2 norm on the other hand requires low UV content of the illumination, resulting in little or no fluorescent effect. We will see later in this work that the light source used with our laboratory LSMSC system is tuned such that this condition is fulfilled, allowing us to adopt the M2 norm for spectral and color measurements.

Spectral reflectance functions can be measured with a *spectroradiometer*, using an external illumination source and defined observation and scene illumination geometry. A device called *spectrophotometer* typically includes an illumination source and implements a defined geometry. Both types are point measurement devices that incorporate an optical component to disperse light (eg. a grating or a prism), typically a photodetector array that responds to light in a quantitative manner producing an analog electrical signal, and electronics to convert the analog signal to a digital signal.

In this work, ground-truth measurements of reflectance factor data are mostly obtained from a Konica Minolta FD7 spectrodensitometer (which in fact is a spectrophotometer). For this device, measurement is ISO 13655 conform under condition M2.

We explicitly note that in later parts of this work, we will refer to *reflectance factor* only when it is important to distinguish between other kinds of reflectance functions and the reflectance factor function. Otherwise, we only refer to *reflectance* for the sake of simplicity.

---

[1] ISO 13655:2009(E): Graphic Technology - Spectral measurement and colorimetric computation for graphic arts images.

[2] CIE is the abbreviation for *Commission Internationale de l'éclairage*, which stands for the International Commission on Illumination.

2.2.3 Discrete representation of reflectance and other spectral data

Considering a surface reflectance as a function of wavelength, this function can be decomposed into its frequency components by Fourier transformation. If the function's Fourier transform is zero outside a certain range of frequencies, it is band limited [8]. According to Nyquist-Shannon sampling theorem, the sampling rate of a function should then be at least twice the maximum frequency component of the function to allow full reconstruction. In the field of color science, spectral signals are usually sampled in a 10 nm interval, which has been found to be sufficiently precise assuming the spectra are smooth [9]. Unless specified otherwise, we consider 10 nm sampling for spectral signals and adapt a discrete representation. The spectral measurements of the Konica Minolta FD7 are in accordance with this sampling and a measured spectral reflectance function is therefore represented as a vector $\mathbf{r} = [r_1, r_2, ...r_m]^T \in \mathbb{R}^{m\times 1}$, where $m = 36$ and the vector elements are corresponding to wavelengths from 380 nm to 730 nm. Other spectral signals are defined in a similar manner.

2.2.4 Spectral reflectance orthogonalization

The span of a set of surface reflectance vectors is a subspace in $\mathbb{R}^m$. Various researchers have statistically analyzed large sets of reflectance data and shown that this subspace is much smaller than $\mathbb{R}^m$ [8, 10, 11, 12].

Using statistical procedures based on orthogonal transformations, a more compact representation of the reflectance data can be obtained. By that, the correlated wavelength components of reflectances in $\mathbb{R}^m$ can be converted to a set of linearly uncorrelated variables. The orthonormal basis vectors of the orthogonalized spectral reflectance space are denoted $\mathbf{u}_i$, where $i = 1, 2, ...m$. Having these vectors ordered according to the variance of reflectances in $\mathbb{R}^m$ allows a ranking of the importance of the orthogonal vectors with respect to the amount of information each vector carries. The ordered orthogonal basis vectors stacked in columns of matrix $U = [\mathbf{u}_1, \mathbf{u}_2, ...\mathbf{u}_m] \in \mathbf{R}^{m\times m}$ define a transformation to the orthogonalized space

$$\mathbf{c} = U^T\mathbf{r}, \tag{2.6}$$

where $\mathbf{c} \in \mathbb{R}^{m\times 1}$ is a coefficient vector containing coordinates in the orthogonal space corresponding to reflectance vector $\mathbf{r} \in \mathbb{R}^{m\times 1}$. This process is referred to as the decomposition of $\mathbf{r}$ into a weighted sum of linearly uncorrelated basis vectors (where the weights are given by the coefficient vector $\mathbf{c}$).

As $UU^T = I$ Reflectance $\mathbf{r}$ can be reconstructed from U and the coefficient vector $\mathbf{c}$ by evaluating

$$\mathbf{r} = U\mathbf{c}. \tag{2.7}$$

The reconstruction of $\mathbf{r}$ is exact, up to rounding errors in numerical computation. Hence, representing $\mathbf{r}$ in terms of U and $\mathbf{c}$ is also not more compact.

However, because the reflectance subspace is smaller than $\mathbb{R}^m$, it can be approximated quite accurately with less than m basis vectors. The corresponding subspace is truncated and in practice, only the b most significant basis vectors are retained. Scalar b takes values from 1 to m. As basis vectors in columns of U are in decreasing order of significance, we define $\hat{U} = [\mathbf{u}_1, \mathbf{u}_2, ...\mathbf{u}_b]$. The corresponding transformation matrix to the truncated orthogonal space is

$$\hat{\mathbf{c}} = \hat{U}^T\mathbf{r}, \tag{2.8}$$

and reflectance $\mathbf{r}$ can be reconstructed from $\hat{U}$ and the coefficient vector $\hat{\mathbf{c}}$ by evaluating

$$\hat{\mathbf{r}} = \hat{U}\hat{\mathbf{c}}. \tag{2.9}$$

where $\hat{\mathbf{r}} - \mathbf{r}$ is a reconstruction error due to dimensionality reduction of the orthogonalized spectral reflectance space by truncation. For large values of b, this error is low, as the information corresponding to the least significant basis vectors is often mostly noise.

There are various ways to obtain orthogonal vectors that span the subspace in which a set of observations (ie. known reflectances) are contained. In linear algebra this process is called orthogonalization and the most popular approaches are based on principal component analysis (PCA) or singular value decomposition (SVD). Both methods are intimately related and their names are often used interchangeable [13]. It follows a brief introduction to the computation of SVD and PCA. Unless mentioned differently, we use SVD for spectral reflectance orthogonalization in this dissertation.

A Singular value decomposition (SVD)

Let $M \in \mathbb{R}^{m \times l}$ be a real matrix of rank k. The rows of M are components of the multidimensional data variables and the columns are observations such that $M = [\mathbf{m}_1, \mathbf{m}_2, ...\mathbf{m}_l]$, where $\mathbf{m}_i = [m_{1,i}, m_{2,i}, m_{m,i}]^T \in \mathbb{R}^{m \times 1}, i = 1, 2, ...l$. The SVD of M is a matrix factorization of form

$$M = U \Sigma V^T, \tag{2.10}$$

where $U \in \mathbb{R}^{m \times m}$ is the orthogonal matrix of left singular vectors $\{\mathbf{u}_i \in \mathbb{R}^{m \times 1} | i = 1, 2, ...m\}$ in columns, matrix $\Sigma \in \mathbb{R}^{m \times l}$ is the singular value matrix containing decreasingly ordered singular values $\{\sigma_i \in \mathbb{R} | i = 1, 2, ... \min(m, l)\}$ on the diagonal and zero elsewhere. All entries of Σ are nonnegative and exactly k of them are strictly positive. The orthogonal matrix $V \in \mathbb{R}^{l \times l}$ contains right singular vectors $\{\mathbf{v}_i \in \mathbb{R}^{l \times 1} | i = 1, 2, ...l\}$ in columns. [14, pp. 18-22]

B Principal component analysis (PCA)

SVD is performed directly on the data matrix M. PCA is performed on the mean-corrected covariance matrix of M, which is the $m \times m$ matrix $K = \mathrm{cov}(M - \overline{A})$. Matrix $\overline{A} = [\overline{\mathbf{a}}, \overline{\mathbf{a}}, ...\overline{\mathbf{a}}] \in \mathbb{R}^{m \times l}$ contains the mean values of each component over all observations, ie. $\overline{\mathbf{a}} = [\overline{a}_1, \overline{a}_2, ...\overline{a}_m]^T$ and $\overline{a}_j = \frac{1}{l} \sum_{i=1}^{l} m_{j,i}$, where $j = 1, 2, ...m$. Assuming observations in M are from a multivariate random variable, we can use the unbiased estimate of the covariance matrix described in Appendix A.4. Performing Eigenvector decomposition, we obtain

$$U^{-1} K U = D, \tag{2.11}$$

where $U = [\mathbf{u}_1, \mathbf{u}_2, ...\mathbf{u}_m] \in \mathbb{R}^{m \times m}$ is the matrix of m Eigenvectors $\{\mathbf{u}_j \in \mathbb{R}^{m \times 1} | j = 1, 2, ...m\}$ of K in columns. Matrix $D \in \mathbb{R}^{m \times m}$ contains the Eigenvalues $d_1, d_2, ...d_m$ of K along the diagonal and is zero elsewhere. Eigenvectors and Eigenvalues are paired and matrices U and D usually ordered according to decreasing Eigenvalues. The columns in U are the principal components of M. The diagonal elements in D are the variances corresponding to the principal components.

C Applications

Orthogonalization allows dimensionality reduction for transmission and storage of spectral data with little or no loss in information [15, 16]. Linear models of spectral data were also used in spectral estimation [17, 18] (we come back to this in Chapter 5). Apart from that, an orthogonalized spectral reflectance space is of use in multi-spectral imaging system design to quantify the information loss in the acquisition process (we come back to this in Chapter 3).

2.2.5 Physical model of color perception

We revise briefly the well-known link between radiometric quantities describing electromagnetic signals in the visible range of the spectrum of light and the physiologically perceived color in the human visual system. According to the tri-chromatic nature of photoreceptors in the human eye, the range of perceivable colors for an average human observers can be described by coordinates in a three dimensional color space. The spectral functions that define the mapping from spectral space to the color space are specific primaries associated with a color space and all colors can be expressed by mixtures of the three primaries. One particular important set of primaries is defined by the CIE and called CIE-XYZ. It is the space of chromatic responses of the so-called standard observer. These responses are referred to as *tristimulus values.* A numerical descriptor of tristimulus values (X, Y and Z) to a color signal resulting from a surface with spectral reflectance function $r(\lambda)$, illuminated by a source with SPD $l(\lambda)$ is computed with the CIE standard observer color matching functions ($\overline{x}(\lambda)$, $\overline{y}(\lambda)$ and $\overline{z}(\lambda)$) [19, p.156]:

$$\begin{aligned} X &= \int_{380}^{780} r(\lambda)l(\lambda)\overline{x}(\lambda)d\lambda \\ Y &= \int_{380}^{780} r(\lambda)l(\lambda)\overline{y}(\lambda)d\lambda \\ Z &= \int_{380}^{780} r(\lambda)l(\lambda)\overline{z}(\lambda)d\lambda. \end{aligned} \tag{2.12}$$

By using a particular normalization, the Y-tristimulus value can be related to Luminance of a color stimuli. We note that we do not include this normalization factor in the previous equation because we mainly consider ratios of tristimulus values and the normalization is hence without effect.

Considering discrete sampling, we can rewrite this expression in matrix notation and obtain

$$\mathbf{x}_{XYZ} = Y_{CMF}^{T}\ \operatorname{diag}(\mathbf{l})\ \mathbf{r}, \tag{2.13}$$

where $\mathbf{x}_{XYZ} = [x, y, z]^T \in \mathbb{R}^{3\times 1}$ is a vector of tristimulus values, $\operatorname{diag}(\mathbf{l}) \in \mathbb{R}^{m\times m}$ a matrix with values of the SPD $\mathbf{l} \in \mathbb{R}^{m\times 1}$ on the diagonal. Vector $\mathbf{r} \in \mathbb{R}^{m\times 1}$ denotes a spectral reflectance and matrix $Y_{CMF} \in \mathbb{R}^{m\times 3}$ the discrete CIE Standard Observer color matching functions (CMFs).

Lack of perceptual uniformity when considering geometric distances between pairs of colors at different locations in CIE XYZ space led to the development of the CIE 1976 L*a*b* color space (CIE-Lab) among several others. The non-linear transformation of a vector of tristimulus values $\mathbf{x}_{XYZ} \in \mathbb{R}^{3\times 1}$ from XYZ to vector $\mathbf{x}_{Lab} = [L^*, a^*, b^*]^T \in \mathbb{R}^{3\times 1}$ in CIE-Lab space is defined as[3]

$$\begin{aligned} L^* &= 116\ f\left(\frac{y}{y_n}\right) - 16 \\ a^* &= 500\ \left[f\left(\frac{x}{x_n}\right) - f\left(\frac{y}{y_n}\right)\right] \\ b^* &= 200\ \left[f\left(\frac{y}{y_n}\right) - f\left(\frac{z}{z_n}\right)\right]. \end{aligned} \tag{2.14}$$

---

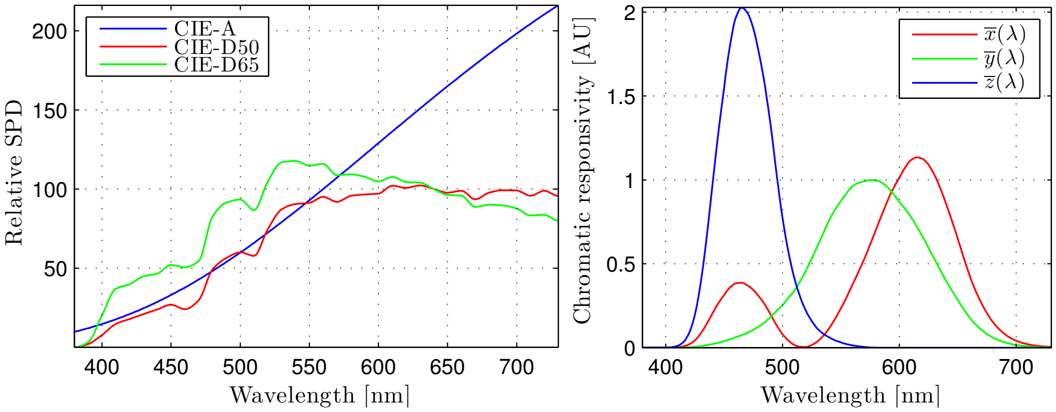
[3] Definition from ISO 13655:2009(E): Graphic Technology - Spectral measurement and colorimetric computation for graphic arts images.

The scalars x_n, y_n and z_n are the CIE XYZ tristimulus values of the reference white point and

$$f(t) = \begin{cases} t^{\frac{1}{3}} & \text{if } t > (\frac{6}{29})^3 \\ \frac{841}{108}\, t + \frac{4}{29} & \text{otherwise.} \end{cases} \tag{2.15}$$

The components of the color vector $\mathbf{x}_{Lab}$ are the lightness L^* of the corresponding reflectance assuming a given scene illumination and chromaticity coordinates a^* and b^*. Whenever CIE-Lab coordinates are computed in this work, we use the CIE 1964 10° Standard Observer as CMFs, which was derived from works of Styles and Burch [20] and Speranskaya [21]. By selecting the 10° Standard Observer, we consider a viewing condition assuming the objects to be viewed are close to the observer. Color coordinates are computed using CIE-D65 Standard Illuminant. Both are shown in Figure 2.4.

The above definitions of the CIE-Lab color space and corresponding color coordinates act as quantitative descriptors for color based on the CIE standard observer. Color scientists have derived transformations to compute *hue* and *chroma* from CIE-Lab color coordinates, which allow quantifying color in perceptual terms.



(a) **(b)**

Figure 2.4: Illustration of CIE Standard Illuminant A, D50 and D65 (a) and CIE 1964 10° Standard Observer (b).

2.2.6 Color measurement

The color stimuli of a surface for a given standard illuminant can be obtained by direct or indirect measurement. With direct measurement we refer to the measurement of tristimulus values by means of a photosensitive detector and color filters that implement a form of the CMFs. These kind of devices are usually called *Colorimeters* or *Tristimulus colorimeters*. A device is said to be colorimetric if it satisfies the so-called *Luther condition* or *Maxwell-Yves criterion*, requiring the product of spectral responsivity of the sensor and the spectral transmittance of the filters to be a linear combination of the CMFs.

An indirect approach to obtain tristimulus values for a color stimuli is to measure spectral reflectance of the surface for a known scene illumination and to calculate $\mathbf{x}_{XYZ}$ according to Equation 2.13. This approach is much more flexible than the direct measurement, as it allows to change the measurement condition (ie. standard observer and standard illuminant) a posteriori. Indirect color measurement can be considered as the state of the art in tristimulus value measurement. The spectral imaging based measurement considered in this work is one kind of indirect color measurement that is estimation based.

2.3 Spectral image acquisition

Subsequently, we are interested in using a digital camera system for spectral measurements, specifically spectral reflectance factor measurements. A spectral image can be represented in form of a 3-dimensional matrix $C \in \mathbb{R}^{r \times c \times n}$, where r and c denote the spatial image dimensions and n the spectral image dimension (ie. the number of discrete spectral image channels).

There exist various technological approaches for spectral imaging and we are interested in those acquiring information of the visible part of the spectrum of light. They are all based on image sensors with a 2-dimensional array of photosensitive cells and follow one of the following four basic principles (illustrated schematically in Figure 2.5):

Spatial scanning: The acquisition of multiple images is synchronized with a translation movement (spatial scanning) of either camera system or image scene with respect to the other. Each acquired image contains spatial information of the scene in one image dimension, and spectral information in the other image dimension. The spectral separation of electromagnetic radiation can for instance be achieved with an optical diffraction grating (eg. [22]).

Spectral scanning: Each of multiple, typically sequentially acquired images contains spatial scene information with distinct spectral content. The spectral separation can for instance be achieved using multiple optical filters (eg. [23, 24]), a liquid-cristal tunable filter (*LCTF*) (eg. [25]) or an acousto-optic tunable filter (eg. [26]).

Spatio-spectral scanning: There exists no separation of spectral and spatial content in a single acquired image. Typically, the intensity in one spatial dimension of the image is also a function of the wavelength of the electromagnetic radiation sensed, and not only of the spatial coordinate. For instance, spatio-spectral image filtering can be achieved with holographic Bragg-gratings (eg. [27, 28]). Multiple images with different spectral content are acquired sequentially and resampled in a post-processing step to achieve the final spectral image cube.

Snapshot (non-scanning): Snapshot imaging refers to techniques that do not require moving parts and acquire multiple spectral image channels typically at the same moment in time.

One type of snapshot system is a multi-sensor camera that divides the optical path using one or several beam-splitters. Distinct optical filters can be attached to each sensor to achieve the spectral filtering of the spatial images and the acquisition of a spectral image cube from multiple sensors can be performed time-synchronized (eg. [29]).

A conventional RGB camera can be considered as a 3-channel snapshot system. The spectral separation is achieved by a Bayer-pattern filter array in front of the sensor. It is also possible to achieve spectral separation without a filter array using specific image sensors that make use of the wavelength-dependent penetration depth of photons in photosensitive material [30]. The same physical principle can be used to create image sensors with responsivities that are tunable at runtime. This has been

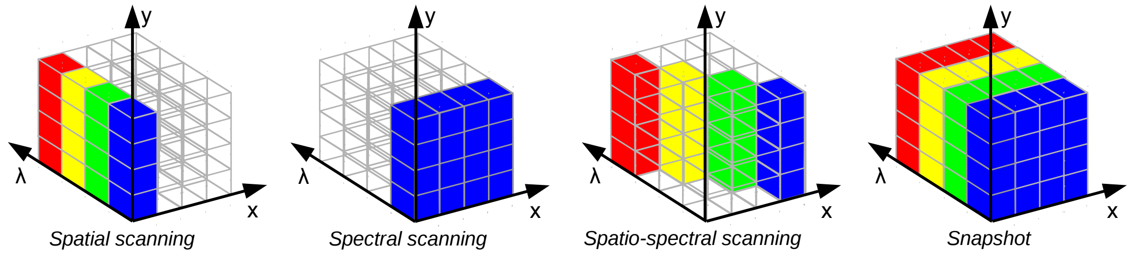


Figure 2.5: Schematic illustration of the four basic principles of spectral imaging, adopted from our previous work [28].

> shown for a sensor with few pixels [31], but not yet for a sensor with sufficient spatial resolution to perform spectral imaging. Because acquisition of multiple spectral images is performed sequentially, this approach could also be classified as spectral scanning.

Until now, we have not specified what we refer to exactly when stating that a spectral image cube contains spatial information of an image scene and spectral information corresponding to *discrete spectral image channels*. In fact, depending on the spectral imaging technique in focus, the meaning of a *spectral channel* can be quite different. Usually, spectral image channels are characterized by their spectral responsivity function, which describes how the channel responds to electromagnetic radiation as a function of wavelength. Without being any more specific, this can mean a spectrally narrow Gaussian shaped filter function as well as a spectrally wide filter function that has any other shape.

As stated earlier, we are interested in pixel-wise measurement of spectral reflectance factor functions from multi-spectral image data. Our requirement on the spectral sampling of these functions is a 10 nm step size. For instance, spectral responsivities of image channels of a well characterized LCTF based spectral imaging system can fulfill this requirement, because the filter functions are spectrally narrow Gaussian-shaped and can be tuned to sample the spectral curve in a 10 nm interval.

In this work, we are interested in an alternative approach to obtain the spectral reflectance function that is distinct from the aforementioned sampling scheme. We want to use camera response data from a multi-channel imaging system in combination with mathematical models to either recover the spectral reflectance factor function based on a model of the image acquisition process, or to estimate it using machine learning techniques. We will come back to this problem in Chapter 5.

In the following subsection, we will illustrate the working principle of digital cameras in general and then introduce the LSMSC system that is considered in this work. After that, we will formally define a model of the general image transfer function, describing the relation of scene radiance sensed by the camera and the corresponding camera response in digital counts.

2.3.1 Digital camera working principle

Digital cameras, with their basic functionality to sense light, belong to the class of radiometric measurement devices. The general working principle and corresponding information processing chain of a digital camera are illustrated in Figure 2.6. Accordingly, a scene is

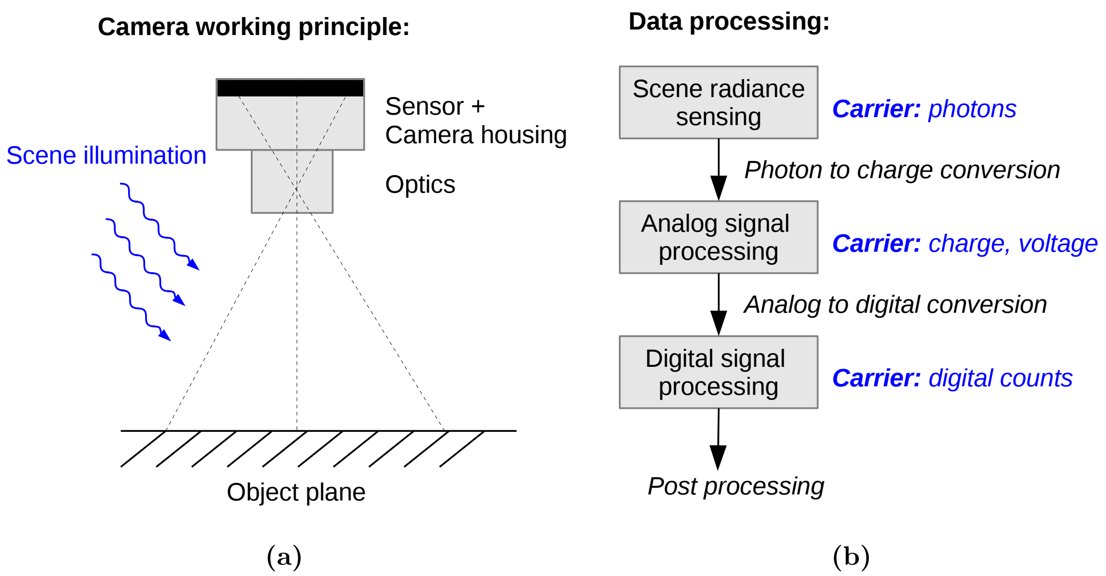


Figure 2.6: Illustration of the digital camera working principle (a) and information processing chain (b).

illuminated by a light source and a portion of the light irradiating the scene is reflected from the surface. A certain amount of this (reflected) light falls within a solid angle of observation with respect to the camera and the sensor responds to it. To get a sharp image of the scene at the sensor plane, the distances between image scene, camera optics and sensor plane have to be adjusted according to the focal length of the camera lens or vice versa. We refer to the distance between camera sensor and image scene as *working distance* of a camera system.

Most conventional cameras contain an area sensor with a rectangular array of photosensitive cells. The cells receive the reflected spectral radiance of the scene and by that spatially sample the scene in a single exposure. The magnitude of the response of the sensor to light as a function of wavelength is described by the sensor responsivity[4].

The imaging principle considered in this dissertation is based on a line sensor. Such a sensor has pixels arranged in a line and therefore only samples one line of the scene in a single exposure. A spatial image is obtained by translating either camera or scene with respect to the other such that the movement is perpendicular to the line sensor. The acquisition is performed synchronous to the movement and multiple image lines are concatenated to form the final image[5].

A CCD sensor technology

The most common types of camera sensor architectures are CCD and CMOS. CCD is the abbreviation for *charge coupled device* and refers to the type of movement of electrical charges, as we will see below. CMOS is the abbreviation for *complementary metal-oxide-semiconductor*, which is a technology for constructing integrated circuits. Even though

---

[4] In related literature this function is also called *sensitivity*. We use the term *responsivity* to avoid confusion with another common meaning for sensitivity, which is the minimum stimulus level required for a detector to respond.

[5] Note that this principle is described in more detail in Section 2.3.2 and also illustrated in Figure 2.9a.

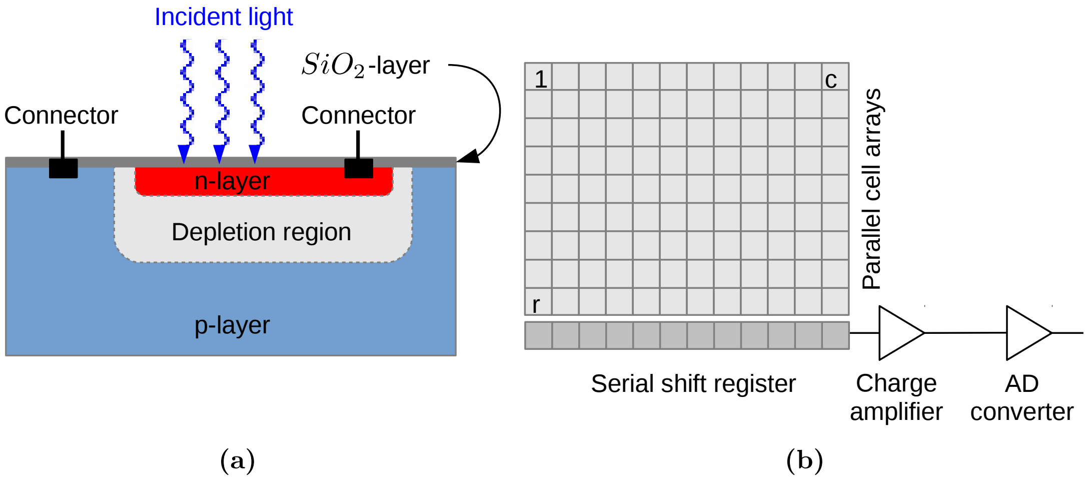


Figure 2.7: Schematic illustration of a CCD sensor cell (a) and an imaging array (b).

both technologies are based on metal-oxide semiconductor circuits, there exists a great difference in the sense that a CMOS sensor converts electronic charge to voltage at a pixel level, while charge-voltage conversion in a CCD is typically performed after shifting the charges to the sides of the sensor. A detailed comparison of the characteristics of these types of sensors can be found elsewhere, for instance in [32, 33]. In what follows, we concentrate on CCD technology, which is the sensor type used in the laboratory LSMSC system considered in this work.

The conversion from scene radiance to digital counts in a CCD sensor consists of 4 steps: charge generation, charge collection, charge transfer and charge measurement. The following basic introduction to these processes is based on [34].

The sensor cell contains a positive doped semiconductor substrate (p-layer), usually silicon, and above a thin negative doped silicon layer (n-layer) and a layer of isolating silicon-dioxide (SiO_2-layer). The SiO_2-layer is transparent for light photons. On top of this compound follow optically transparent electronic conductors, usually made of polycrystalline silicon. Considering the sensor cell in the sensor array compound, conductors form a regular structure on the top layer of the sensor. The layer composition is illustrated in Figure 2.7a.

The thin doped n-layer is often referred to as the buried-channel. Without applying voltage to the sensor cell, free electrons from the n-layer recombine with holes from the p-layer and a small depletion region is created at the junction of the p- and n-layer. By applying a reverse bias (ie. connecting the p-layer to the negative source pole and the n-layer to the positive source pole), the depletion region is increased and remains virtually free of charge. A small current that crosses the pn-junction regardless of the reverse bias is called dark current and depends on the ambient temperature.

If the sensor cell is exposed to light, photons that impinge on the sensor surface partially penetrate into the silicon. If the photon energy is larger than the band-gap of the sensor material, the interaction causes electron-hole pair generation based on the photoelectric effect. The charge generated in this process is linearly related to the light intensity impinging on the cell and to the time in which the cell is exposed to light (ie. the integration time).

The maximum amount of charge that can be generated is an important sensor property called full-well capacity. This parameter describes the amount of electrons a sensor cell can hold (usually between 50 000 and 1 000 000). A larger full-well capacity means a larger dynamic range (ie. the ratio between the most and the least bright object in a scene that can be simultaneously discriminated).

The efficiency of electron-hole pair generation in the silicon semiconductor, and consequently the efficiency of the sensor in light sensing, is depending on the energy they carry, ie. their wavelength. The energy of a photon is

$$E_\lambda = \frac{hc_0}{\lambda}, \tag{2.16}$$

where h is Plank's constant, c_0 the speed of light in vacuum and λ the photon's wavelength. This quantity is usually denoted in unit Electronvolt (eV) and $1\,\mathrm{eV} = 1.602 \times 10^{-19}\,\mathrm{J}$.

Assuming $c_0 \approx 2.99 \times 10^8\,\frac{\mathrm{m}}{\mathrm{s}}$ and $h \approx 6.626 \times 10^{-34}\,\mathrm{Js}$, we see that $E_{380\,\mathrm{nm}} = 3.25\,\mathrm{eV}$ and $E_{780\,\mathrm{nm}} = 1.59\,\mathrm{eV}$. The band-gap of a mono-crystalline silicon lattice is $E_{Si} = 1,1\,\mathrm{eV}$, and because $E_{380\,\mathrm{nm}} > E_{780\,\mathrm{nm}} > E_{Si}$, we can conclude that this material is sensitive in the spectral range from 380 nm to 780 nm considered in this work. Obviously, the sensor is also sensitive in the near-infrared, but the responsivity is bounded by the silicon band-gap towards higher wavelengths.

Another limitation of responsivity is related to the photon penetration depth in the semiconductor material, which is larger for more energetic photons (ie. shorter wavelength photons)[6]. The relation of photon energy and penetration depth can be approximated by Lambert-Beer law. Accordingly, the further away from the *pn*-junction a photon penetrates, the less likely electron hole pair generation occurs and the more likely is the recombination of free charges after electron-hole pair generation. The only region where recombination can not occur is the charge-free depletion region. The sensor responsivity is therefore the largest in the depletion region and decreases for wavelengths beyond this range.

Until now, charge generation and collection were discussed on a sensor cell level. Charge transfer is the process of moving accumulated charge packages from a single sensor cell to the amplifier that measures the charge. This is achieved by changing electromagnetic fields induced by clocked gate voltage, thermal diffusion and self-induced drift.

We explain the read-out process based on a schematic illustration of a sensor with $r \times c$ sensor cells (pixels) in Figure 2.7b. The transfer of charge packages is performed by shifting them sequentially in vertical direction through neighboring cells, synchronous for each of the c parallel sensor cell arrays. The efficiency of the charge transfer process (measured in percentage of preserved charge after shifting from one cell to another) is very high, usually higher than 99.999% [34]. A serial shift register collects charge packages shifted out of the sensor cells. Its charge transfer direction is orthogonal to the cell array charge transfer direction and an output amplifier is located at the end of this register. The shift frequency of the serial shift register has to be at least c times higher than the transfer rate through sensor cells, or several output amplifiers have to be used. Often, odd and even cell arrays are processed with separate output amplifiers to increase the transfer rate.

The output amplifier electronic usually consists of a small capacitor that receives the charge and a MOSFET amplifier. This amplifier creates a voltage that is proportional to the transferred cell charge. The voltage can then be converted to digital counts by analog-to-digital conversion.

---

[6] We note that this principle was already mentioned earlier in Section 2.3. To be more specific at this point, the aforementioned sensor detects photons at various depths in the photosensitive material.

B Sources of image noise

In a consumer camera, post-processing of the raw camera response is performed to make the acquired images appear visually pleasant to the observer. Post-processing is often non-linear with respect to the image intensities. Opposed to that, in a scientific camera, the linear relation of scene radiance and raw camera response is highly desired and the degree of linearity is usually considered as a quality factor. For this type of camera, effects resulting in a deviation from linearity are undesired and usually referred to as noise. Mechanisms that contribute noise to the measured camera response are photon arrival statistics, noise from the CCD array (transfer noise, dark current, fixed pattern noise and high energy radiation), amplifier noise (thermal noise, shot noise, flicker noise), analog-to-digital converter noise, electrical interference and the noise generated in the signal processing chain [5].

For this work, we consider noise types that are most relevant for the laboratory LSMSC system considered in this dissertation, which are *dark current noise*, *shot noise* and *quantization noise*. We describe them briefly as follows [34, 5]:

Dark current noise is a CCD array type noise resulting from thermal agitation of electrons accumulating in the potential well during the exposure time. This current source is termed *dark current*. It is almost proportional to the exposure time and it is a sensitive function of the sensor temperature. Dark current noise can be reduced drastically by cooling the CCD sensor. In most imaging applications, dark noise correction is performed by subtracting an average dark frame from the image data.

Shot noise refers to the random photon arrival statistics at the sensor. The arrival follows Poisson distribution and the variance is equal to the mean number of photons reaching the sensor per unit time. For low signals (ie. few photons arriving at the sensor), the signal fluctuation is correspondingly higher. Shot noise can not be accounted for by system design other than increasing the amount of photons reaching the sensor.

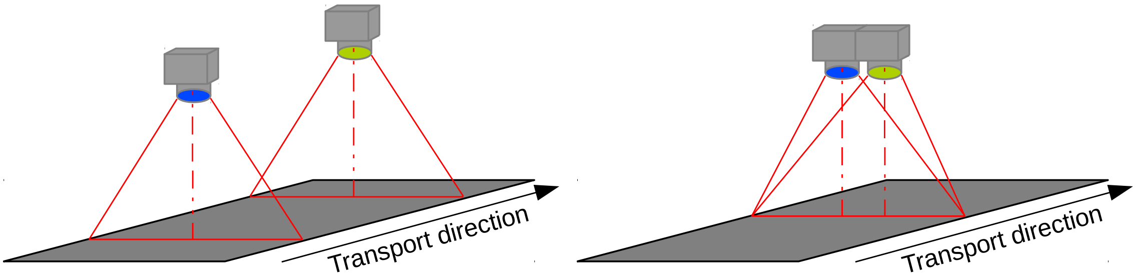
Quantization noise is a type of noise from the analog-to-digital converter and effectively a rounding error between the analog input voltage and the digitized output value. This type of noise is non-linear and signal dependent.

2.3.2 Line-scan multi-spectral camera system

The multi-spectral camera system in focus of this work is a line-scan system. Conventional area cameras acquire a 2D image in one exposure, while a line-scan camera acquires a 1D image line per exposure. As mentioned earlier, successive acquisition of such image lines with respect to a movement perpendicular to the sensor line samples the other dimension spatially to obtain a 2D image. This so-called *scanning principle* has many applications in consumer electronics, manufacturing industry and science. Examples of devices or applications using line-scan cameras are office scanners, quality control machine vision systems in assembly lines and air-borne or satellite imaging.

The length of a line-scan image in scanning direction is only limited by the available memory or computational constraints. Further, line sensors typically have larger photosensitive cells than area sensors and therefore allow for shorter exposure times. This is an advantage over area sensors in applications when fast continuous moving objects are to be measured or inspected.

There exist various possible mechanical configurations for line-scan multi-spectral imaging. We illustrate two cases schematically in Figure 2.8. System A consists of multiple camera modules that are sequentially arranged along the direction of movement of the scanning object (ie. transport direction). System B consists of multiple camera modules that are arranged in parallel at one position along the transport direction. The optical components of all cameras are designed such that each one acquires an image of the entire scan line. Further, camera modules can either contain multichannel line-scan cameras, or monochrome line-scan cameras with additional color filter. In either case, light sources are required to illuminate each scan line uniformly. Obviously, system A requires multiple light sources, whereas System B requires only one.



(a) System A **(b)** System B

Figure 2.8: Schematic illustration of line-scan multi-spectral imaging principles for the case of two camera modules.

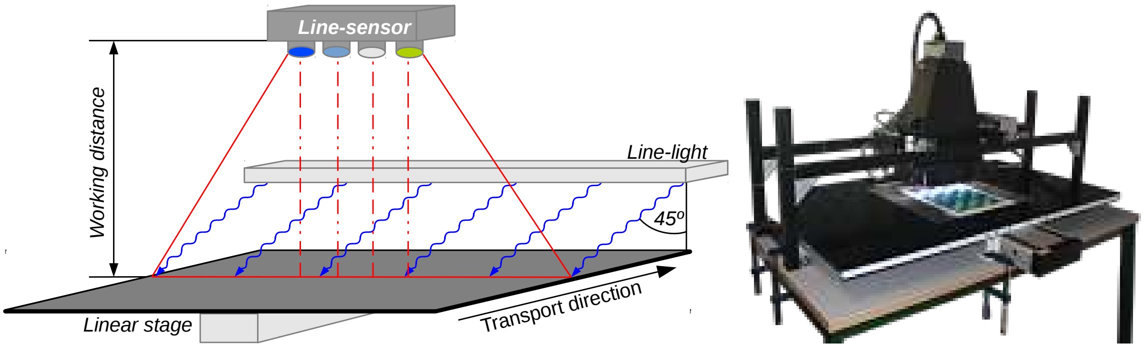
The imaging principle that is considered in this dissertation is similar to System B. Specifically, we focus on a particular system design recently invented by Chromasens GmbH: instead of using multiple camera modules, a single RGB line-scan camera module is used. Multiple lenses with additional color filters are attached to the line sensor such that corresponding image regions of the entire scan line are projected on individual parts of the line sensor. From now on, we refer to this as LSMSC system.

A System components and scene geometry

The acquisition set-up of the LSMSC is depicted schematically in Figure 2.9a. It consists of a motorized scanning table (linear stage), a line-scan camera at an angle of 0 ° with respect to the scan surface normal and a line light source at an angle of 45 °. On the right side in the same figure, a photo of the corresponding laboratory set-up is shown.

The linear stage is a motorized spindel drive actuator by Isel AG, which is used to translate the scanning object under the camera. In our setup, the stage is controlled via native movement commands sent from a serial communication controller implemented in the numerical computing environment Matlab by MathWorks, Inc., or a software user interface provided by Chromasens GmbH.

The line-scan camera module in our setup is a prototype by Chromasens GmbH called truePIXA. It consists of a 7k-pixel RGB CCD line sensor which allows line-scanning rates of up to 22.1 kHz. Attached to the camera housing are 4 lenses that project identical spatial image scene content onto 4 distinct parts along the sensor line as depicted in Figure 2.10a. Each lens has a distinct color glass filter attached. The 4 RGB sub-images therefore contain different spectral information of the image scene. A 12-channel image is formed by splitting the raw 3-channel RGB image into 4 parts and consecutive alignment of corresponding



(a) (b)

Figure 2.9: Non-proportional schematic illustration of the line-scan principle (a) and photo of the corresponding laboratory set-up (b).

spatial areas in each sub-image. The final scan resolution of our laboratory LSMSC is 100 dots per inch (dpi). Figure 2.10b illustrates the combined effect of camera spectral responsivity, spectral filter transmittances and scene illumination spectrum (which we call from now on *effective responsivity* of the camera system). The responsivity measurements were obtained using a Bentham M300 monochromator[7] and filter transmittances measured with a Konica Minolta CS2000 spectroradiometer[8].

Data acquired with the truePIXA camera is transferred via a *camera link* interface to a PC frame-grabber. Camera link is a serial communication protocol standard which allows a raw data transfer speed of up to 680 MB/s.

Image lines are acquired synchronously with the constant movement of the linear stage and concatenated to a 2D image. Constant movement is not required if a position encoder connected to the linear stage is used to trigger the line acquisition at constant steps of movement.

The object scene (ie. the scanning line of the camera) is illuminated by a line light source of type Corona 2e, developed by Chromansens GmbH[9]. This light source is composed of 4 different types of LEDs which are spatially alternated on a electronic circuit board and which can be controlled individually. The LED currents are operated to approximate CIE-D50 (see Figure 2.4a), but are tunable to other configurations as well. We illustrate the SPDs of individual LED types at two current levels in Figure 2.11. High illuminance of the source of up to 3 000 000 lux is achieved by high-power LEDs and a focusing back reflector design. This allows operating the camera with low gain and short exposure times while maintaining high scanning rates.

B Sensor architecture

The sensor in the truePIXA camera is a CCD RGB line sensor[10]. The general electronic working principle of the CCD sensor was described before but the sensor architecture of

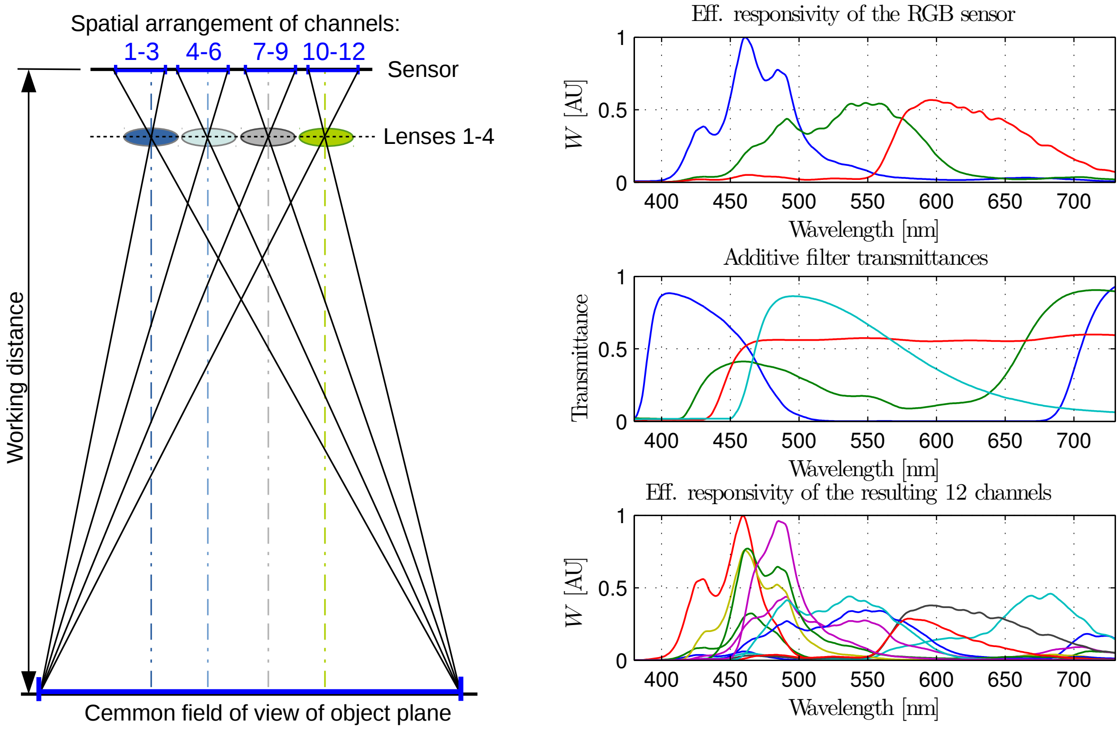
---

[7] Reading from the web: `http://www.bentham.co.uk/m300.htm`; retrieved: 23.03.2015

[8] Reading from the web: `http://sensing.konicaminolta.asia/products/cs-2000-spectroradiometer/`; retrieved: 23.03.2015

[9] Reading from the web: `http://www.chromasens.de/sites/default/files/chromasens-corona2-en-2014-f.pdf`; retrieved: 23.03.2015

[10] Note that most of the content of this subsection is summarized from allPIXA camera manual CD40067 Version 3.4.



(a) (b)

Figure 2.10: Optical principle of the truePIXA camera: generation of 12 channels with RGB line sensor and 4 serial lenses with additional color filters (a). The corresponding effective responsivity of the camera and filter spectral transmittances are shown in (b).

line sensors differ from area sensors markedly. Most obvious, sensor cells are arranged in serial fashion rather than the rectangular array architecture of the area CCD illustrated in Figure 2.7b. Except for that, the sensor implements full RGB at a pixel level by using parallel lines of R, G and B sensor cells. This is distinct from conventional color CCD sensors in which color filter arrays (eg. a Bayer filter mosaic) and demosaicing are used, which results in a lower true image resolution.

The parallel R, G, and B sensor lines are separated by lines of cells which are masked off to be protected from incident light. These cells act as shift registers to transfer the generated charge in the photosensitive cells to charge amplifiers at the end of the sensor line. This kind out read-out process is referred to as *interline transfer* and allows using an electronic shutter. As compared with charge transfer through the photosensitive cells and mechanical shutter, interline transfer features much shorter exposure times [5, p.496].

To speed up data processing, the read-out path of each sensor line of the truePIXA is split half in so-called rear and front tabs, and odd and even pixels use separate shift registers. The masked cells neighboring the photosensitive cells on both sides are used for that. Accordingly, 12 charge amplifiers are required to process the charges. We illustrate the line sensor architecture schematically in Figure 2.12.

C Data processing

We focus our discussion on the process performed after analog-to-digital conversion, i.e. once the acquired image is transferred to a PC. Raw image data is available in 8-bit and 16-

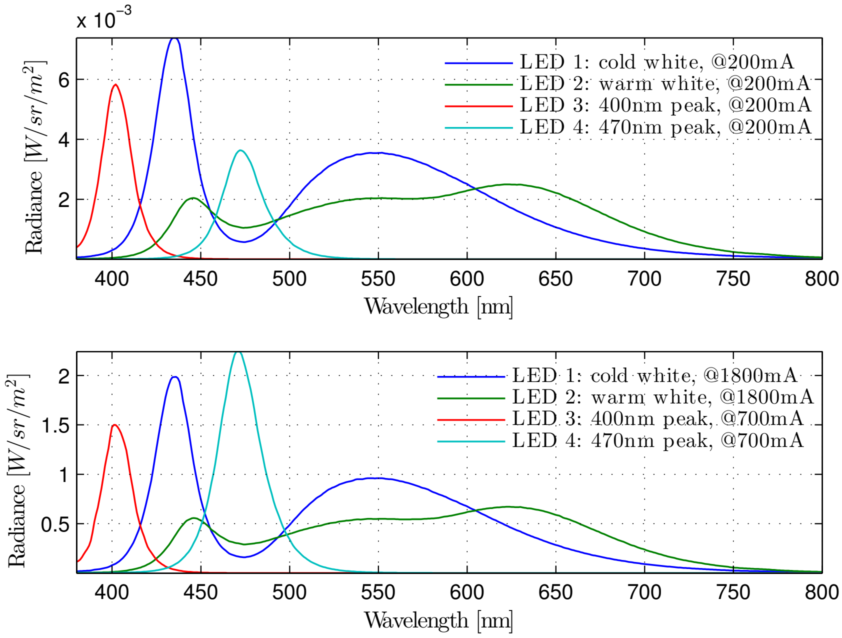


Figure 2.11: Spectral power distribution of the 4 LED types of the Corona 2e line light. The figure contains spectroradiometric measurements corresponding to 45 ° illumination- and 0 ° observation angle with respect to the measurement surface normal and 0.2 ° aperture at 52 cm distance. The radiances were measured over a Sphere Optics Zenith Polymer calibrated diffuse reflectance target, using a Photo Research PR75 spectroradiometer. The reflectance target was discounted from the measurement.

bit precision. Preliminary experiments by us have shown that there is no clear advantage of using 16-bit over 8-bit precision when the entire data processing pipeline for spectral reflectance measurement is taken into consideration, hence we consider raw data in 8-bit range. The main steps in sequence of processing are dark image subtraction, white-level correction, signal averaging and geometrical calibration.

Dark image subtraction is also called *dark frame subtraction* and its purpose is to account for dark current noise. As mentioned before, this type of noise is signal independent and is due to random generation of electrons and holes in the depletion region of photosensitive cells in a sensor. By acquiring an image in dark environment, this noise pattern can be quantified and subtracted later from any scene image. As dark current depends on the camera gain, exposure time and sensor temperature, the dark image has to be captured with the same settings as the scene image, and renewed in case ambient temperature changes severely.

White-level correction is also called *shading correction* or *flat-field correction* and mainly accounts for two effects. The first effect is pixel response non-uniformity (PRNU), an effect related to the image sensor due to which identical radiances reaching two sensor cells at different locations can produce distinctive amount of charges. PRNU can be caused by non-uniformly doped semiconductor material, variation of absolute pixel area in the manufacturing process, pixel-to-pixel variation of dark current or non-uniformity of sensor substrate thickness [5, p.498]. The PRNU can be corrected for by camera response scaling.

The second effect is not related to the sensor architecture, but to non-uniformity of

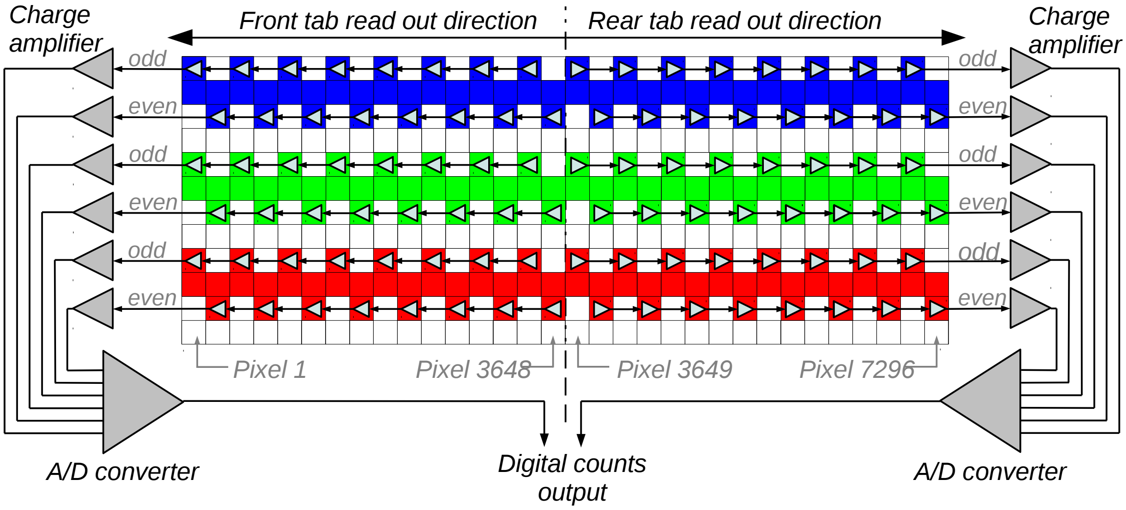


Figure 2.12: Schematic illustration of the line sensor architecture.

the scene illumination. If we assume this non-uniformity to be independent of the wavelength of light, it can also be corrected for by camera response scaling.

The camera response scaling can be achieved by applying a pixel-wise scaling factor to every line of the scanned image. Our laboratory setup includes a uniform white reference surface in form of a stripe that extends over the whole length of the scan line and is attached to the scanning table. Hence, every scanned image contains an area corresponding to the white reference surface. This area is extracted and averaged in scanning direction. Pixel-wise scaling factors correspond to reciprocal values of the averaged white response. After the white-level correction, the image intensity range is from 0 to 1 and precision is limited by the data type used (in our case floating numbers with 64-bit precision).

Oversampling/signal averaging is a process used to increase the image signal-to-noise ratio (SNR). Assuming signal and noise are uncorrelated, the signal level is constant over multiple acquisitions and that noise is a random variable with zero mean and constant variance, the SNR increases with the square root of the number of measurements that are averaged [35, p.60].

The assumption of constant signal level in multiple acquisitions generally does not hold for line-scan systems due to the continuous scanning movement of the sample with respect to the camera. However, if the exposure time is relatively short compared with the scan velocity, constant signal level can be assumed.

As an example we consider our laboratory LSMSC: let $v = 0.075\,\mathrm{m/s}$ be the scan velocity and $t = 50\,\mu\mathrm{s}$ be the exposure time. The distance that the measurement surface travels under the camera in one exposure is $s = vt = 3.75\,\mu\mathrm{m}$. The spatial resolution r of the system in scan direction is 100 dpi, or $r = 254\,\mu\mathrm{m/pixel}$, according to which $\lfloor \frac{r}{s} \rfloor = 67$ measurements can be averaged.

Geometrical calibration is the process of converting the raw 3-channel RGB image data to a 12-channel multi-spectral image cube. As mentioned before, a raw sensor RGB image contains 4 filtered sub-images. Each RGB sub-image contains an image of the scanned surface, projected over one of the lenses attached to the camera housing. The sub-images have to be registered spatially and corrected for several misalignment effects:

- a line shift (perpendicular to the scan direction) due to the displacement of the R, G and B sensor lines. For our LSMSC system, the resulting displacement is 4 pixels between consecutive channels.
- a fixed displacement of RGB sub-images from each lens due to the sequential arrangement of the lenses along the sensor line.
- observation angle deviation in perpendicular direction to the sensor due to precision limitations in the assembly of the 4 lenses in front of the camera. See Figure 2.13b for an exaggerated illustration of this effect.
- a non-linear image distortion along the sensor line which is distinct for each of the 12 channels due to lens aberrations and off-optical-axis magnification change (barrel distortion).
- a scene-adaptive pixel displacement if non-planar scanning objects are considered.

The 1st and 2nd item are global effects and can be corrected by a fixed translation. The last three items are more complicated to correct. We devote Chapter 3 to an extensive discussion about image registration of multi-channel line sensor data and present a registration framework for planar and non-planar scanning objects.

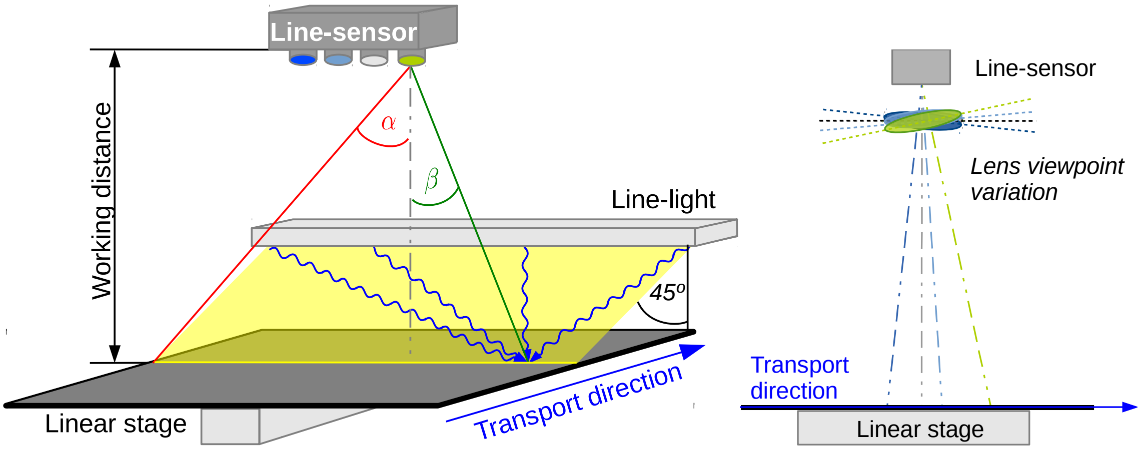
D Limitations of the LSMSC system with respect to color measurement

The specific interest in measuring reflectance factor with the LSMSC system motivates further discussion. We introduced in Section 2.2.2 a set of 4 different measurement conditions that conform with the norm ISO 13655 for reflectance factor and color measurement. The LSMSC imaging principle imposes some limitations in the realization of these conditions, which we will discuss next:

Illumination SPD: As mentioned before, we restrict our work to measurements conform with the M2 condition. Hence, it is crucial to ensure that the light source must not exhibit substantial power in the spectral range below 400nm. This is achieved by switching the current of the UV type LED off (LED 3 in Figure 2.11).

Measurement geometry: The ISO 13655 and related norms define geometrical requirements for color and spectral measurement. With the detector centered at zero degree with respect to the measurement surface normal, the light source has to illuminate the sample at 45° circumferentially or annularly. The LSMSC system uses the Corona 2e line light which features directional illumination from multiple angles. Accordingly, radiance at a certain location on the scan line is an integrated quantity over each angle of illumination as seen from that location. We indicate this effect in Figure 2.13a by depicting several rays of light in blue color for a location on the scan line corresponding to observation angle β and the most right lens.

Another limitation is a deviation of observation angle from 0° that depends on the raw sensor pixel location. If we consider the projected sub-image corresponding to one lens, a maximum observation angle deviation for distinct image points can be up to $\alpha \approx 17^\circ$ (see Figure 2.13a). The final 12-channel image is composed by concatenating 4 sub images and hence, image intensities in the final image that result from different sub-images correspond to distinct observation angles.



(a) perspective view **(b)** side view

Figure 2.13: Limitations for spectral and color measurements with respect to illumination and observation geometry. We illustrate in (a) the deviations of the observation angle from 0 ° (in red and green) and the deviations of the illumination angle from 45 ° (in blue) and lens viewpoint variations in (b).

2.3.3 General image transfer function

A The camera response model

We consider the products of spectral responsivity of the 3-channel RGB line sensor and each of the four filter transmittances of the lenses attached to the camera housing as 12 individual camera channel responsivities. The acquisition system captures the physical stimuli (ie. the spectrum of light incident on each camera sensor cell) and creates responses corresponding to the spectral responsivities of the 12 channels. We model the response of one line of a n-channel[11] line-scan camera system as

$$x_i(v) = \gamma_i \int_{\Lambda} s_i(\lambda, v) y_i(\lambda, v) d\lambda + \eta_i(v), \;\; i = 1, 2, ...n, \tag{2.17}$$

where $x_i(v)$ denotes the camera response of channel i at sensor pixel location $v \in \{1, 2, ..., c\}$, and c is the image width[12]. Further, $\Lambda = [380, 390, ...730]$nm is the wavelength range of light considered in this study and $y_i(\lambda, v)$ is the i-th function of camera spectral responsivity at wavelength λ[13]. Function $s_i(\lambda, v)$ describes the scene radiance received at sensor pixel location v (which is also referred to as the color signal) and the scalar $\eta_i(v)$ is a camera response noise term. The scaling factor γ_i is related to camera channel gain and exposure time.

The functional dependence of s_i, y_i, η_i and consequently x_i on the sensor location v can be modeled as a non-linear function Γ_i. Further, when considering discrete sampling of continuous functions, we can rewrite Equation 2.17 and obtain

$$x_i = \Gamma_i \left(\gamma_i \sum_{\lambda=1}^{m} s_i(\lambda) y_i(\lambda) + \eta_i \right), \;\; i = 1, 2, ...n. \tag{2.18}$$

---

[11] Accordingly, for our LSMSC, $n = 12$.

[12] The location referred to is that of the final 12-channel image and not the raw 3-channel image

[13] The spectral responsivity y_i is modeled as a function of v to account for the color glass filter transmittance variation with observation angle for each sensor pixel location.

Assuming Γ_i is the identity function and adapting vector notation, we obtain

$$\mathbf{x} = Y^T \mathbf{s} + \boldsymbol{\eta}, \tag{2.19}$$

where $\mathbf{x} = [x_1, x_2, ...x_n]^T \in \mathbb{R}^{n\times1}$ is a vector of camera responses and vector $\mathbf{s} = [s_1, s_2, ...s_m]^T \in \mathbb{R}^{m\times1}$ is the color signal, ie. radiance reflected from the object, illuminated by the light source. Further, $Y \in \mathbb{R}^{m\times n}$ is a matrix with n-channel responsivities in columns, each scaled by γ_i. The vector $\boldsymbol{\eta} \in \mathbb{R}^{n\times1}$ denotes camera response noise.

Because we assume Γ_i to be the identity function, we consider every pixel as a detector at 0 ° with respect to the surface normal, capturing the radiance of the sample illuminated at 45 °. This means that we ignore the effect of channel-dependent deviation of the observation angle and scene illumination.

If the sample surface is characterized by reflectance factor $\mathbf{r} = [r_1, r_2, ...r_m]^T \in \mathbb{R}^{m\times1}$, we can express the color signal as the product of reflectance factor and illumination SPD $\mathbf{l} = [l_1, l_2, ...l_m]^T \in \mathbb{R}^{m\times1}$, hence $\mathbf{s} = [x_1l_1, x_2l_2, ...x_ml_m]^T \in \mathbb{R}^{m\times1}$. For convenience, we define a matrix $W = \mathrm{diag}(\mathbf{l})Y \in \mathbb{R}^{m\times n}$ that combines the effect of camera responsivity and scene illumination, and was introduced earlier as effective responsivity of the camera system. These conventions allow us to rewrite Equation 2.19 such that

$$\mathbf{x} = W^T \mathbf{r} + \boldsymbol{\eta}. \tag{2.20}$$

B Camera response simulations

The general framework for camera response simulations is in accordance with Equation 2.20. The SPD of the scene illumination, filter transmittances and the responsivity of the raw RGB sensor were obtained by measurement and W is computed based on these measurements. The channel gain factors γ_i are generally adjusted such that an expected maximum color signal does not result in a camera response that exceeds 80% of the dynamic range of the camera system. This operation condition is considered optimal with respect to SNR and utilization of the camera dynamic range.

We shall see that for the LSMSC system, channel gain factors can not be adjusted independently for each of the 12 channels. In fact, in later parts of this work we will also consider a 3- and a 6-channel system configurations and each system configuration constrains the selection of channel gains in a unique manner. To clarify we consider Figure 2.14, which illustrates the relation of physical location of projected sub-images on the camera line sensor and image channels for the three system configurations. From this figure it can be seen that for the 3- and 6-channel system configurations, each channel gain can be tuned to its optimal operation point because there exist independent signal amplifiers for each corresponding channel (ie. amplifiers for front and rear tab, R, G and B line sensor channels, as well as odd or even pixels[14]. For the 12-channel system configuration, two channels are influenced by the same gain factor (C1 and C4, C2 and C5, C3 and C6, C7 and C10, C8 and C11, C9 and C12). We model this dependence when simulating camera responses.

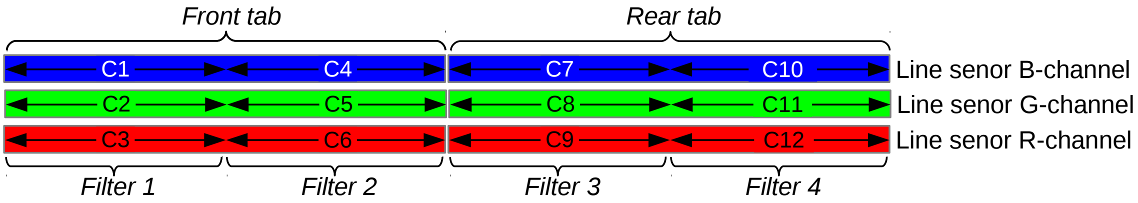
When it comes to simulating the noise term in Equation 2.20, we use a parametric noise model defined as

$$\boldsymbol{\eta} = \mathrm{diag}(\mathbf{a})\ \boldsymbol{\eta}_d(W\mathbf{r}) + \boldsymbol{\eta}_i, \tag{2.21}$$

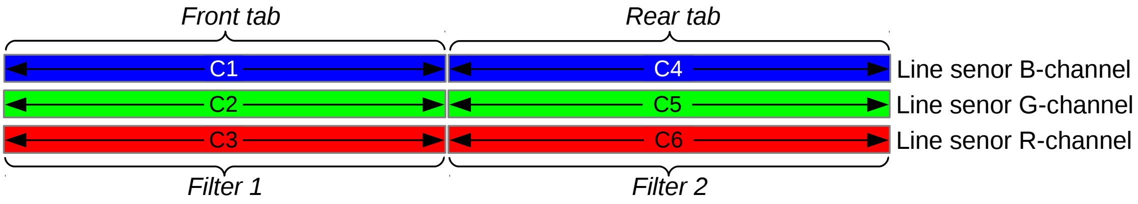
where $W \in \mathbb{R}^{n\times m}$ and $\mathbf{r} \in \mathbb{R}^{m\times1}$ are as before and elements in vector $\mathbf{a} = [a_1, a_2, ...a_n]^T \in \mathbb{R}^{n\times1}$ scale the influence of signal dependent noise (first term on the right in Equation 2.21)

---

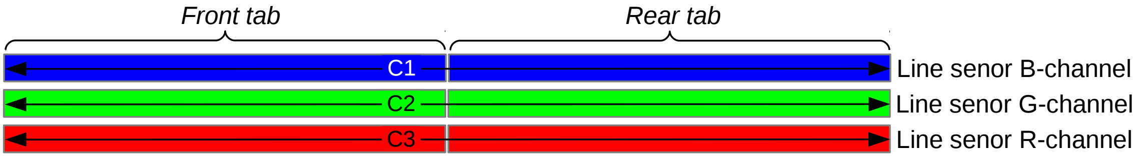
[14] See Section 2.3.2.B and Figure 2.12



(a) 12-channel configuration



(b) 6-channel configuration



(c) 3-channel configuration

Figure 2.14: Relation of physical location of projected sub-images on the camera line sensor and image channels for various LSMSC system configurations considers in this work.

with respect to signal independent noise (second term on the right in Equation 2.21) for each channel.

The signal dependency is mainly due to shot noise, which is a Poisson type noise and proportional to the square root of the signal ($\sqrt{W\mathbf{r}}$). We use a normal distribution to approximate the shot noise distribution. We also assume the signal independent noise term to be normally distributed and model both terms as statistically independent random variables with zero mean, so

$$\begin{aligned} \boldsymbol{\eta}_d(W\mathbf{r}) &\sim \sqrt{W\mathbf{r}}\,\mathcal{N}(0,\sigma_d^2) \\ \boldsymbol{\eta}_i &\sim \mathcal{N}(0,\sigma_i^2), \end{aligned} \tag{2.22}$$

where σ_d^2 and σ_i^2 are the variances of the normally distributed variables and "$\sim$" means proportional[15]. We distinguish two scenarios for noise simulations:

Noise Scenario 1: simulating a LSMSC at various levels of noise

This scenario allows us to study the influence of various levels of noise on the system performance explicitly. We make use of this in Section 5.5.1, when comparing noise properties of several spectral estimation approaches. For this scenario, parameter $\mathbf{a}$ in Equation 2.19 is adjusted such that equivalent amounts of $\boldsymbol{\eta}_d$ and $\boldsymbol{\eta}_i$ have an approximately similar influence on the average SNR of a particular simulated set of data.

---

[15] To avoid confusion, we explicitly note that σ_d^2 is the variance of a normal distribution and not the variance of the signal dependent noise term. This variance would be $\mathrm{var}(\sqrt{W\mathbf{r}}\,\mathcal{N}(0,\sigma_d^2))$, where $\mathrm{var}(\cdot)$ denoted variance.

Noise Scenario 2: simulating a LSMSC at a realistic fixed noise level

The camera manufacturer has characterized the noise properties of the LSMSC RGB sensor in great detail and provided parameters for $\boldsymbol{\eta}_d$. Noise variances of the signal dependent noise term were extracted from a raw RGB image of a white reference surface using the camera sensor without additional filters, and a variance of approximately 1 in 8-bit operation mode was found[16]. The channel-wise scale factors in $\mathbf{a}$ of Equation 2.19 are then adjusted such that the noise variances for the maximum signal (ie. the white reference surface) are scaled with the fraction of the signal that is left after introducing the additional color filters.

As signal independent noise (ie. dark current noise) is corrected for by dark image subtraction in practice, we assume $\boldsymbol{\eta}_i = 0$. Analog-to-digital converter related quantization noise is included in the model by a flooring operation on the real-valued simulated noisy camera responses. The effect of temporal averaging by over-sampling and spatial averaging homogeneous image regions is accounted for in the model as well, as it also reflects a realistic application scenario.

We consider this particularly realistic operation point of the system for instance in Chapter 3.4, where simulated camera response data is required for filter selection.

2.4 Framework for system evaluation

In the design and development of multi-spectral imaging systems, performance evaluation is a key aspect for a successful system implementation. *Performance* is a general expression, indicating *how well* some task is performed. Measurable performance parameters with respect to multi-spectral imaging systems are for instance image quality (spatial and spectral), color accuracy, computational complexity of computations or acquisition and/or data processing speed.

In this work, we barely consider computational complexity or speed related parameters, but mainly focus on design parameters related to spectral and color measurement accuracy. The next subsections introduce corresponding measures, the model validation technique used and the framework for model parameter selection.

2.4.1 Spectral estimation error

In spectral reflectance estimation, the estimation error is defined as the residual between an estimated reflectance $\widetilde{\mathbf{r}} = [\widetilde{r}_1, \widetilde{r}_2, ...\widetilde{r}_m]^T \in \mathbb{R}^{m\times 1}$ and its true[17] counterpart $\mathbf{r} = [r_1, r_2, ...r_m]^T \in \mathbb{R}^{m\times 1}$. This residual is a multi-dimensional quantity and not easy to assess individually when large sets of data are considered. It is a common practice to map this multi-dimensional vector to a single number and there are various ways to do this mapping. When considering a metric that operates on the spectral data, we refer to *spectral error*.

For color measurement, the spectral data is transformed to colorimetric data (for instance tristimulus values or color coordinates in CIE-Lab color space). We can express the aforementioned spectral residual vector as well in terms of a residual vector in the colorimetric space. Again, a mapping function can be employed to compute a single

---

[16] We note that for our experiments, we individual noise variance terms for each of the R, G and B channel.

[17] "True" in the sense of a ground-truth measurement. In our work we use the Konica Minolta FD7 spectrophotometer for that.

number, which is usually referred to as *color difference*. Hence, when considering a metric that operates on colorimetric data, we refer to *colorimetric error*.

Some authors have found it useful to combine spectral and color metrics for specific applications. An example is the work by López-Álvarez *et al.* [36] on multi-spectral imaging system design, in which a multivariate optimization problem is treated as univariate using a combined metric. There are no limitations in designing a function combining several metrics, but it can be difficult in practice to define a meaningful way of combining individual metrics. This is specifically the case when typical value ranges of individual metrics are different from each other.

In general, the selection of an appropriate error metric is application dependent. Because the focus of this work is reflectance and color measurement, we consider metrics from both categories.

A Quantifying spectral signal difference

The square root of the average of squared residuals between components in $\mathbf{r} \in \mathbb{R}^{m\times 1}$ and $\widetilde{\mathbf{r}} \in \mathbb{R}^{m\times 1}$ is called root-mean-square error and defined as

$$RMSE(\mathbf{r}, \widetilde{\mathbf{r}}) = \sqrt{\frac{1}{m}\sum_{i=1}^{m}(r_i - \widetilde{r}_i)^2}. \tag{2.23}$$

In this metric, each wavelength element is considered equally important.

Modified versions have been proposed to introduce a spectral weighting, which allows emphasizing spectral bands that are more important in a specific application than others. For color measurement, Imai *et al.* [37] proposed a weighting function that emphasizes those bands that contribute more to the tristimulus values of the signal.

Closely related to $RMSE$ is a logarithmic version called $RMSLE$, defined as

$$RMSLE(\mathbf{r}, \widetilde{\mathbf{r}}) = \sqrt{\frac{1}{m}\sum_{i=1}^{m}(\ln(r_i + 1) - \ln(\widetilde{r}_i + 1))^2}, \tag{2.24}$$

where $\ln(\cdot)$ denotes the natural logarithm. The $RMSLE$ weights spectral residuals with low reflectance value higher than corresponding residuals of the same magnitude with high reflectance value. We find this specifically useful in applications for color measurement from estimated reflectance data. In such applications, residuals from low magnitude spectral reflectances often contribute to large colorimetric errors and therefore should not result in a similar error value as corresponding residuals from high magnitude reflectances, as is the case for the $RMSE$ metric. An illustrative example of this effect is given in Figure 2.15, in which we also compare other metrics defined below.

Another way of mapping the residual vector to a single number is the Pearson distance, defined as

$$d_p(\mathbf{r}, \widetilde{\mathbf{r}}) = 1 - \frac{\mathbf{r}^T\,\widetilde{\mathbf{r}}}{\|\mathbf{r}\|\|\widetilde{\mathbf{r}}\|}, \tag{2.25}$$

which is one minus the normalized scalar product of the two vectors $\mathbf{r}$ and $\widetilde{\mathbf{r}}$. This measure can also be interpreted geometrically as the sine of the angle between the two vectors. If the vectors are parallel, the angle between them is zero and $d_p = 0$. The maximum Pearson distance is $d_p = 1$ and corresponds to vectors that are orthogonal. The geometrical interpretation helps understanding that this measure is invariant to scale changes of the vectors. However, large values of Pearson distance do not have a practical meaning when considering this metric to quantify spectral errors because the angular differences are usually very low. The measure is still highly informative because it is sensitive to changes in the vectorial shape and invariant to a signal offset.

The Pearson distance is also known as the complemented GFC (goodness of fit measure). To our knowledge, the GFC metric was first used by Romero *et al.* [38] to quantify spectral differences and is now commonly used for spectral measurements.

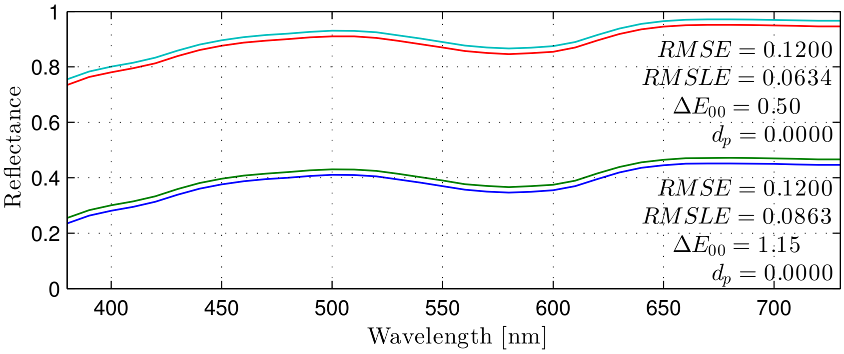


Figure 2.15: Comparison of $RMSE$, $RMSLE$, d_p and ΔE_{00} of sample reflectances.

B Quantifying color difference

The formulation of color differences aims at predicting the magnitude of the perceived difference between two color stimuli. Historically, the study of difference of color stimuli is nearly as old as the study of color itself and yet, remains a very active field of research until this date. The complexity of the task relates to the fact that the observer's judgment varies greatly with the conditions of the observation and the kind of stimuli (e.g. size, shape and texture of the stimuli and the display situation, as well as the spectral power distribution of the scene illumination). Apart from that, intra-observer variations occur [19].

The probably most commonly used color difference metric is the Euclidean distance of the residual vector of color coordinates in CIE-Lab space, denoted ΔE_{76} and computed as

$$\Delta E_{76}(\mathbf{x}_{Lab}, \widetilde{\mathbf{x}}_{Lab}) = \|\mathbf{x}_{Lab} - \widetilde{\mathbf{x}}_{Lab}\| \tag{2.26}$$

where $\mathbf{x}_{Lab} \in \mathbb{R}^{3\times1}$ and $\widetilde{\mathbf{x}}_{Lab} \in \mathbb{R}^{3\times1}$ are CIE-Lab coordinates corresponding to reflectances $\mathbf{r} \in \mathbb{R}^{m\times1}$ and $\widetilde{\mathbf{r}} \in \mathbb{R}^{m\times1}$ and $\|\cdot\|$ denotes Euclidean norm.

Deficiencies with respect to perceptual uniformity of the CIE-Lab color space resulted in the development of various improved color difference equations. For the case of the ΔE_{94} metric, modifications were introduced to account for the effect of a perceived decrease in chroma- and hue-difference with increasing chroma [39]. For the case of the ΔE_{00} metric, an additional hue rotation term was developed to correct the problematic blue region for hue angles in the neighborhood of 275 °. Apart from that, compensation for neutral colors, lightness, chroma and hue were introduced [40]. The ΔE_{00} metric is described by the CIEDE2000 color-difference formula, which is the latest recommendation of the CIE.

For this work, if not noted differently, the formulation of color difference is based on the ΔE_{00} metric. The computation for two color samples with known CIE-Lab color coordinates is described in Appendix A.3.

2.4.2 Model validation

In later sections, we will use learning-based models to estimate reflectance data. One criterion to validate the performance of these methods is generalization, ie. how well do the results generalize to an independent dataset. Because we sometimes do not consider an independent dataset for testing, we can use a cross-validation scheme to evaluate estimation performance and by that reduce the risk of over-fitting the model.

Cross-validation is a technique in which for consecutive rounds of evaluation, portions of the dataset are used for model training and the rest for testing. By that, all available data is used for training and testing (but never at the same time). The overall performance is typically computed as the average performance over all evaluation rounds.

The type of cross-validation technique used is *k-fold cross-validation* [41]: in each of k rounds of evaluation, a subset of $\lfloor \frac{l}{k} \rfloor$ of the initially shuffled l samples of the validation dataset are used for testing, and the rest for training the model. The process is illustrated schematically for $k = 4$ in Figure 2.16. In this dissertation, $k = 10$ is used, ie. 10-fold cross-validation.

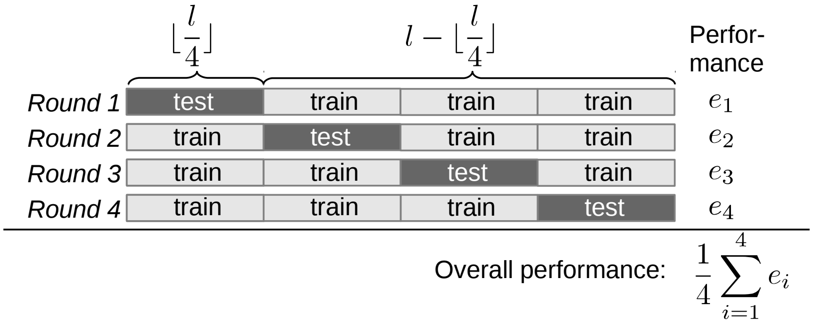


Figure 2.16: Schematic illustration of $k = 4$-fold cross-validation process for a validation dataset with l samples.

2.4.3 Parameter selection

Various spectral estimation models contain parameters. The most common way to set these parameters is grid search, which in its simplest form is a process that consists of empirically evaluating performance of the model at predefined values in a parameter grid, and subsequent selection of the parameter combination that minimizes a certain loss function. We note that the computational demand for grid search increases exponentially with the number of model parameters. However, parameter combinations are independent from each other and the grid search can therefore be parallelized to speed up computations.

Simple grid search is prone to under-sampling the associated loss as a function of the parameter grid values. The risk of under-sampling is specifically high if the parameter search space is continuous and the relation of parameter range and step size with respect to model performance is not clear a priori. Hence, we use simple grid search followed by grid refinement to reduce the risk of under-sampling. Accordingly, after simple grid search, the parameter search range is replaced by a refined grid and the grid search is repeated. The refined grid is bounded by neighboring parameter values around the selected parameters, and we use a linear grid refinement with 10 refined parameter steps per parameter.

Accordingly, assuming vector $\mathbf{p}_h \in \mathbb{R}^{n \times 1}$ contains n monotonically increasing values for parameter h, scalar o is the index of the parameter found in simple grid search that minimizes the loss, and $\Delta p = \frac{\mathbf{p}_h(o+1) - \mathbf{p}_h(o-1)}{10}$ is the parameter spacing for 10 step grid refinement, then

$$\mathbf{p}_{r,h} = [\mathbf{p}_h(o-1), \mathbf{p}_h(o-1)+\Delta p, \mathbf{p}_h(o-1)+2\Delta p, ..., \mathbf{p}_h(o-1)+10\Delta p]^T \in \mathbb{R}^{11\times 1} \tag{2.27}$$

denotes the refined parameter grid vector if $1 < o < n$. If $o = 1$, the parameter spacing is $\Delta p = \frac{\mathbf{p}_h(o+1)-\mathbf{p}_h(o)}{10}$ and the grid vector

$$\mathbf{p}_{r,h} = [\mathbf{p}_h(o), \mathbf{p}_h(o)+1\Delta p, \mathbf{p}_h(o)+2\Delta p, ..., \mathbf{p}_h(o)+10\Delta p]^T \in \mathbb{R}^{11\times 1}. \tag{2.28}$$

Otherwise $o = n$, the parameter spacing is $\Delta p = \frac{\mathbf{p}_h(o)-\mathbf{p}_h(o-1)}{10}$ and the grid vector is as in Equation 2.27. We illustrate the parameter selection process schematically for the case of 2-parameter grid search and 5-step parameter grid refinement in Figure 2.17.

The loss function considered in this work is the average $RMSE$ associated to reflectance estimation and the data considered is the training data in the cross-validation process. To minimize risk of over-fitting parameters to the training data, we use again a 10-fold cross-validation scheme for parameter selection.

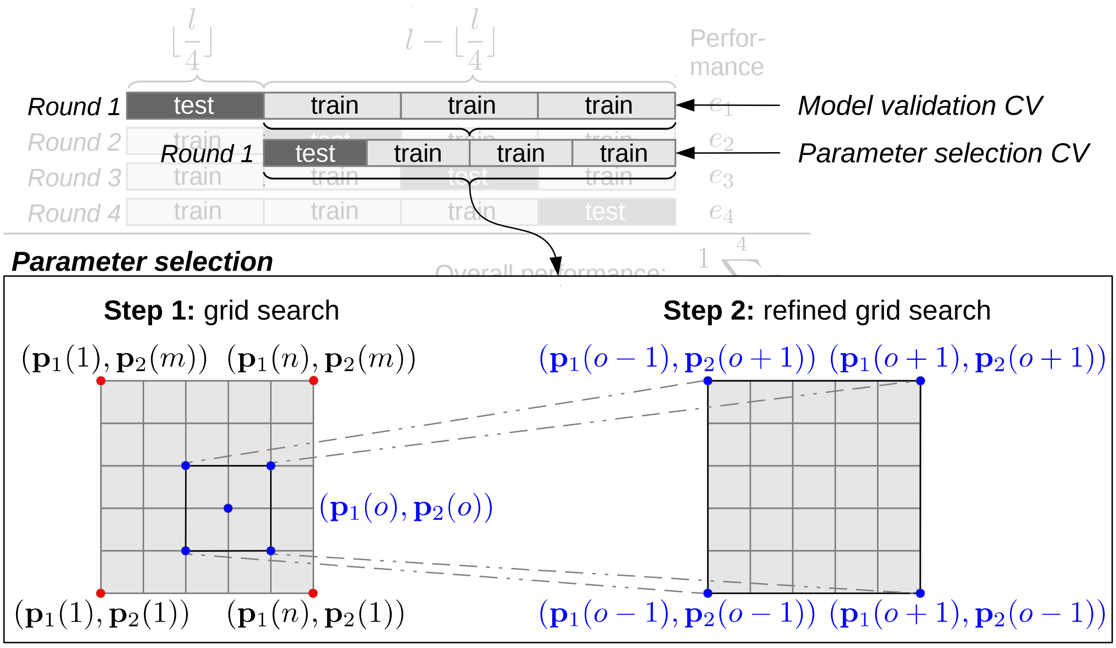


Figure 2.17: Schematic illustration of 2-parameter grid search with 5-step parameter grid refinement in a 4-fold cross-validation scheme. Parameter combinations are denoted $(\mathbf{p}_1(i), \mathbf{p}_2(j))$, where $\mathbf{p}_1 \in \mathbb{R}^{n\times 1}$ and $\mathbf{p}_2 \in \mathbb{R}^{m\times 1}$ are vectors containing values of predefined parameter ranges and $i = 1, 2, ...n,\ j = 1, 2, ...m$.

2.5 System configurations and datasets considered in this work

Many design parameters of a multi-spectral imaging system for reflectance measurements are application specific. For instance, the selection of an algorithm for recovery/estimation of the spectral signal depends on the availability of certain data. Another example is that the required measurement performance for a certain application will influence the system design because performance is closely linked to the number of imaging channels. In fact, the type and number of spectral channels available is also considered in the selection of a suitable algorithm.

We take these aspects as motivation for considering various LSMSC system configurations in this dissertation. The following list describes the systems used[18]:

System configuration 1 ($\mathbf{SC1_{12C}}$): 12-channel LSMSC manufactured by Chromasens GmbH (see Section 2.3.2).

System configuration 2 ($\mathbf{SC2_{12C}}$): 12-channel LSMSC with 4 filters selected as part of this work (see Chapter 3.4.4).

System configuration 3 ($\mathbf{SC3_{6C}}$): 6-channel LSMSC with 2 filters selected as part of this work (see Chapter 3.4.3). The effective scan resolution is a factor of 2 higher than that of the 12-channel system.

System configuration 4 ($\mathbf{SC4_{3C}}$): 3-channel LSMSC with 1 filter selected as part of this work (see Chapter 3.4.2). The effective scan resolution is a factor of 4 higher than that of the 12-channel system.

The data considered in experimental evaluations consists of several sets of measured spectral reflectance factor data and corresponding camera responses. Whenever simulated camera responses are considered, we refer to noise scenario 1 or 2 of the model described in Section 2.3.3. Whenever measured camera responses are considered, we refer to data extracted from multi-channel images of color charts acquired with our laboratory LSMSC system ($\mathbf{SC2_{12C}}$), or explicitly describe something else. These charts contain color patches of size 11 mm × 11 mm, arranged in a rectangular array on A3 size cardboard. The color patches are manually punched from color swatches of Pantone, HKS and RAL paints. Camera responses for each patch were extracted from the charts by averaging pixels in an area that corresponds approximately to the size of the measurement aperture of the FD7 spectrophotometer, which is a circle with 3.5 mm diameter.

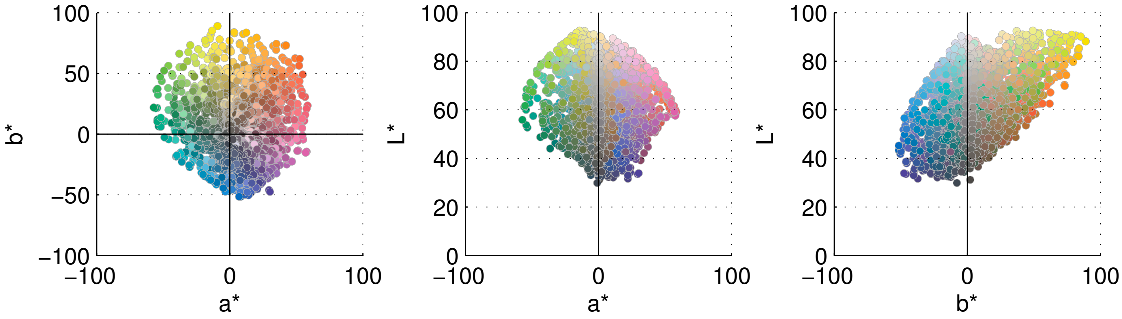
A summary of relevant properties of the datasets is given in Table 2.1. Spectral reflectance factor data was obtained with the Konica Minolta FD7 spectrophotometer. The data was measured with 3 repetitions and the average taken. We illustrate this spectral data in Figure 2.18 and the corresponding CIE-Lab coordinates in Figure 2.19.

---

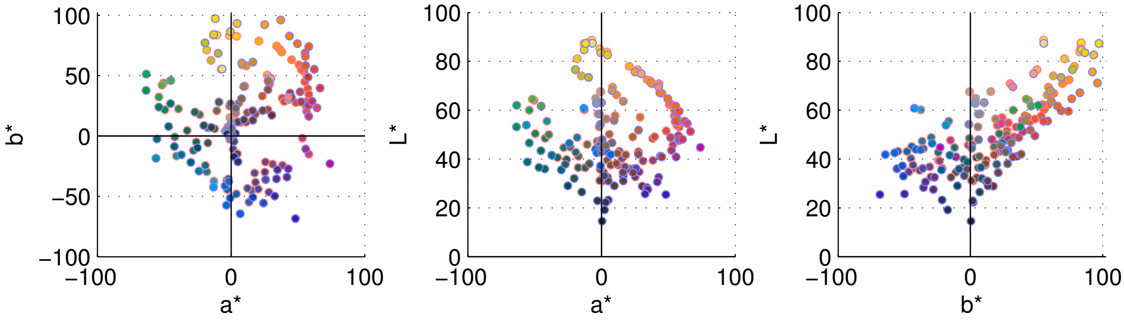
[18] Refer to Figure 2.14 for an illustration of the physical relation of projected sub-images on the camera line sensor and image channels.

Table 2.1: Datasets considered in experimental evaluations of this dissertation.

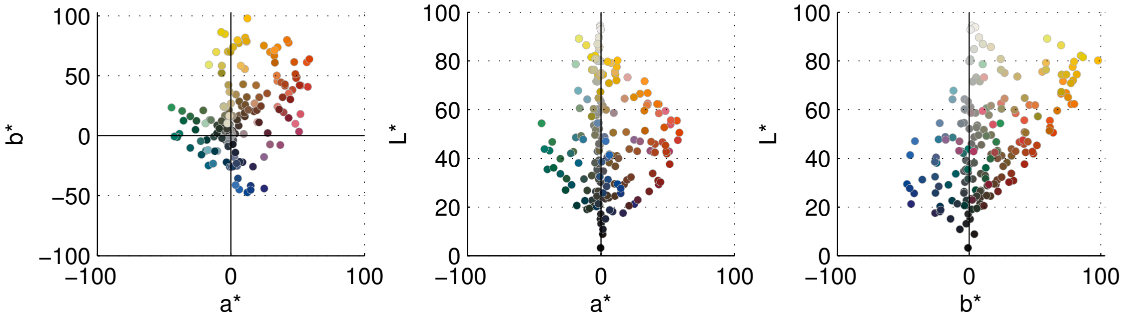
Name	**Material / Production**	#	**Surface**
Pantone solid chips (**DS1**)	Offset printed using 39 Pantone basic color inks on uncoated 118 g/m^2 premium grade paper including optical brighteners	1761	matt
HKS-N (**DS2**)	Offset printed on 100 g/m^2 wood-free uncoated paper, using 11 HKS basic inks + metallic and gold single component colors	90	matt
HKS-K (**DS3**)	Offset printed on 115 g/m^2 wood-free premium grade paper, using 11 HKS basic inks + metallic and gold single component colors	88	glossy
RAL classic 840-HR (**DS4**)	Paper substrate and hidrosoluble varnish	213	semi matt



(a) DS1: Pantone solid chips

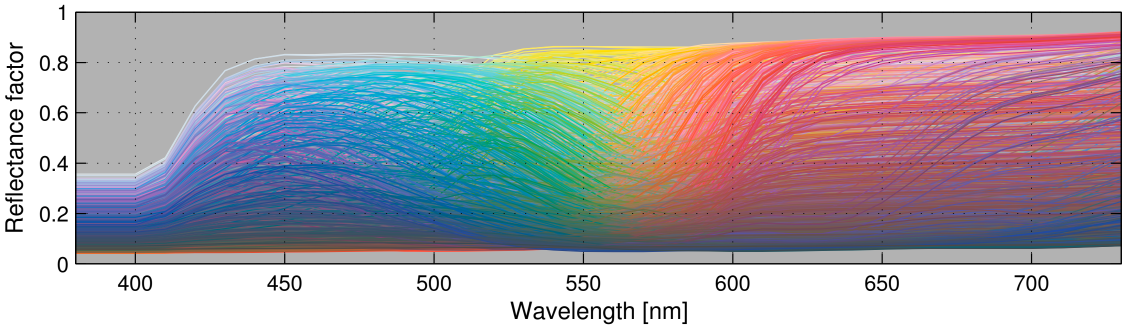


(b) DS2: HKS-N (blue circles) and **DS3**: HKS-K (red circles)

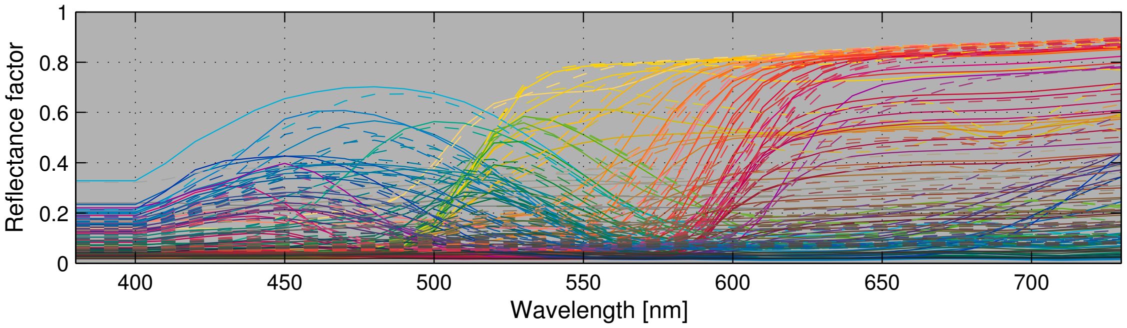


(c) DS4: RAL classis 840-HR

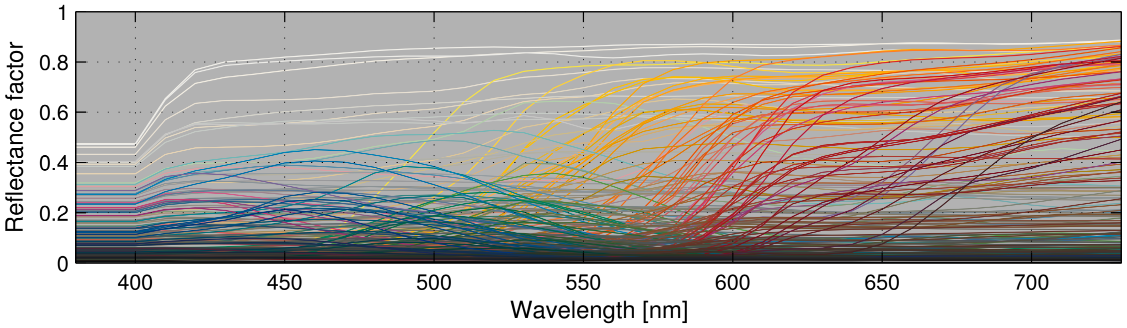
Figure 2.18: Color coordinates of measured reflectance data **DS1-DS4**, converted to CIE-Lab using CIE-10°standard observer and CIE-D65 standard illuminant.



(a) DS1: Pantone solid chips



(b) DS2: HKS-N (solid line) and **DS3**: HKS-K (dashed line)



(c) DS4: RAL classis 840-HR

Figure 2.19: Measured reflectance data **DS1-DS4**.

"Zu einem vollkommenen Menschen gehört die Kraft des Denkens,
die Kraft des Willens, die Kraft des Herzens."

From German:
To a complete human belongs the power of thought,
the power of will, and the power of heart.
Ludwid Feuerbach (1804–1872)

3
Acquisition system design: filter selection

Contents

3.1 Preface

The development of multi-spectral imaging systems involves the design of various system components, such as the camera module, light source and in line-scan applications a translation stage for the scanning movement of camera or scan object. It is obvious that a successful system implementation requires expertise in the fields of mechanical, opto-mechanical and electrical engineering and beyond. In this dissertation, we constrain our discussion to optical components of a LSMSC system, namely camera system responsivity (including filter transmittances) and scene illumination. Assuming the camera response model presented in Section 2.3.3, modification of the camera system spectral responsivities can have a similar influence on the generation of camera responses as modification of the scene illumination. Hence, LSMSC system design should involve both components with equal importance.

In this dissertation, we regard the system design task as an optimization problem and hence need to define a measure of system optimality, subject to which a solution is developed. The following section addresses system optimality in detail. Until this point, we do not make any assumption whether optimization of scene illumination or the camera module's spectral responsivity (for instance by modifying filter transmittances) is considered.

In Section 3.3, we continue more specifically by considering only the filter selection problem for a system with fixed illumination SPD. This is the problem of optimizing the system by selection of a set of filters that, in combination with the spectral responsivity of the camera system and the fixed scene illumination, minimize one or multiple objective functions of system optimality. An exhaustive evaluation of all possible filter combinations from a set of available filters can be very time consuming and hence heuristic techniques are motivated to simplify the task. A heuristic approach is a problem solving strategy that is not guaranteed to be optimal, but sufficient for a predefined goal. In Section 3.4, we first introduce the application specific heuristics used in this dissertation and then exemplify filter selection for various LSMSC system configurations in form of three case studies (ie. filter selection for a 3-, 6- and 12-channel LSMSC). This chapter is concluded with a summary and a discussion about potential future research.

3.2 System optimality criteria

As mentioned earlier, the system design optimization problem requires us to define system optimality in form of an objective function, subject to which an optimal solution is developed. The selection of an objective function is problem specific. In brief, if for instance the measurement of device-independent color is considered as an application of the multi-spectral imaging system, it is desirable to optimize the system for a low prediction error of color coordinates. If the spectral estimation of surface reflectances is considered, the objective function has to reflect prediction error of spectral reflectance functions. This trivial view of the problem can be described mathematically in a more general form using linear algebraic concepts of subspace matching and vector projection. Trussel, Vora and Sharma [42, 43, 44] conducted extensive research related to imaging system design. In [45], Ng *et al.* summarized the aforementioned and other related works and presented a methodology for imaging system filter design. The following subsection is therefore closely related to these works.

Remarks on the notation
We recall from Section 2.1 that $\mathcal{X} = C(X)$ denotes the column space of a matrix $X \in \mathbb{R}^{m\times n}$, which is the linear span of the column vectors of X, and that $\mathcal{X}$ is a subspace of $\mathbb{R}^m$. Further, let $P_{\mathcal{X}} \in \mathbb{R}^{m\times m}$ be a projection matrix that projects a vector $\mathbf{x} \in \mathbb{R}^{m\times 1}$ onto subspace $\mathcal{X}$. We can denote O_X as the orthonormal basis of $\mathcal{X}$, accordingly $C(O_X) = C(X)$ and $O_X^T O_X = I$, where I denotes the identity matrix. Such a basis can be obtained by Gram-Schmitt orthogonalization [46, p. 307].

3.2.1 Subspace projection error

The effective responsivity of the multi-spectral acquisition system is a matrix $W \in \mathbb{R}^{m\times n}$. It spans a subspace of the spectral space $\mathcal{R} \subset \mathbb{R}^m$, which we denote as $\mathcal{W} = C(O_W) = C(W) \subset \mathcal{R}$ and refer to as sensor subspace. It can be shown that $\mathcal{R} = \mathcal{W}_{\|} + \mathcal{W}_{\perp}$, where $\mathcal{W}_{\|} = W$ and $\mathcal{W}_{\perp}$ is the nullspace of W. Further, each surface reflectance vector $\mathbf{r} \in \mathcal{R}$ can be written uniquely as $\mathbf{r} = \mathbf{r}_{\|} + \mathbf{r}_{\perp}$, where $\mathbf{r}_{\|} \in \mathcal{W}_{\|}$ and $\mathbf{r}_{\perp} \in \mathcal{W}_{\perp}$. The decomposition of $\mathbf{r}$ into $\mathbf{r}_{\|}$ and $\mathbf{r}_{\perp}$ holds an important interpretation: $\mathbf{r}_{\|}$ is the fundamental part of $\mathbf{r}$ with respect to W, which contributes to the formation of the camera response. Vector $\mathbf{r}_{\|}$ is called the orthogonal projection of $\mathbf{r}$ onto $\mathcal{W}$. The vector $\mathbf{r}_{\perp}$ is the the part of $\mathbf{r}$ that lies in the nullspace of the system, ie. which is not *seen* by the camera. We call $\mathbf{r}_{\perp}$ the *metameric black* part [1].

The orthogonal projection matrix [46, p. 429] $P_{\mathcal{W}} \in \mathbb{R}^{m\times m}$ that projects onto $\mathcal{W}$ is defined as

$$P_{\mathcal{W}} = W(W^T W)^{-1} W^T, \tag{3.1}$$

where the columns of $W \in \mathbb{R}^{m\times n}$ are a set of effective responsivity vectors of the multi-spectral image acquisition system, which are usually linearly independent. This projection matrix can also be expressed in terms of the orthonormal basis $O_W \in \mathbb{R}^{m\times n}$ of the effective responsivity as

$$P_{\mathcal{W}} = O_W O_W^T. \tag{3.2}$$

The projection matrix $P_{\mathcal{W}}$ is the unique linear operator such that $P_{\mathcal{W}}\mathbf{r} = \mathbf{r}_{\|}$. It follows as well that $\mathbf{r}_{\perp} = \mathbf{r} - P_{\mathcal{W}}\ \mathbf{r}$. We could now decompose a reflectance into fundamental and metameric black part.

For conceptual reasons, we define another subspace of $\mathcal{R}$, which we call *application subspace* and denote as $\mathcal{A}$. The motivation for its definition will become clear later and the meaning of the application subspace is explained by two examples: For a multi-spectral system specifically designed for color measurement applications, subspace $\mathcal{A}$ is spanned by the vectors of the CMFs in $Y_{CMF} \in \mathbb{R}^{m\times 3}$ weighted by the standard illuminant $\mathbf{l} \in \mathbb{R}^{m\times 1}$ considered for color measurement (ie. $\mathcal{A}$ is the set of all linear combinations of weighted CMF vectors). For a multi-spectral system specifically designed for spectral reflectance measurements of a certain application domain, this space is the set of all linear combinations of reflectance vectors related to the application.

From a general perspective, we can distinguish three important cases that describe the relation of $\mathcal{W}_{\|}$, $\mathcal{W}_{\perp}$ and $\mathcal{A}$ with respect to image acquisition. These cases are graphically depicted in Figure 3.1 and described in the following:

---

[1] The notion of *metameric black* was introduced by Wyszecki in the 1950s. He considered the CMFs weighted by a particular standard illuminant as the system and referred to $\mathbf{r}_{\perp}$ as metameric black because the response of this system (the tristimulus values) are equal to zero [19].

Case 1: the application subspace $\mathcal{A}$ intersects with the sensor subspace $\mathcal{W}_{||}$ and is not a subspace of $\mathcal{W}_{||}$. The acquisition system therefore captures partial information of reflectances in $\mathcal{A}$ and this particular information can be recovered[2].

Case 2: the application subspace $\mathcal{A}$ is a subset or equal to the sensor subspace $\mathcal{W}_{||}$. All information of reflectances in $\mathcal{A}$ is retained in the acquisition process and hence, all information can be recovered.

Case 3: the application subspace $\mathcal{A}$ does not intersect the sensor subspace $\mathcal{W}_{||}$. Image acquisition does not capture any information related to reflectances in $\mathcal{A}$ and hence, no information can be recovered.

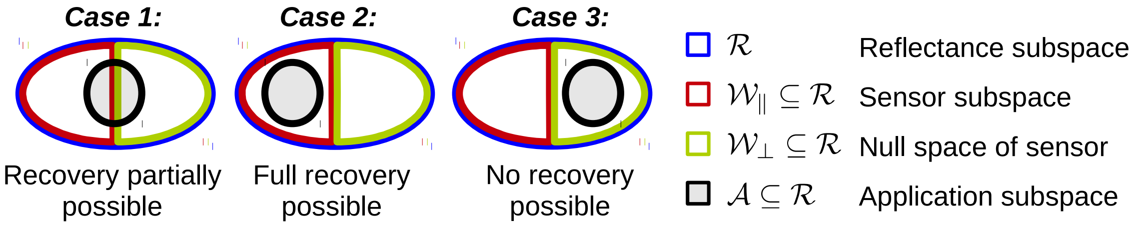


Figure 3.1: Graphical illustration of the relation of $\mathcal{W}_{||}$, $\mathcal{W}_{\perp}$ and $\mathcal{A}$ with respect to image acquisition.

The orthogonal projection matrix $P_{\mathcal{A}} \in \mathbb{R}^{m \times m}$ that projects onto $\mathcal{A} \subset \mathcal{R}$ is defined as

$$P_{\mathcal{A}} = A(A^T A)^{-1} A^T, \tag{3.3}$$

where the columns of A are a set of linearly independent vectors that define a basis for the application subspace[3]. As before, we can define

$$P_{\mathcal{A}} = O_{\mathcal{A}} O_{\mathcal{A}}^T, \tag{3.4}$$

where $O_{\mathcal{A}} \in \mathbb{R}^{m \times m}$ is an orthonormal basis of $\mathcal{A}$.

Given previous definitions, we can establish a measure of projection error for reflectance $\mathbf{r} \in \mathcal{A}$, projected via $\mathcal{W}$ onto $\mathcal{A}$ as

$$\boldsymbol{\epsilon}_{\mathbf{r}} = \left(P_{\mathcal{A}} - P_{\mathcal{A}} P_{\mathcal{W}}\right) \mathbf{r}, \tag{3.5}$$

which is the difference between direct projection of $\mathbf{r}$ on the application subspace and the projection via the effective responsivity space. The projection error $\boldsymbol{\epsilon}_{\mathbf{r}} \in \mathbb{R}^{m \times 1}$ is a vector in $\mathcal{R}$ space that quantifies the spectral information that is lost from $\mathbf{r}$ in the image acquisition process.

Obviously, if the application subspace $\mathcal{A}$ is the spectral space $\mathcal{R}$, this residual is the part of $\mathbf{r}$ that lies in the nullspace of the effective responsivity, ie. $\mathcal{W}_{\perp}$. If $\mathcal{A} \subseteq \mathcal{W}$, full recovery of $\mathbf{r}$ is possible and $\boldsymbol{\epsilon}_{\mathbf{r}} = 0$ (Case 2 in Figure 3.1).

According to equation 3.5, a measure of system optimality can be defined as

$$\begin{aligned} \mathcal{E} &= \sum_{i=1}^{l} \left\| \left(P_{\mathcal{A}} - P_{\mathcal{A}} P_{\mathcal{W}}\right) \mathbf{r}_i \right\|^2 \\ &= \operatorname{Tr}\left([(P_{\mathcal{A}} - P_{\mathcal{A}} P_{\mathcal{W}})R]^T [(P_{\mathcal{A}} - P_{\mathcal{A}} P_{\mathcal{W}})R] \right). \end{aligned} \tag{3.6}$$

---

[2] With recovery we refer to mathematically inverting the image acquisition model. This description is vague and shall only illustrate the process conceptually. We will come back to recovery of spectral information from camera responses by a physical model in Section 5.2.

[3] See Section 2.2.4

This is the general expression for the sum of squared projection error E for an ensemble of l reflectances $R = [\mathbf{r}_1, \mathbf{r}_2, ...\mathbf{r}_l] \in \mathbb{R}^{m \times l}$, under projection to a certain application subspace.

A Application scenarios and related literature

The general framework of quantifying subspace projection errors as a criterion for system optimality can be defined more specifically for particular applications. In the following, we present some application scenarios and derive expressions for subspace projection error from Equation 3.6:

- If the multi-spectral imaging system is considered for colorimetric measurement (ie. measurement of the device independent XYZ color coordinates), the application subspace is $A = W_{CMF} = \text{diag}(\mathbf{l})\ Y_{CMF} \in \mathbb{R}^{m \times 3}$. The expression for the subspace projection error becomes:

  $$\mathcal{E} = \sum_{i=1}^{l} \|(O_{W_{CMF}} O_{W_{CMF}}^T - O_{W_{CMF}} O_{W_{CMF}}^T O_W O_W^T)\ \mathbf{r}_i\|^2, \tag{3.7}$$

  This error measure was proposed by Vora *et al.* [42, Eqn. 14] and applied as a measure of goodness of a set of color scanning filters.

  It is obvious that this subspace projection error is data dependent, as it quantifies the part of the reflectances in $\mathcal{R}$ in the application subspace $\mathcal{A}$ that lies in the nullspace of sensor sub-space $\mathcal{W}$. A potential danger in considering the data structure in the system optimality measure is that the data available for system design varies structurally from the data later on used in the application. Based on the previous expression of projection error, Vora *et al.* derived a measure that is data independent. They take the assumption that any spectrum $\mathbf{r}$ can be expressed as a sequence of independent, identically distributed random variables and by that obtain the normalized measure

  $$\upsilon = \frac{\sum_{i=1}^{n} {\lambda_i}^2}{n}, \tag{3.8}$$

  where λ_i is the i-th singular value of the matrix $O_W O_{W_{XYZ}}^T$ [42, Eqn. 23]. Matrix $O_W \in \mathbb{R}^{m \times n}$ is the orthonormal basis of the effective responsivity of the acquisition system matrix $W \in \mathbb{R}^{m \times n}$, the orthonormal basis of the CMFs is $O_{W_{XYZ}} \in \mathbb{R}^{m \times 3}$ and n is the number of sensor spectral channels. This measure is 1 if the projection error is zero. Contrary to the aforementioned, including no information on the statistical distribution of the reflectance data in spectral space might be misleading, as a sensor corresponding to a low υ-value might be a good spectral subspace match, but not necessarily a good match for a particular set of reflectances of a certain application. In most practical cases, statistical information of the application reflectances are actually known and can therefore be included in the computation of subspace projection error.

- If the system is considered for estimation of reflectances and we assume that spectral reflectance vectors can be expressed as a sequence of independent and identically distributed (i.i.d) random variables, the application subspace is $\mathbb{R}^m$ and $P_{\mathcal{A}} = I_m \in \mathbb{R}^{m \times m}$ the identity matrix. Substituting $P_{\mathcal{A}}$ in equation 3.6, we obtain

  $$\mathcal{E} = \sum_{i=1}^{l} \|(I - P_{\mathcal{W}})\ \mathbf{r}_i\|^2. \tag{3.9}$$

Ng *et al.* [45, Eqn. 15] follow a similar approach to define a measure that sums the projection error for an available set of reflectances. The i.i.d. assumption is not true for natural surface reflectances[4]. So instead of considering $\mathbb{R}^m$ as application subspace, one can use

$$P_{\mathcal{A}} = O_R O_R^T, \tag{3.10}$$

where O_R is an orthonormal basis of the subspace spanned by a set of l spectral reflectances $R = [\mathbf{r}_1, \mathbf{r}_2, ...\mathbf{r}_l]$ that are representative for the application in consideration. Following this approach, statistical information about the nature of the reflectance data is taken into account when defining the projection onto application subspace. Accordingly, Equation 3.6 becomes

$$\mathcal{E} = \sum_{i=1}^{l} \|(O_R O_R^T - O_R O_R^T O_W O_W^T)\ \mathbf{r}_i\|^2, \tag{3.11}$$

which equals to

$$\mathcal{E} = \sum_{i=1}^{l} \|O_R O_R^T (I - O_W O_W^T)\ \mathbf{r}_i\|^2, \tag{3.12}$$

which, in turn, is similar to the mean-square error expression derived by Vora *et al.* [42, Eqn. 12]. This time, we do not consider the subspace spanned by the color matching functions as application subspace, but the subspace spanned by the application surface reflectance data.

3.2.2 Spectral estimation error

If the application of the multi-spectral imaging system under consideration is based on spectral reflectance estimation, system optimality can also be defined in terms of an estimation error for a particular set of data. The link of such a measure to system optimality is indirect, because the estimation error depends not only on system properties, but also on the estimation algorithm considered. The spectral estimation problem was mentioned in Chapter 2, but not defined any further. We will come back to this later and for now consider the process of spectral estimation as a black box.

From a practical point of view, there exists a particular advantage of using estimation error as optimality criterion, namely that it allows to take into account specific application constraints. Examples of constraints are camera parameters such as the range of feasible exposure times, camera gain levels and the camera system noise model. Each of those influences the generation of camera responses and consequently the estimation error, while none of them has an influence on the basic subspace projection based error criteria defined before.

Several measures for quantifying spectral signal differences as well as differences between color stimuli were introduced in Section 2.4.1. The selection of a suitable metric, or a combination of several metrics, is application dependent. A general expression of the average error for a given dataset of surface reflectances is

$$\mathcal{E} = \frac{1}{l} \sum_{i=1}^{l} \epsilon(\mathbf{r}_i, \tilde{\mathbf{r}}_i), \tag{3.13}$$

---

[4] In [47], it was found that reflectances for a range of real-world environments are well described by a beta or mixture of Gaussian distribution.

where $\mathbf{r}_i \in \mathbb{R}^{m\times 1}$ and $\tilde{\mathbf{r}}_i \in \mathbb{R}^{m\times 1}$ are the i-th measured and estimated surface reflectance respectively, l is the number of reflectance samples in the application dataset and ϵ is one of the error metrics from section 2.4.1.

Related literature

Using estimation error as a system optimality criterion was considered by various researchers. Ng *et al.* [45, Eqn. 30] used a colorimetric error term among others in least squares estimation with linear feature mapping for filter selection. Shen *et al.* [48, Eqn. 12] minimized a RMSE cost function in a binary differential evolution based filter selection of a multi-spectral imaging system. Their reflectance estimation is based on Wiener estimation (which will be introduced in Section 5.2.2). Also Haneishi *et al.* [49] used a Wiener model and evaluated their filter selection by spectral and colorimetric measures.

3.2.3 Subspace projection and estimation error comparison

Subspace projection error as well as estimation error based system optimality evaluations have certain advantages but also come with drawbacks. To illustrate some important properties of the two categories of optimality measures, we consider the following toy examples:

System 1 and System 2 are 3-channel imaging systems for device independent color measurements. Their effective spectral responsivities are depicted in color in Figure 3.2. The observation condition considered for color measurement relates to the CIE 10° Standard Observer and CIE Standard Illuminant D65. We illustrate the corresponding CMFs alongside System 1 and 2 in gray. We note that these spectral curves span the application subspace, which means $A = W_{CMF} = \text{diag}(\mathbf{l})\, Y_{CMF} \in \mathbb{R}^{m\times 3}$. System 1 is characterized by spectral channels that are narrow-band and whose peaks are located approximately at the peaks of the CMFs. System 2 is characterized by two spectral channels that are identical with the CMFs and a third narrow-band channel with peak wavelength at 400 nm.

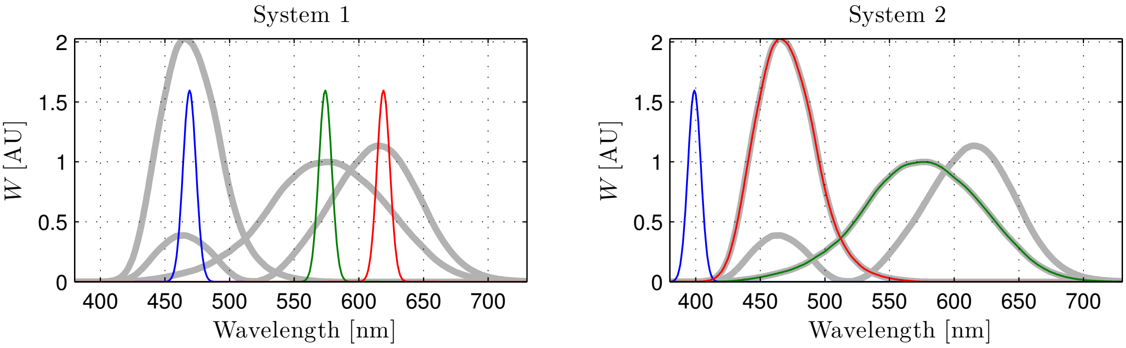


Figure 3.2: Effective spectral responsivities of System 1 and 2 (in color) considered in a toy example for comparing subspace projection error and estimation error. Illustrated in gray are the CMFs corresponding to observation condition considered for color measurement.

An intuitive interpretation when comparing System 1 and 2 is that System 2 can not perform well in color measurement because of missing information in the red part of the spectrum of visible light, even though the other two channels are identical with the CMFs. System 1 is clearly not ideal either, as the narrow-band channels are quite distinct from the CMFs.

We proceed by comparing the two systems with respect to previously defined system optimality criteria. We denote $\mathcal{E}_1$ as the data dependent subspace projection error measure defined in Equation 3.7. Matrix $R = [\mathbf{r}_1, \mathbf{r}_2, ...\mathbf{r}_l] \in \mathbb{R}^{m \times l}$ contains reflectance factor data from dataset **DS1**. Further, let $\mathcal{E}_2 = 1 - \upsilon$ be a data independent measure for system optimality, which is related to Equation 3.8. We also consider an estimation error based metric $\mathcal{E}_3$, which is the average ΔE_{76} color difference computed from estimated reflectances of System 1 and 2 with respect to the measured reflectances[5]. Camera response data was simulated according to **Noise Scenario 1** in Section 2.3.3.B, with noise variances $\sigma_d = \sigma_i = 0$, ie. noise free. The corresponding numerical values of the three error measures for System 1 and 2 are listed in Table 3.1.

Table 3.1: Comparison of subspace projection and estimation based error metrics in a toy experiment.

System	$\mathcal{E}_1$	$\mathcal{E}_2$	$\mathcal{E}_3$
System 1	29669	0.68	**7.54**
System 2	**7441**	**0.33**	15.20

Like expected, the color measurement performance of System 1 is better than for System 2. This is manifested by a lower colorimetric error $\mathcal{E}_3$ for System 1. Not so obvious is why System 2 has a lower subspace projection error $\mathcal{E}_1$, even though the application subspace is based on the color measurement task. This result can be understood when recalling that this error measures the summed spectral residual that lies in the sensor nullspace with respect to the application subspace for a given set of reflectance data. By design of the two systems, this spectral residual is zero for two channels of System 2, but not for System 1. Hence, the optimal system determined by $\mathcal{E}_1$ does not coincide with the optimal system for the given task. Even though $\mathcal{E}_2$ is data independent, we see the same effect. As noted by Vora *et al.* [42], this measure can be related to the principal angles between the sensor subspace $\mathcal{W}$ and application subspace $\mathcal{A}$. So from the point of view of subspace projection error, there is a greater loss in information for System 1 as compared with System 2.

Another drawback of subspace projection based optimality is related to the previously mentioned property that corresponding measures can not account for specific application constraints. We illustrate this with a second toy example. Let's consider the effective responsivities of two 2-channel camera systems illustrated in Figure 3.3. The two camera response components of System 3 to an achromatic reflectance are of the same value, as both channels are equally responsive (even though in different spectral ranges). In System 4, the channel illustrated in red is a factor of 10 less responsive than the channel illustrated in blue. The subspace projection error (Equation 3.6) for the achromatic reflectance is identical for System 3 and 4. This is because the responsivity scale factor is only a linear operator on the responsivities that does not affect the subspace projection. However, the influence of noise in the less responsive channel might result in a larger spectral estimation error. Hence, if a model for simulating camera responses is available and spectral estimation error is used as a measure of system optimality, candidate systems whose camera responses are heavily corrupted by noise should not be favored.

Even though considering spectral estimation error as a system optimality criterion seems to be the preferable choice for its practical reasons, care has to be taken. This

---

[5] The evaluation scheme was based on 10-fold cross-validation as described in Section 2.4.2.

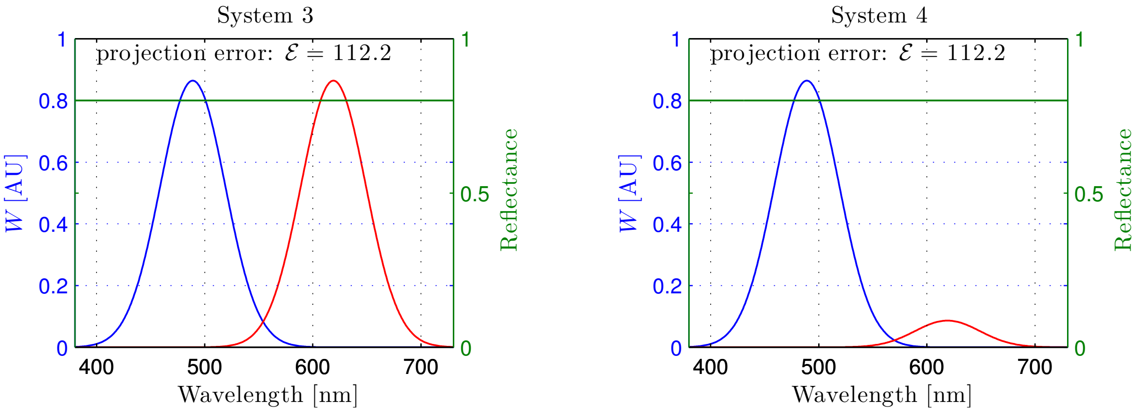


Figure 3.3: Illustration of channel scale invariance property of subspace projection error for two acquisition systems: the channel illustrated in red for System 4 is scaled to 10% of the channel illustrated in red in System 3. Projection errors of an achromatic sample reflectance are equal in both systems.

discussion did not consider potential risks introduced by wrongly using the estimation algorithm (eg. wrong model parameter selection or data over-fitting), which might bias estimation based optimality measures. Further, for a large number of candidate effective responsivities to be evaluated, the computational overhead of camera response simulation, spectral estimation and potential cross-validation might make this approach not feasible in practice.

3.3 A framework for filter selection

We elaborate on the filter selection task, which is the problem of finding a set of filters that, in combination with the spectral responsivity of the camera system and a fixed scene illumination, minimize one or multiple objective functions that quantify the system performance.

Filter selection is a specific form of filter design, which is the more general task of finding (and subsequently manufacturing) filters for multi-spectral systems. Considering a selection from existing filters (or combinations of filters) is, from an economical point of view, more interesting than filter design. A considerable large amount of work has been devoted specifically to filter design [43, 44, 50, 51, 45, 52], as well as filter selection [53, 45, 49, 54, 43, 55, 56, 53, 48] of optical imaging systems.

3.3.1 Brute force selection - an exhaustive approach

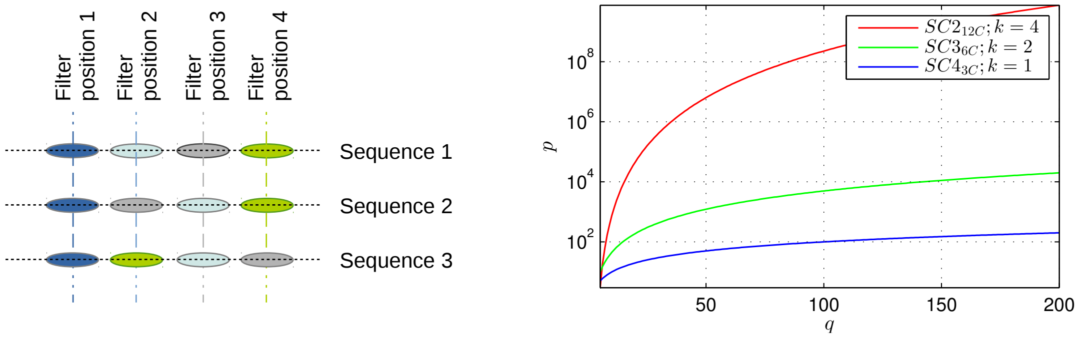
Let $\mathcal{T} = \{\mathbf{t}_1, \mathbf{t}_2, ...\mathbf{t}_q\}$ be the set of q filters with transmittances $\mathbf{t}_i \in \mathbb{R}^{m\times 1}, i = 1, 2, ...q$. The number of elements contained in $\mathcal{T}$ is its cardinality, denoted by $|\mathcal{T}|$. We consider the problem of selecting k filters from $\mathcal{T}$. So for a LSMSC system, k depends on the system configuration. For $\mathbf{SC2_{12C}}$, $k = 4$; for $\mathbf{SC3_{6C}}$, $k = 2$, and for $\mathbf{SC4_{3C}}$, $k = 1$ (see Section 2.5). Further, let $\mathcal{F}_k$ be the set of possible combinations of k filters and $|\mathcal{F}_k| = p$. The cardinality of $\mathcal{F}_k$ is given by the binomial coefficient:

$$p = \binom{q}{k} = \frac{q!}{k!(q-k)!}. \tag{3.14}$$

For $\mathbf{SC2_{12C}}$ (4 additional filters), we might also consider the case of adding only 3 additional filters. Further, for this system the order of placing the filters in front of the sensor matters because combinations of two channels share the same gain controller[6]. Taking this into account, the number of distinct combinations for $\mathbf{SC2_{12C}}$ is

$$p = 3\binom{q}{k+1}. \tag{3.15}$$

We illustrate in Figure 3.4a the only three possible ways of placing 4 filters in front of the sensor, resulting in different systems. Any other sequence of ordering 4 filters is identical to one of the previous three sequences with respect to the optimal gain setting.



(a) **(b)**

Figure 3.4: Possible sequences of ordering 4 filters in front of $\mathbf{SC2_{12C}}$ (a). Number of filter combinations p that are possible for selecting from a pool of q filters for systems with varying number of additional filters (b).

The relation of p, q and k is shown in Figure 3.4b. As an example, to select 2 filters from $q = 100$ for $\mathbf{SC3_{6C}}$, we need to evaluate $p > 4000$ filter combinations. For $\mathbf{SC2_{12C}}$ and $q = 100$, more than 200 million combinations are possible.

3.3.2 Trading of completeness for speed - a heuristic approach

Depending on the system configuration, exhaustive search for filter selection might be time consuming or even unfeasible. Heuristic techniques trade off either optimality of the optimization problem solution with respect to the error function, or completeness with respect to evaluating all candidate solutions for speed. The heuristic scheme followed in this work trades off completeness for speed and is a two-step process.

We note that the heuristic rules presented in the following are specific for the application, the cardinality of the initial filter set and the type of system considered. While the principle of the heuristic scheme can be considered general, the actual implementation might differ for other cases from the ones considered in this dissertation.

A Filter pre-selection

The first step is a filter pre-selection to reduce the cardinality of the initial set $\mathcal{T}_0$ containing q_0 filters to a smaller set $\mathcal{T}_1$ containing q_1 filters. Set $\mathcal{T}_0$ can be reduced by imposing

---

[6] The gain is tuned such that the dynamic range of the system is used in an optimal way and hence, the order of placing the filters is important. See Section 2.3.3.B.

constraints on spectral properties of the filter transmittance functions. Accordingly, we remove filters for which any of the following statements is true:

- the average transmittance of the filter is less than 40% and includes no value larger than 60%.
- the filter transmittance is mostly flat (having a standard deviation of transmittance values smaller than 0.05) and average transmittance larger than 90%.

The first constraint removes filter functions with an overall low transmittance while retaining those with low average transmittance but some spectral band with high transmittance. The second constraint removes spectrally flat filters with an overall high transmittance. Such filters are not useful, as they mainly scale the effective responsivity of the system. A similar effect as using such a filter can be achieved by adjusting channel gains or the exposure time of LSMSC.

If these pre-selection criteria do not reduce the number of filters enough with respect to the resulting number of filter combinations that have to be evaluated, an additional filter reduction scheme can be employed. This reduction is achieved by performing k-means clustering [57, p.424] with q_1 clusters on the filter transmittance data. A reduced set of filters is then the q_1 filters with minimum Euclidean distance to the cluster centers.

B System plausibility check

The constraints on spectral properties of filter transmittances can be extended to the physical properties of the resulting responsivity of the camera system. This is a reasonable step because some filters might alter the effective responsivity of the camera system such that the information loss in image acquisition is too great to consider the system further. An example is given in Figure 3.5 on the left, where a bandpass filter with cut-off wavelength 530 nm (indicated by a solid red line) is applied to a 3-channel system (indicated by blue lines). It can be seen that the resulting system (illustrated on the right) is almost not sensitive below 500 nm. If color measurement is considered without taking assumptions about the type of colors to be measured, the missing information in the suppressed spectral range inevitably leads to large colorimetric error.

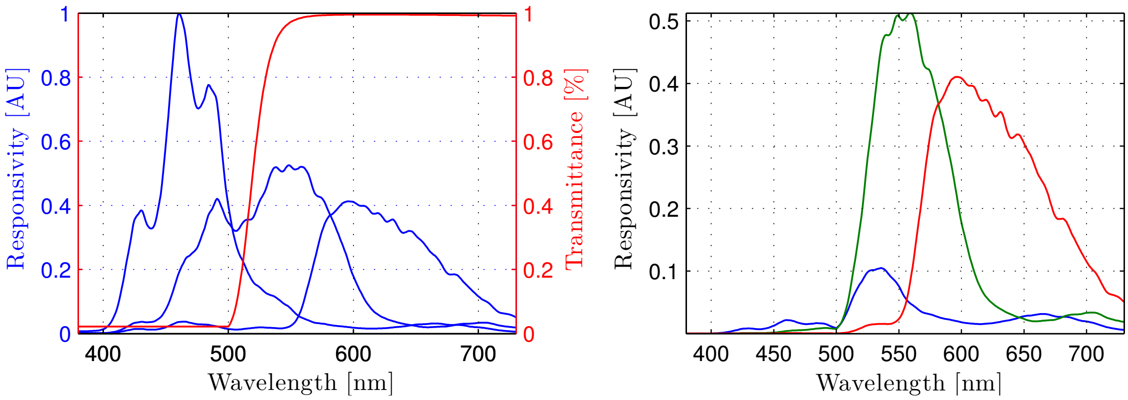


Figure 3.5: System plausibility check for heuristic filter selection: combining a filter transmittance (red line) with an RGB sensor responsivity (blue lines) results in the system on the right with low responsivity below 500 nm.

The system plausibility constraints are imposed on the effective responsivities of the system after applying the filters that passed the pre-selection in set $\mathcal{T}_1$. For each channel of a candidate system, we compute the centroid wavelength. If a spectral curve of the effective system responsivity is considered as an area, the centroid wavelength is defined as the wavelength that corresponds to the centroid of that area. Filters that correspond to effective responsivities for which any of the following statements is true are removed:

- the responsivity is less than 40% at all wavelengths and all channels.
- the integrated effective responsivity divided by the integrated raw sensor responsivity is less than 20% for all channels.
- the centroid wavelengths of all channels withing each RGB sub-image are less than 25 nm apart from each other.
- the centroid wavelengths of all R, G or B channels from all sub-images are less than 10 nm apart from each other.

The first two conditions are trivial. The third condition can be understood as rejecting those candidate systems for which the information within any RGB sub-image is redundant (ie. the R, G and B channel responsivities corresponding to a particular sub-image are rather similar). The last condition only affects systems with more than one lens ($k > 1$), for instance the 6- or 12-channel system configurations. Again, candidate systems that acquire redundant information are removed. This time, redundancy between all channels of RGB sub-images is considered.

Resulting from this step is the final set of q_2 filters, which we denote as $\mathcal{T}_2$. Following the heuristics, we obtain a reduced set $\mathcal{F}_{k,H}$ of filter combinations that are subsequently evaluated for optimality.

3.3.3 Selection of Pareto optimal filters

We consider the filter selection task as a multi-objective optimization problem [58]

$$\arg\min_{f_i} \left(\mathcal{E}_1(f_i), \mathcal{E}_2(f_i), ...\mathcal{E}_d(f_i)\right), \tag{3.16}$$

where $\mathcal{E}_j(f_i)$ denotes the j-th objective function measuring system optimality of filter combination $f_i \in \mathcal{F}_{k,H}, i = 1, 2, ...p$. Let $\mathcal{F}_P$ denote a set of so called Pareto optimal filter combinations, for which there is no other filter combination $f \in \mathcal{F} \backslash \mathcal{F}_P$ that satisfies

$$\mathcal{E}_j(f) \leq \mathcal{E}_j(\hat{f}), \text{ for all } j = 1, 2, ...d \text{ and any } \hat{f} \in \mathcal{F}_P. \tag{3.17}$$

Accordingly, a Pareto optimal solution is a solution for which minimizing any of the objective functions further would lead to an increase in at least one of the other objective functions. We illustrate this concept in Figure 3.6 for the case of a bivariate optimization problem with objective functions $\mathcal{E}_1$ and $\mathcal{E}_2$. Gray points are from $\mathcal{F}_{k,H}$ and red points from $\mathcal{F}_P$.

The solution of the multi-objective filter optimization considered shall be understood as computing the Pareto set $\mathcal{F}_P$ from $\mathcal{F}_{k,H}$. The final selection of a preferred filter combination from $\mathcal{F}_P$ can be achieved by a decision maker, which is either a human expert, or a function composed of a set of rules defining optimality, described by a human expert. As we will see later, the Pareto set is in practice often much smaller as compared with the initial set of candidate solutions, which makes it feasible for a human expert to be a decision maker for this task.

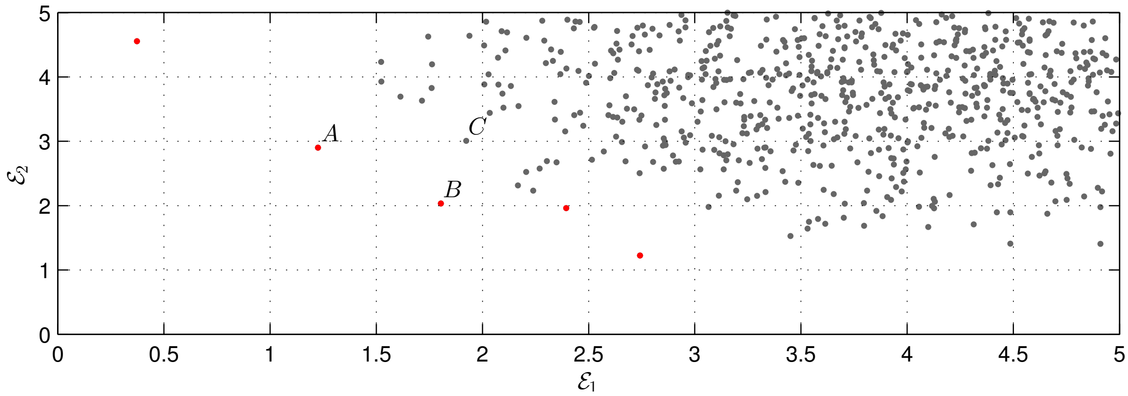


Figure 3.6: Illustration of Pareto optimality principle: A dominates B when $\mathcal{E}_1$ is considered ($\mathcal{E}_1(A) < \mathcal{E}_1(B)$), but is dominated by B for $\mathcal{E}_2$ ($\mathcal{E}_2(A) > \mathcal{E}_2(B)$). C is dominated by A and B ($\mathcal{E}_1(A) < \mathcal{E}_1(C)$ & $\mathcal{E}_1(B) < \mathcal{E}_1(C)$ & $\mathcal{E}_2(A) < \mathcal{E}_2(C)$ & $\mathcal{E}_2(B) < \mathcal{E}_2(C)$). Point C is therefore not Pareto optimal.

3.4 Experimental case studies

Because there exist various possible application cases for LSMSC systems, we demonstrate filter selection for different application scenarios:

Case 1: 3-channel system for colorimetric measurements

Case 2: 6-channel system for spectral reflectance factor measurements

Case 3: 12-channel system for spectral reflectance factor measurements

For all three cases, we used spectral reflectance estimation error based metrics as a measure of system optimality in order to account for application constraints in the filter selection. The filter selection for case 3 was considered for practical implementation and the corresponding experiment is therefore closest to a real application scenario.

As mentioned before, we considered linear least squares regression with linear feature mapping as spectral estimation method, which will be discussed in a later chapter (Section 5.3.1.A). Compared with more advanced estimation methods, this algorithm does not have parameters and performs considerably well in practice, as will be shown in Chapter 5. Whenever colorimetric error measures are considered in this chapter, ΔE_{76} is used instead of the CIE recommended ΔE_{00}, because of its faster computation. Preliminary experiments on smaller sets of filter combinations have shown that this compromise did not affect the final filter selection.

The spectral reflectance dataset used for computing system optimality consists of 2000 samples from a Pantone chart, an Inkjet printed IT8 chart and an HKS offset print color chart[7]. Because it is not feasible to fabricate all filter combinations and acquire corresponding camera response data, responses required for the computations were simulated in accordance with the camera response model in Section 2.3.3.B, **Noise Scenario 2** (ie. a realistic level of simulated camera response noise). The filters used for the selection are off-the-shelf Schott color glass filters. The transmittance of glass filters is affected by additive compounds that absorb light of some wavelengths while transmitting the other

---

[7] We note that this dataset is different from the datasets defined in Section 2.5, because the filter selection was performed at an earlier stage of the work than the collection of data for the other datasets.

part of the light. Glass filters can be manufactured with different thickness, which also influences the spectral transmittance of the filter[8].

We retrieved two sets of 60 filters from the manufacturer's web-page[9] corresponding to 0.5 mm and 1 mm glass thickness respectively. The filter transmittances are illustrated in Figure 3.7. Filter stack transmittances were computed by the dot product of any combination of two out of the 120 initial filters. We also considered the unstacked filters and hence, 7260 filter transmittances are contained in the initial filter set $\mathcal{T}_0$ in total. Unlike in other places of this work, we considered 5 nm sampling of the spectral data instead of 10 nm sampling. While 10 nm sampling is adequate for color calculus or most applications that consider spectral reflectance data (which are generally rather smooth), it is not adequate for filter selection because the responsivities of the camera sensor in combination with additional filters might be rather steep curves. We did not evaluate the optimal sampling rate for spectral data in the filter selection task, but merely consider 5 nm as a trade-off between accuracy and computational cost when considering finer sampling.

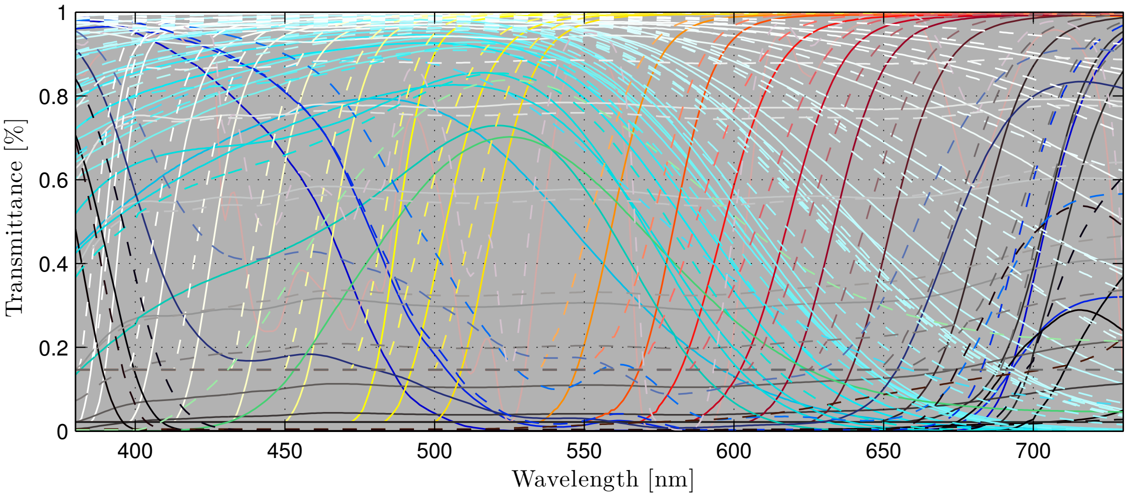


Figure 3.7: Illustration of two sets of color glass filter transmittances: the solid lines correspond to 1 mm glass thickness and the dashed lines correspond to 0.5 mm glass thickness.

For each of the three application cases, we show the selection of a filter combination that is optimal in Pareto sense with respect to a set of predefined system optimality criteria. In order to illustrate that the filter selections were not influenced by the dataset considered in the selection process, we perform an independent evaluation on dataset **DS1**. This subsequent analysis is based on 10-fold cross-validation and we measured system performance in terms of logarithmic root-mean-square error $RMSLE$[10] and ΔE_{76}

---

[8] In fact, the deviation of viewing angle when imaging an object through a color filter also influences the filter transmittance. This is because the distance that light travels through the filter increases with larger deviation angles with respect to the filter surface normal. The effect is not modeled in camera response simulations, as we assume a viewing angle of 0 ° (see Section 2.3.3).

[9] Reading from the web: `http://www.schott.com/advanced_optics/english/filter/index.html`; retrieved: 01.03.2014.

[10] As mentioned in Section 2.4.1.A, $RMSLE$ weights spectral residuals with low reflectance value higher than corresponding residuals of the same magnitude with higher value. When color measurement is based on spectral measurement, this metric is useful because low spectral residuals often contribute to large colorimetric errors and should therefore be weighted higher.

between estimated and measured reflectance data.

From a computational point of view, filter selection in accordance with our framework can be parallelized to a high degree and hence be solved more rapidly using parallel computing. In fact, we will see in the next subsection that for the 6- and 12-channel system configurations, the number of filter combinations to evaluate is in the scale of several millions. Because of this, we implemented the filter selection framework in GNU Octave[11] programming language, using the Message Passing Interface (MPI) extension[12] to allow distributing the workload in parallel processes. The computations were then performed on the high performance computing system *Alhambra UGR*, which comprises 1808 cores and has a computational power of approximately 40 Terra FLOPS[13]. Accordingly, filter selection for the 12-channel system $\mathbf{SC2_{12C}}$ was performed in less than 24 hours. On a conventional personal computer, it was approximated that months would have been required for the same task.

3.4.1 Heuristic reduction of filter combinations to evaluate

Filter pre-selection performed on the aforementioned set of 7260 filters in $\mathcal{T}_0$ resulted in a reduced set $\mathcal{T}_1$ with $q_1 = 4658$ elements. For the 12-channel configuration, the corresponding number of possible filter combinations is $p_1 = 5.4702197 \times 10^{16}$ (see Equation 3.15), which is clearly still too large to be considered in practice. We therefore employed the k-means based filter reduction scheme and set $q_1 = 150$, which resulted in a reasonable low p_1.

We illustrate a summary of the heuristic filter combination reduction achieved for the case of the 3-, 6- and 12-channel system configurations considered in case studies in Table 3.2.

Table 3.2: Heuristic reduction of the number of possible filter combinations for the case of 3-, 6- and 12-channel system configurations: $q_0 = |\mathcal{T}_0|$ is the cardinality of the initial filter set, $q_1 = |\mathcal{T}_1|$ is the number of filters after pre-selection, p_1 denotes the number of filter combinations to evaluate before system plausibility check and p_2 is the final number of combinations to evaluate for system optimality.

System	$\mathbf{q_0}$	$\mathbf{q_1}$	$\mathbf{p_1}$	$\mathbf{p_2}$
3-channel	7260	4658	4658	3526
6-channel	7260	4658	10846153	5893577
12-channel	7260	150	20823198	7049365

3.4.2 Filter selection for a 3-channel color measurement system

The system design considered in this experiment is for a 3-channel system corresponding to $\mathbf{SC4_{3C}}$. Accordingly, we considered the problem of selecting an additional filter that is placed in front of the RGB sensor and optimizing the system performance with respect to the selected optimality criterion.

---

[11] Reading from the web: `https://www.gnu.org/software/octave/`; retrieved: 14.04.2015

[12] Reading from the web: `http://octave.sourceforge.net/mpi/`; retrieved: 14.04.2015

[13] Reading from the web: `https://alhambra.ugr.es`; retrieved 14.04.2015. The author thanks the Centro de Servicios de Informática y Redes de Comunicaciones (CSIRC), Universidad de Granada, for providing the computing time.

One measure for system optimality for a color measurement system is colorimetric estimation error and as mentioned before, we used ΔE_{76} as such. We selected median error (Mdn. ΔE_{76}) and maximum error (Max. ΔE_{76}) as optimality criteria and considered filter selection as a bivariate optimization. We therefore computed the Pareto optimal subset of filters from the $p_2 = 3526$ possible filter combinations that remained after heuristic pre-selection. For illustration purposes, median as well as maximum errors for all p_2 filter combinations as well as the Pareto optimal subset are shown in Figure 3.8.

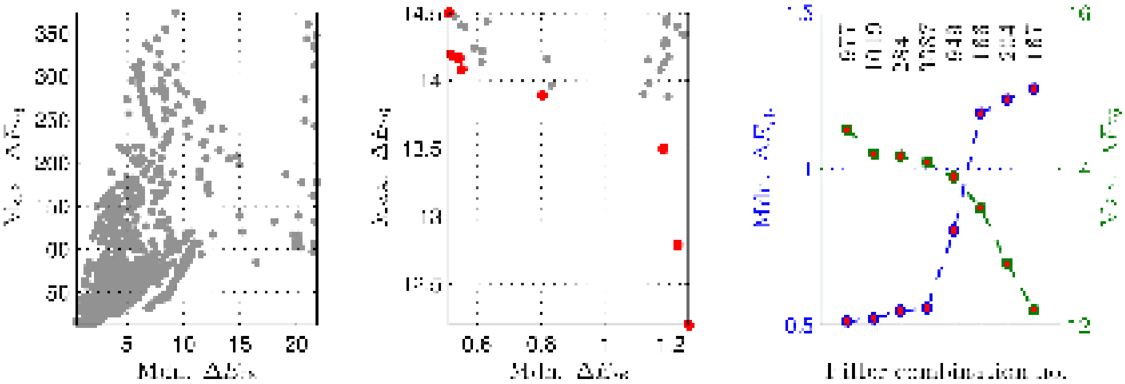


Figure 3.8: Median and maximum ΔE_{76} error of the q_2 filter combinations evaluated for the design of $\mathbf{SC4_{3C}}$. Illustrated in red are the Pareto optimal filter combinations. The plot in the middle is a zoomed version of the plot on the left. In the right plot, Pareto optimal combinations are shown in order of increasing median ΔE_{76} error.

We have selected filter combination 3387 from the Pareto set, which is a trade-off between reasonably low median and maximum colorimetric error (Mdn. $\Delta E_{76} = 0.55$, Max. $\Delta E_{76} = 14.08$). This filter combination is a stack of a 1 mm thick BG63 and a 1 mm thick GG420 filter glass. The transmittances of these filters and the corresponding filter stack as well as the resulting effective responsivity of the 3-channel camera system in this configuration are shown in Figure 3.9.

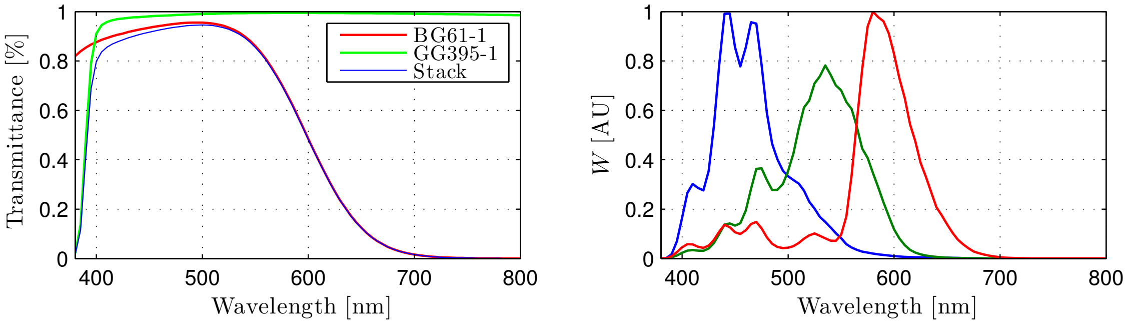


Figure 3.9: Left: Filter stack of combination 3387. Right: Effective system responsivity of $\mathbf{SC4_{3C}}$.

A Performance evaluation via simulations

We have compared the performance of the 3-channel system with the additional filter with the same 3-channel system without the additional filter. We use boxplots[14] to summarize

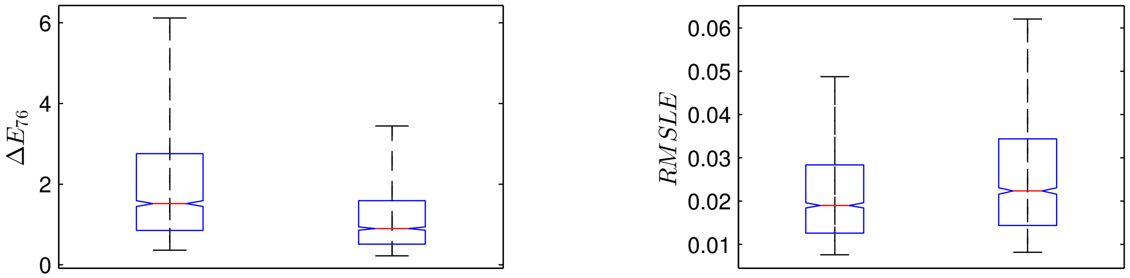
---

[14] See Appendix A.1 for boxplot conventions.

the numerical results of the system performance evaluation. Boxplots are used rather than tabulated data because they are convenient tools for comparing the systems visually. The corresponding illustrations are shown in Figure 3.10 and we summarize the findings as follows: Considering the colorimetric error (ΔE_{76}), it can be observed that the system with additional filter results in a lower median, mean, maximum and minimum error. Further, the non-overlapping notches of the boxes offer evidence of a statistical significant difference (on the 5% significance level) between the medians. The evaluated spectral error ($RMSLE$) shows an overall decrease in performance. This does not come to much surprise, as spectral error was not considered as optimality criterion in the filter selection of this system.

We extended our analysis: according to [3], a system is said to be colorimetric, if the effective system spectral responsivity is a linear combination of the CMFs for a particular observation condition. We show normalized responsivity plots of the two systems in Figure 3.11, in addition to the spectral properties of the standard observer weighted by the standard illuminant. Neither of the two systems shows a perfect resemblance. However, the red channel of the system with additional filter is less sensitive from 600 nm to 800 nm and hence seems more similar to the corresponding color matching function. We confirmed this observation by expressing system optimality in terms of subspace projection error and considered the data independent measure proposed by Vora *et al.* [42], denoted in Equation 3.8. Accordingly, we found $\upsilon = 0.9353$ for the system without additional filter and $\upsilon = 0.9386$ for the system with additional filter. Values closer to 1 indicate higher resemblance with the weighted CMFs.

In summary, we can conclude that within the scope of limitations due to considering only simulations in the experiment, the system with additional filter offers a clear improvement in color measurement applications as compared with the system without additional filter.



Without filter	**With filter**
Mean: 2.11	Mean: 1.24
Min: 0.03	Min: 0.03
Max: 11.91	Max: 8.07

Without filter	**With filter**
Mean: 0.022	Mean: 0.027
Min: 0.003	Min: 0.003
Max: 0.098	Max: 0.128

Figure 3.10: Box plot diagram of numerical results from spectral estimation experiment for the 3-channel configuration $\mathbf{SC4_{3C}}$ (left: ΔE_{76}, right: $RMSLE$).

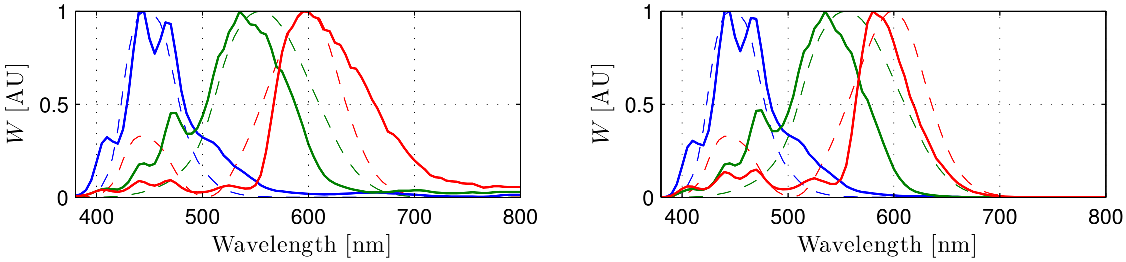


Figure 3.11: Normalized effective system responsivity of $\mathbf{SC4_{3C}}$ without additional filter (left) and with filter stack combination 3387 (right). The additional dashed black lines correspond to the normalized standard observer weighted by the standard illuminant used for color measurements.

3.4.3 Filter selection for a 6-channel system

We also considered the design of a 6-channel system corresponding to $\mathbf{SC3_{6C}}$. One aim of this experiment was to illustrate the big influence of the selection of optimality criterion on the overall system performance. We therefore considered two design options:

Option 1: system for spectral reflectance measurements using median RMSE (Mdn. $RMSE$), maximum RMSE (Max. $RMSE$) and median Pearson distance (Mdn. d_p) as optimality criteria. The filter selection problem can hence be considered as a trivariate optimization.

Option 2: system for colorimetric measurements using median colorimetric error (Mdn. ΔE_{76}) and maximum colorimetric error (Max. ΔE_{76}) as optimality criteria. The filter selection problem can be considered as a bivariate optimization[15].

Even though the number of filter combinations that were evaluated for this system is large ($p_2 = 5893577$), the set of Pareto optimal filters is quite small (8 filter combinations for Option 1 and 39 filter combinations for Option 2). It is often true in practice that the cardinality of the Pareto set decreases with increasing number of objective functions considered in the optimization.

Like in the previous example, we illustrate the Pareto sets for Option 1 and 2 in form of figures (Figure 3.12 and Figure 3.13). Our choice of optimal filter combinations were the ones with number 3209253 for Option 1 and number 6476237 for Option 2. The transmittances of these filter stacks as well as the resulting effective responsivities of the two systems are shown in Figure 3.14 and Figure 3.15.

A Performance evaluation via simulations

The corresponding boxplots for Option 1 and Option 2 are shown in Figure 3.16. We can see that Option 1 results in a significantly higher spectral estimation performance as compared with Option 2, and the opposite for colorimetric error (ΔE_{76}). These results were expected but there is another important conclusion that can be drawn from this: To consider either only color or only spectral metrics as optimality criteria in filter selection is critical, because high performance as measured by one metric might result in low performance by the other.

---

[15] Note that the filters selected for Option 2 are considered for $\mathbf{SC3_{6C}}$ in later parts of this dissertation.

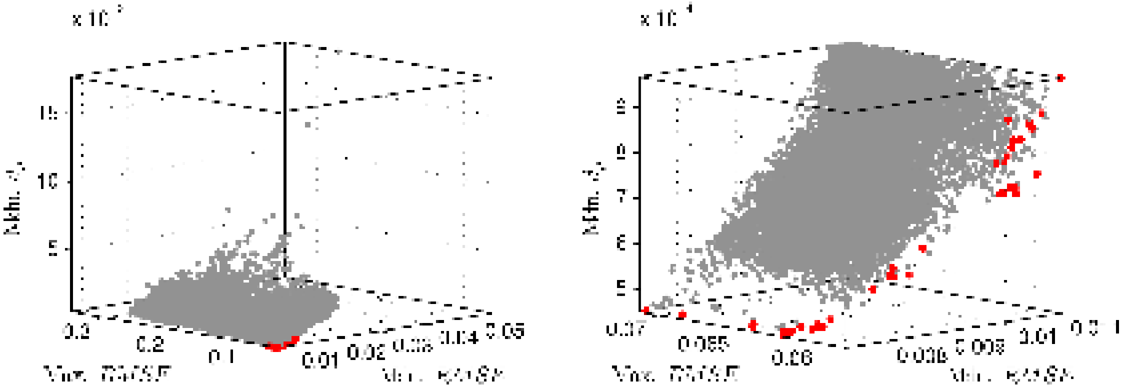


Figure 3.12: Median and maximum $RMSE$ and median d_p of the p_2 filter combinations evaluated for the design of $\mathbf{SC3_{6C}}$. Illustrated in red are the Pareto optimal filter combinations. The plot on the right is a zoomed version of the plot on the left.

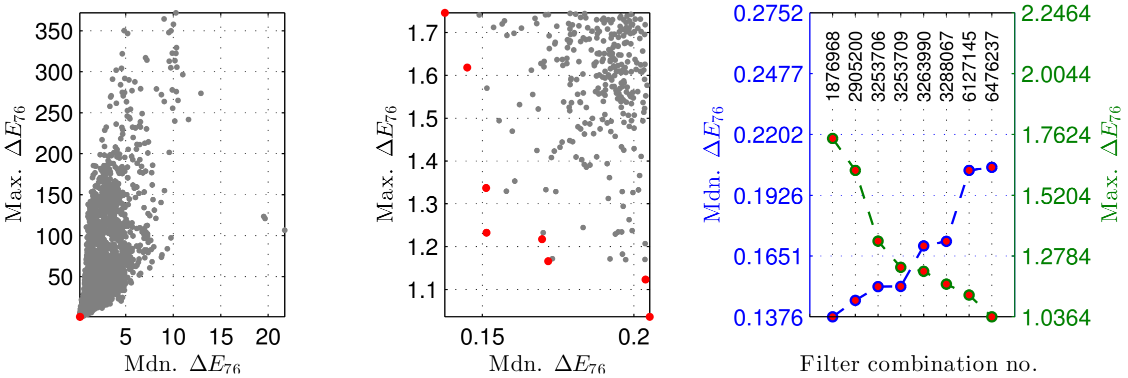


Figure 3.13: Median and maximum ΔE_{76} error of the p_2 filter combinations evaluated for the design of $\mathbf{SC3_{6C}}$. Illustrated in red are the Pareto optimal filter combinations. The plot in the middle is a zoomed version of the plot on the left. In the right plot, Pareto optimal combinations are shown in order of increasing median ΔE_{76} error.

3.4.4 Filter selection for a 12-channel system for spectral reflectance measurements

This subsection relates to the design of a 12-channel system corresponding to $\mathbf{SC2_{12C}}$. For the design of a spectral reflectance measurement system, optimality criteria should of course relate to spectral measurement performance. Because we want to use estimated spectral data from this system for color measurement as well, it is advisable to consider a color metric as was shown in the previous subsection. So we chose $RMSLE$ and ΔE_{76} as optimality criteria. The 95-percentiles (p_{95} $RMSLE$ and p_{95} ΔE_{76}) were chosen as a descriptive measure for performance to reduce the strong influence of outliers on median or mean error. For this system, we are rather interested in good performance over the vast majority of data considered.

The number of possible filter combinations that had to be evaluated was approximately 7 million, which is considerably larger than in the 3- and 6-channel case. The corresponding Pareto set is illustrated in Figure 3.17. From the set of 12 Pareto optimal filter stack combinations, number 2358396 was selected. This combination is a trade-off between reasonably low colorimetric and spectral error. The resulting spectral responsivity of the

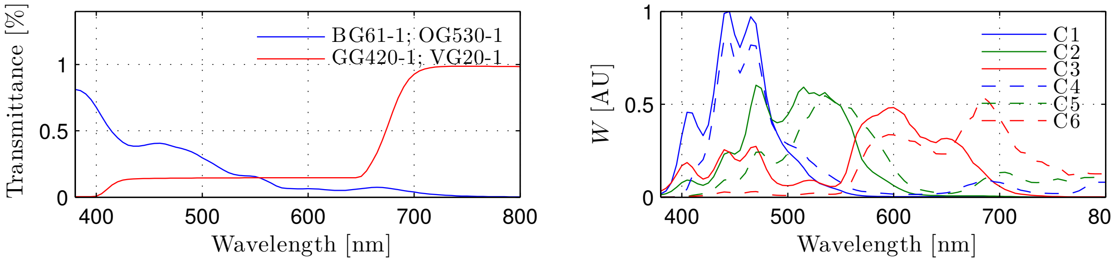


Figure 3.14: Left: Filter stack of combination 3209253. Right: Effective system responsivity of $\mathbf{SC3_{6C}}$ Option 1.

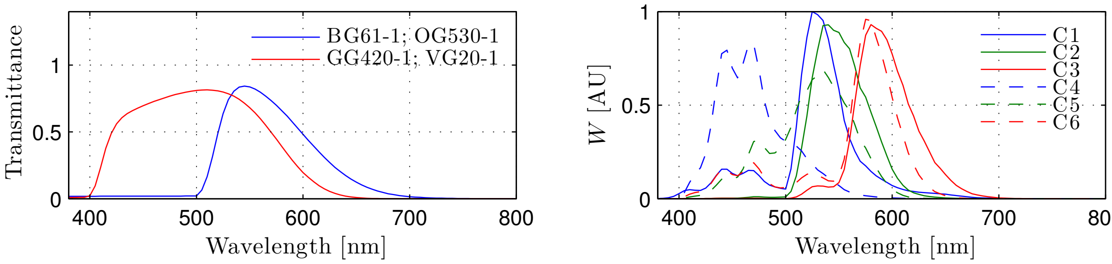


Figure 3.15: Left: Filter stack of combination 6476237. Right: Effective system responsivity of $\mathbf{SC3_{6C}}$ Option 2.

system is shown in Figure 3.18.

A Comparison between manufactured and manufacturer specified filter transmittances

The aforementioned filter stack combination for $\mathbf{SC2_{12C}}$ has been considered for practical implementation. The selected combinations of four filters (each is a stack of two filters) were produced according to our specifications by filter manufacturer *bk Interferenzoptik Elektroik GmbH*. Each filter stack was additionally coated with an anti-reflection layer on both sides. We evaluated the accordance of the simulated filter stack transmittances[16] and measured filter transmittances.

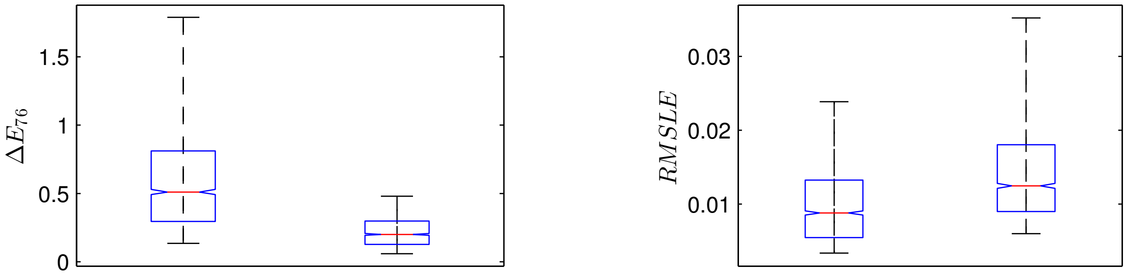
The transmittance measurement was performed as follows: we used a Konica Minolta CS2000[17] spectroradiometer to measure the spectrum of visible light from an incandescent light source transmitted through each filter stack at 0° with respect to the surface normal and divided this measurement by the SPD of this source measured with the same device. Instead of aiming at the source directly, we used an integrating sphere between source and detector. Each measurement was repeated 3 times and the resulting data was averaged.

The spectral data measured by the CS2000 is in the range from 380 nm to 780 nm in steps of 5 nm, and we illustrate the transmittances provided by Schott GmbH in the same range. The corresponding transmittance plots are shown in the first row in Figure 3.19. Illustrated in red are the simulated filter stack transmittances and in blue the measured filter transmittances. Residuals between the two are shown in the 2nd row of the same

---

[16] Recall that filter stack transmittances were simulated computationally by computing the dot product of two filter transmittances specified by the manufacturer.

[17] Reading from the web: `http://sensing.konicaminolta.asia/products/cs-2000-spectroradiometer/`; retrieved: 11.04.2015.



Option 1	**Option 2**	**Option 1**	**Option 2**
Mean: 0.66	Mean: 0.23	Mean: 0.010	Mean: 0.016
Min: 0.03	Min: 0.01	Min: 0.002	Min: 0.002
Max: 4.63	Max: 1.58	Max: 0.049	Max: 0.119

Figure 3.16: Box plot diagram of numerical results from spectral estimation experiment for the 6-channel configuration $\mathbf{SC3_{6C}}$ (left: ΔE_{76}, right: $RMSLE$).

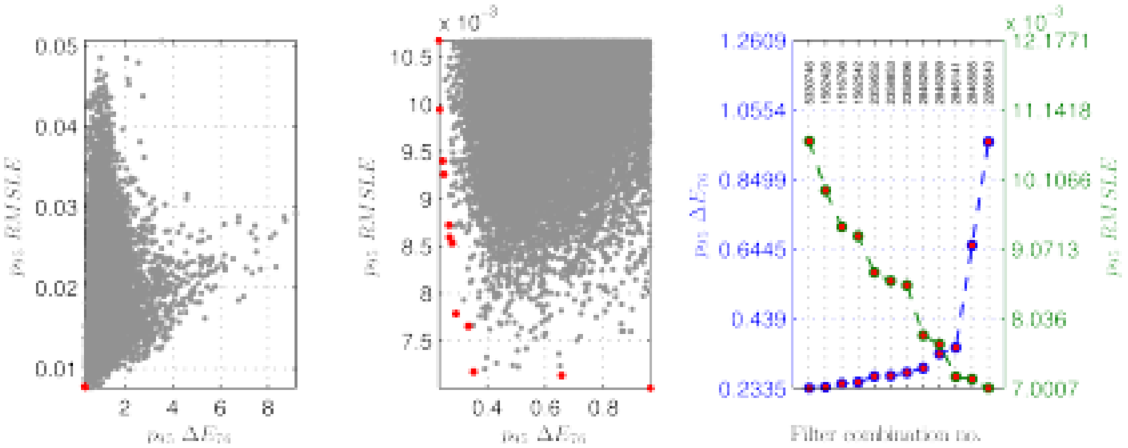


Figure 3.17: 95-percentile for ΔE_{76} and $RMSLE$ error of filter combinations evaluated for the design of $\mathbf{SC2_{12C}}$. Illustrated in red are the Pareto optimal filter combinations. The plot on the right illustrates the Pareto optimal combinations in order of increasing p_{95} ΔE_{76} error.

figure. These residuals are due to fabrication error and to some extend measurement error of the transmittance measurements.

The plots motivate some discussions: the overall shape of the measured and simulated transmittance curves are quite similar for the four filter stacks. The largest deviations are found for the stack of filters GG435 (0.5mm) and BG25 (1mm) and large deviations are mostly at the lower and higher end of the spectral range evaluated. The spikes in residual plots indicate a wavelength shift between filter transmittances. It has to be noted that the height of the spike does not reflect the wavelength shift directly, but the slope of the transmittance curve around the location of the peak. The implications of these findings are discussed within the next subsection.

B Performance evaluation via simulations

We compared the 12-channel system design $\mathbf{SC2_{12C}}$ with $\mathbf{SC1_{12C}}$, which is the reference configuration of our laboratory LSMSC for which filter selection was initially performed by the camera manufacturer. We found that the aforementioned differences between sim-

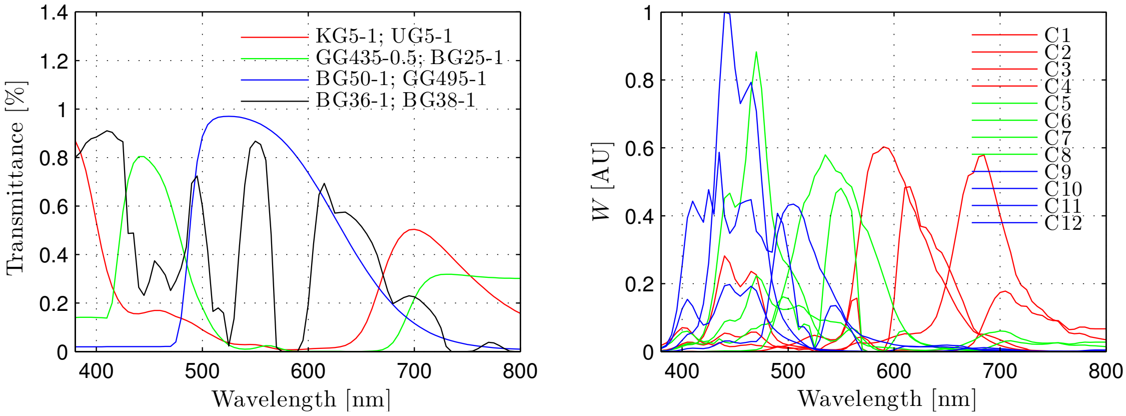


Figure 3.18: Left: Four filter stacks of combination 2358396. Right: Effective system responsivity for $\mathbf{SC2_{12C}}$.

ulated and measured filter transmittances do not impose a significant influence on the overall estimation performance. Therefore, we only considered simulations with filter transmittances derived from manufacturer specification.

A summary of the numerical results is shown in Figure 3.20. The average performance of both 12-channel systems is in general much higher than that of the 3- or 6-channel system evaluated in the previous subsections. The comparison of the two 12-channel systems also reveals a significant improvement of $\mathbf{SC2_{12C}}$ over $\mathbf{SC1_{12C}}$. Colorimetrically, $\mathbf{SC2_{12C}}$ results in a lower median and mean ΔE_{76} error and the minimum color difference is equal for both systems. The maximum color difference is by a factor of more than 2 lower for $\mathbf{SC2_{12C}}$. Considering $RMSLE$, there is only a slight improvement for $\mathbf{SC2_{12C}}$ over $\mathbf{SC1_{12C}}$. Yet, the median $RMSLE$ of $\mathbf{SC2_{12C}}$ is significantly lower than that of $\mathbf{SC1_{12C}}$.

C Performance evaluation via measurements

The conditions of this experiment were similar to the previous one, except that measured camera responses were considered instead of simulations. We therefore installed the filter stacks in front of the LSMSC and adjusted the channel gains such that the system is operated optimally with respect to SNR and utilization of the camera dynamic range[18].

The results of this experiment are again illustrated in form of boxplots (see Figure 3.21). We additionally report estimation performance numerically in Table 3.3. This table includes first order statistics for various metrics (ΔE_{76}, ΔE_{00}, $RMSE$, $RMSLE$, d_p) and also corresponding results for the previous experiment with simulated camera response data for reference.

When comparing the boxplots in Figure 3.21 with the ones corresponding to simulated camera response data, we immediately see that the difference in estimation performance between $\mathbf{SC1_{12C}}$ and $\mathbf{SC2_{12C}}$ is smaller when measured camera response data is considered. This is a strong indication that the camera response model considered does not generalize well to the real acquisition systems. We will come back to this issue in Section 5.5.3 when analyzing experimental results related to the spectral estimation problem, and then also discuss potential sources for the performance discrepancy when considering simulated as compared with measured camera response data. We explicitly note that this aspect illustrates a system limitation and not a limitation of the filter selection framework.

---

[18] See Section 2.3.3.B

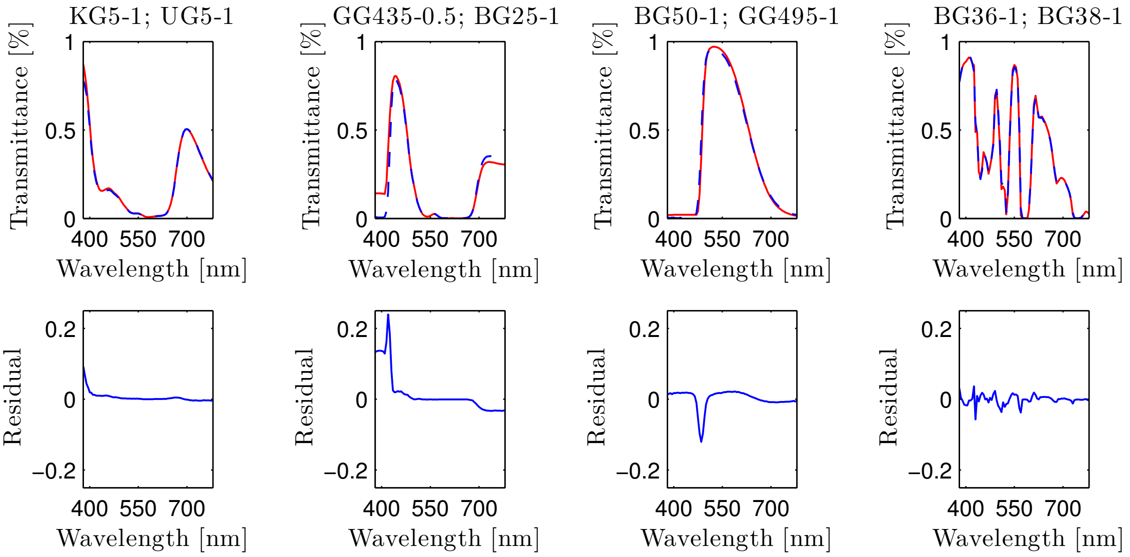
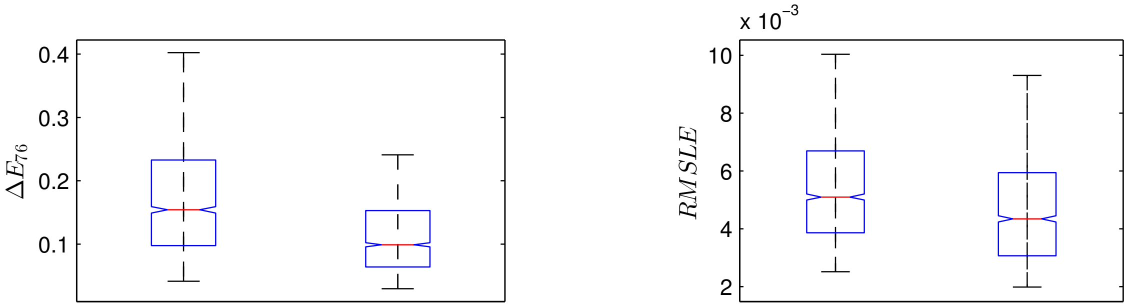


Figure 3.19: Comparison of $\mathbf{SC2_{12C}}$ filter stack transmittances: illustrated in red are the simulated filter stack transmittances and in blue the measured filter transmittances (upper row). The lower row illustrates residuals between the two.

Regardless of the aforementioned, $\mathbf{SC2_{12C}}$ is still found to generally outperform $\mathbf{SC1_{12C}}$ with respect to all colorimetric and spectral metrics evaluated. When considering average estimation error, the performance increase is little. However, with respect to maximum errors, a clear improvement can be seen. As a reference, we illustrate in Figure 3.22 the estimated and measured spectral reflectances corresponding to the maximum $RMSE$ and ΔE_{00}.

The system evaluation based on measured camera response data allows us to draw first conclusions about the applicability of the system for color and spectral reflectance measurements. In terms of colorimetric error (ΔE_{76} and ΔE_{00}), average results are below 1, which indicates remarkable performance. The low average spectral estimation errors confirm the high performance of the LSMSC.



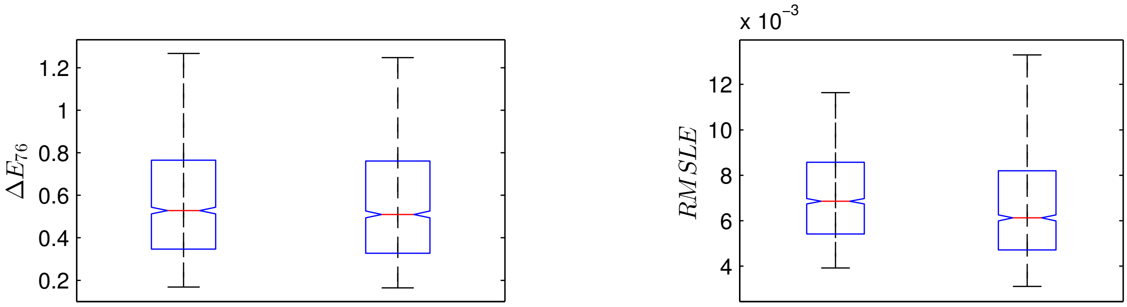
$SC1_{12C}$
Mean: 0.18
Min: 0.01
Max: 1.14

$SC2_{12C}$
Mean: 0.11
Min: 0.01
Max: 0.48

$SC1_{12C}$
Mean: 0.006
Min: 0.001
Max: 0.027

$SC2_{12C}$
Mean: 0.005
Min: 0.001
Max: 0.028

Figure 3.20: Box plot diagram of numerical results from spectral estimation experiment with simulated camera response data for $\mathbf{SC1_{12C}}$ as compared with $\mathbf{SC2_{12C}}$ (left: ΔE_{76}, right: $RMSLE$).



$SC1_{12C}$
Mean: 0.60
Min: 0.05
Max: 3.46

$SC2_{12C}$
Mean: 0.58
Min: 0.03
Max: 2.38

$SC1_{12C}$
Mean: 0.007
Min: 0.002
Max: 0.030

$SC2_{12C}$
Mean: 0.007
Min: 0.002
Max: 0.024

Figure 3.21: Box plot diagram of numerical results from spectral estimation experiment with measured camera response data for $\mathbf{SC1_{12C}}$ as compared with $\mathbf{SC2_{12C}}$ (left: ΔE_{76}, right: $RMSLE$).

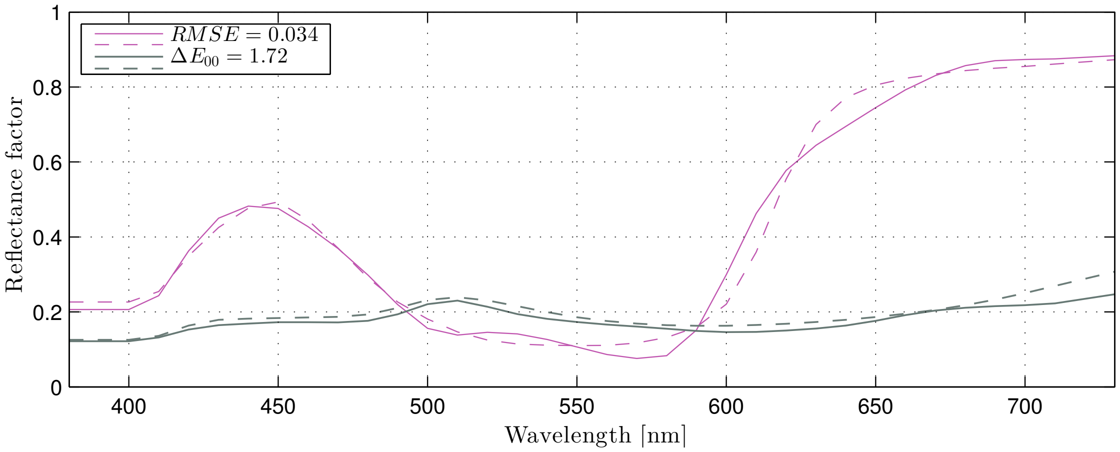


Figure 3.22: Estimated reflectances (solid lines) and measured reflectances (dashed lines) corresponding to the sample with largest $RMSE$ and the sample with largest ΔE_{00} error.

Table 3.3: First order statistics of numerical performance evaluation of $\mathbf{SC1_{12C}}$ and $\mathbf{SC2_{12C}}$, corresponding to the experiment in Section 3.4.4.

(a) Simulated camera response data.

	$\mathbf{SC1_{12C}}$				$\mathbf{SC2_{12C}}$			
	Avg.	**Std.**	**Min.**	**Max.**	**Avg.**	**Std.**	**Min.**	**Max.**
ΔE_{76}	0.18	0.12	**0.01**	1.14	**0.11**	**0.07**	**0.01**	**0.48**
ΔE_{00}	0.12	0.08	**0.01**	0.68	**0.08**	**0.06**	**0.01**	**0.37**
$RMSE$	**0.007**	**0.004**	**0.001**	**0.036**	**0.007**	**0.004**	**0.001**	0.038
$RMSLE$	0.006	**0.003**	**0.001**	**0.027**	**0.005**	**0.003**	**0.001**	0.028
d_p	**0.0003**	**0.0003**	0.0001	**0.0027**	**0.0003**	**0.0003**	**0.0001**	0.0032

(b) Measured camera response data.

	$\mathbf{SC1_{12C}}$				$\mathbf{SC2_{12C}}$			
	Avg.	**Std.**	**Min.**	**Max.**	**Avg.**	**Std.**	**Min.**	**Max.**
ΔE_{76}	0.60	0.37	0.04	3.46	**0.59**	**0.36**	**0.02**	**2.45**
ΔE_{00}	0.42	0.26	0.03	1.81	**0.41**	**0.25**	**0.01**	**1.72**
$RMSE$	0.010	**0.004**	**0.002**	0.041	**0.009**	0.005	**0.002**	**0.036**
$RMSLE$	**0.007**	**0.003**	**0.002**	0.031	**0.007**	**0.003**	**0.002**	**0.025**
d_p	0.0006	**0.0007**	**0.0001**	0.0068	**0.0005**	0.0008	**0.0001**	**0.0062**

3.5 Summary, conclusions and future work

In this dissertation, the design of LSMSC optical component's spectral properties is regarded as an optimization problem. We defined various system optimality criteria, subject to which this optimization problem can be approached. This view of the problem is not new as such. However, the conceptual unification of various methods in related literature within the same mathematical framework and the direct comparison between them by means of illustrative examples can be considered as a novel contributions.

We then narrowed down the scope of LSMSC system design to the filter selection problem. Within this context, we proposed a general framework for solving the task based on a systematic evaluation of possible filter combinations from a set of existing filters. We illustrated that the computational demand for this exhaustive evaluation is increasing drastically for LSMSC system configurations with more than one lens (ie. more than 3 camera channels). This motivated the development of a two-stage heuristic scheme that can be used to reduce the number of filter combinations that have to be evaluated. The heuristics trade off completeness of the evaluation for speed. The heuristic rules that we defined in the experimental part of this chapter were designed such that filter combinations are removed if a low spectral and color measurement performance of a system with these filters is foreseeable.

To systematically evaluate filter combinations, we need to choose adequate system optimality criteria. In practice, it is often desirable to consider various criteria and the strategy followed by other researchers often involves a specific weighting scheme to combine corresponding error metrics to one measure. Defining such a scheme in a meaningful way is not trivial. We proposed a different approach to overcome this limitation by regarding the filter selection task as a multi-objective optimization problem using the Pareto optimality principle. This requires us to evaluate all objective functions that are to be considered individually for every filter combination. The solution to the multi-objective optimization is then considered to be among the set of Pareto optimal filter combinations. This set includes all filter combinations that are optimal in the sense that minimizing any of the individual objective functions further would decrease at least one of the others. The Pareto set often contains only few elements and the selection of an adequate filter combination from this reduced set is much simpler than considering all filter combinations.

We have exemplified filter selection according to our framework for three system configurations, namely $\mathbf{SC4_{3C}}$, $\mathbf{SC3_{6C}}$ and $\mathbf{SC2_{12C}}$. One intention for performing case studies for various systems was to demonstrate the selection of distinct system optimality criteria, which is in practice always application specific.

Filter selection of $\mathbf{SC4_{3C}}$ and $\mathbf{SC3_{6C}}$, and consecutive evaluation of the system performance were solely based on simulated camera response data. For $\mathbf{SC4_{3C}}$, the filter selection is performed to optimize device independent color measurement. We demonstrated that for the particular 3-channel system considered, the color measurement performance could be improved markedly by placing an additional color filter in front of the camera. However, spectral reflectance measurement performance is compromised in consequence.

Filter selection for $\mathbf{SC3_{6C}}$ was performed for both, optimizing spectral and color measurement. The corresponding criteria for system optimality were selected accordingly. By analyzing the performance of the resulting systems, we showed a potential risk involved when considering metrics from a single application domain. For the systems considered, performance was either high for color measurement and specifically low for spectral measurement, or vice versa. In most practical applications of multi-spectral imaging, a trade-off between these conditions is desirable, which is another argument supporting our intention to consider the filter selection task as a multi-objective optimization problem.

Unlike for the other system configurations, we considered practical implementation of the optimal filters for $\mathbf{SC2_{12C}}$, which included an opto-mechanical changeover of our laboratory LSMSC to the new filter configuration. This procedure allowed the analysis of system performance based on measured camera response data and the comparison with the initial filter configuration by the camera manufacturer, which we referred to as $\mathbf{SC1_{12C}}$.

Our findings show a moderate improvement of average spectral and colorimetric measurement performance for $\mathbf{SC2_{12C}}$ over $\mathbf{SC1_{12C}}$, and a significant improvement in terms of a reduction of the maximum errors. We found average colorimetric measurement performance below $\Delta E_{00} = 0.6$ and average spectral measurement performance below $RMSE = 0.01$. The high measurement performance is a strong indication of the efficiency of our filter selection framework. However, we also found a considerable discrepancy between performance when simulated camera response data is used and performance for measured camera responses[19]. This indicates a shortcoming of the image acquisition model, which in turn imposes a severe bias to the filter selection.

An adequate image acquisition model is the basis of filter selection, and we take the aforementioned shortcoming as a motivation for future work. Besides, we have mentioned earlier that illumination optimization is an equally important task in LSMSC design as compared with filter selection. For instance, our laboratory LSMSC system contains a light source that incorporates various types of LEDs. Tuning currents for each LED type individually results in a modulation of the scene illumination SPD. If we consider the spectral mixing of LED SPDs as a function of LED currents, it should be possible to approximate an optimal current setting for an existing LSMSC system. Alternatively, spectrally tuning the light source can be considered as an additional degree of freedom in the filter selection process.

---

[19] An extended discussion of potential reasons for this discrepancy follows in Section 5.5.3.

4

Multi-channel line-scan image registration

Contents

4.1 Preface

When discussing data processing of the LSMSC system in Section 2.3.2.C, we referred to the registration process required to spatially align image channels. Proper image registration means that pixels with the same spatial image coordinates in all channels correspond to the same physical location in the scanning scene. It is a vitally important data processing step because the quality of image alignment is often directly linked to the performance of the acquisition system. Specifically in multi-spectral imaging systems, channel misalignment can reduce the spectral estimation performance at image regions with inhomogeneous spatial content. Besides: RGB images rendered from this channel misaligned data contain color fringes at image locations with sharp edges [28, 59]. For the LSMSC considered in this work, the number of channels to register depends on the system configuration. We focus on $\mathbf{SC2_{12C}}$, so we deal with registration of 12 image channels from filtered RGB sub-images corresponding to the four lenses of the LSMSC. Registration of other LSMSC system configurations can be achieved by following a similar procedure as proposed below.

4.1.1 Planar and non-planar scanning objects

When a planar scanning object is considered, LSMSC image registration can be achieved by geometrical calibration. This is the process of establishing a registration model in a fitting procedure that is based on data extracted from a calibration target image cube. Once established, the registration model can be applied to any other image to correct channel misalignment. The geometrical calibration has to be redone only in case the mechanical setup of the system is modified, or if scanning objects with a different height geometry are considered.

For the case of a non-planar scanning object and a LSMSC with more than one lens, geometrical calibration alone is insufficient to register images. Non-planar scanning objects have a three-dimensional surface structure. Hence, the distance with respect to the camera sensor is variable, resulting in additional scene-dependent image channel misalignment. This problem is illustrated in Figure 4.1. However, because any physical location in the image scene projects to a unique pair of image locations in sub-image pairs of different lenses, this limitation can be overcome by an additional registration step based on correspondence estimation. The method we use to find pixel correspondence consists of a novel color invariant feature mapping and subsequent block matching.

The block diagram in Figure 4.2 illustrates the general framework for image registration of LSMSC image data of planar and non-planar scanning objects. The gray deposited boxes with dashed lines indicate the additional processing steps for non-planar scanning objects.

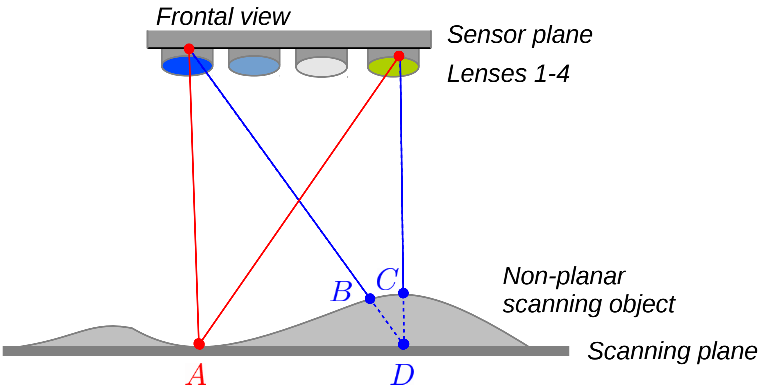


Figure 4.1: Cause of channel misalignment for non-planar scanning objects with $\mathbf{SC2_{12C}}$: in a geometrically calibrated system, scanning object point A which is void of height with respect to the scanning plane maps to the same pixel location in all 12 channels of the image cube. For object point C as observed through lens 4, the virtual point where pixel locations match in all image channels is D, but what we actually observe in image channels corresponding to lens 1 is point B.

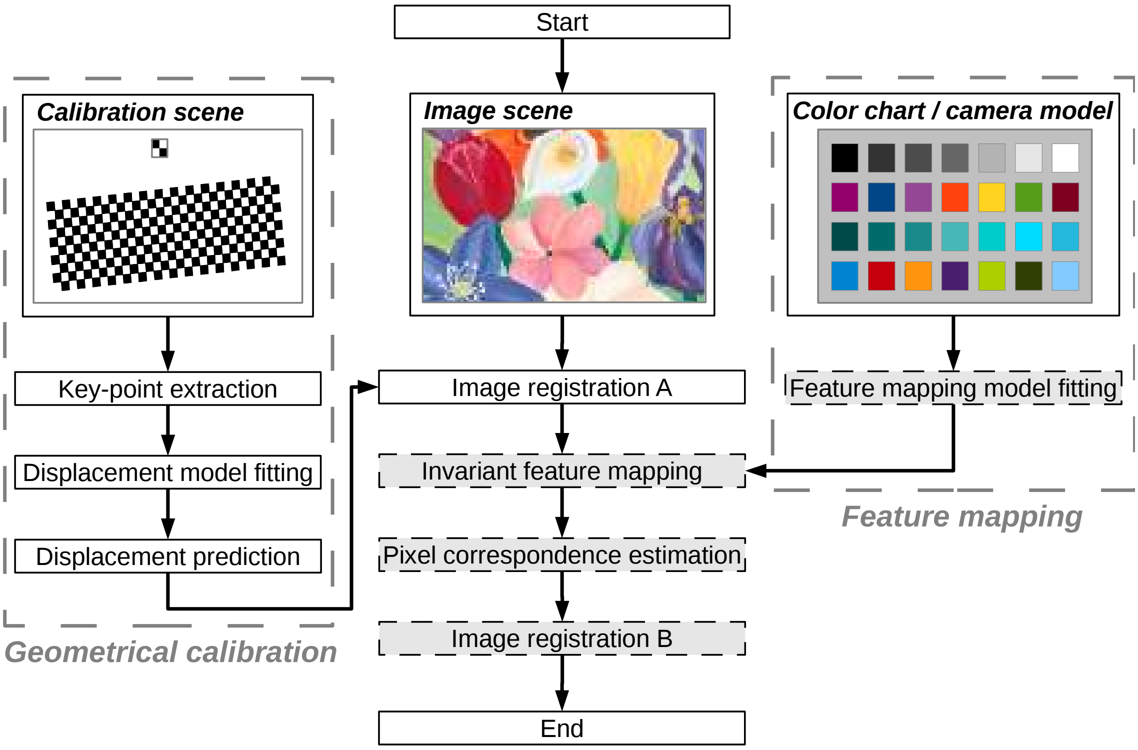


Figure 4.2: Schematic illustration of LSMSC image registration.

4.1.2 Formal definition of multi-channel image registration

Let $\mathcal{C}$ be an unregistered image cube that contains a sequence of image data matrices $\{I_\lambda \in \mathbb{R}^{r\times c} | \lambda = 1, 2, ...n\}$ of an image scene captured by a multi-spectral imaging device. Scalar r is the number of rows, c the number of columns and n is the number of image channels. The domain of image coordinates is $\Omega = \{(x, y) \in \mathbb{Z}^2 | 1 \leq x \leq c, 1 \leq y \leq r\}$. Further, let $\mathcal{C}_0$ be a virtual reference cube $\{I_{\lambda,0} \in \mathbb{R}^{r\times c} | \lambda = 1, 2, ...n\}$ for which the image planes are spatially aligned and

$$I_\lambda \left(x + z_{\lambda,u}(x, y), y + z_{\lambda,v}(x, y)\right) = I_{\lambda,0}(x, y), \tag{4.1}$$

where functions $z_{\lambda,u}(x, y)$ and $z_{\lambda,v}(x, y)$ describe the horizontal and vertical misalignment at any pixel location $(x, y) \in \Omega$ for image channel λ. In that sense, multi-channel image registration is the process of aligning each image I_λ in the multi-spectral image cube $\mathcal{C}$ spatially to its corresponding image $I_{\lambda,0}$ in the reference cube [60]. We note that in the general case, misalignments in horizontal and vertical direction in an image are functions of two variables, namely the image pixel location x and y respectively.

The remainder of this chapter is structured as follows: In Section 4.2, we introduce the geometrical calibration process and describe the two displacement curve fitting models that were considered. After that, we introduce two methods of color invariant feature mapping, followed by a detailed description of the pixel correspondence estimation for non-planar scanning objects in Section 4.3. We evaluate each of the aforementioned processes experimentally in Section 4.4 and summarize this chapter in Section 4.5.

4.2 Geometrical calibration

In the specific case of line-sensor image data and planar scanning surfaces, image misalignment only depends on the pixel location along the sensor line (horizontal image direction), while being constant in scanning direction. This simplifies the registration process because consequently, horizontal and vertical misalignment are only functions of one variable, namely the sensor pixel location. We note that this does not imply that there exists no misalignment in the vertical direction of the image.

We denote $\Pi = \{x \in \mathbb{Z} | 1 \leq x \leq c\}$ as the domain of sensor pixel locations. The image registration model is then

$$I_\lambda \left(x + z_{\lambda,u}(x), y + z_{\lambda,v}(x)\right) = I_{\lambda,0}(x, y), \tag{4.2}$$

where functions $z_{\lambda,u}(x)$ and $z_{\lambda,v}(x)$ describe the misalignment at any sensor pixel location $x \in \Pi$ along the sensor line for image channel λ.

Because the displacement functions can be considered independently for each image channel, we omit subscript λ until Section 4.3, unless otherwise specified. Accordingly, we denote reference image I_0 as the one to which all other channels are to be aligned spatially, and arbitrarily select $I_0 = I_1$. This definition implies that potential image distortion in I_1 is not corrected[1].

We consider geometrical calibration as the problem of finding the horizontal and vertical displacement functions required for image registration of planar scanning surfaces (ie. $z_u(x)$ and $z_v(x)$). In practice, a calibration target is used to obtain corresponding

---

[1] We note that in practice, the distortion along the sensor line direction is much smaller than other misalignment errors.

locations of a set of discrete points (key-points) in a calibration image from all image channels. These points are then used to compute the channel-wise displacement functions in a fitting process.

4.2.1 Design of a geometrical calibration target

The calibration target used in this work contains a regularly spaced checkerboard pattern. Key-points are defined as the locations where black checker patterns intersect. There are various aspects to consider in the development of a geometrical calibration target. We discuss the most important ones in the following list and define the target that we use later on:

Key-point correspondence over all spectral image channels:

Because each image channel of the multi-spectral camera has distinct spectral response characteristics, it has to be ensured that the target is visible in each image channel to allow extracting corresponding key-points. For the spectral sensitive range of the camera sensor considered in this work (380 nm to 730 nm), black toner from a laser printer printed on white paper substrate is suitable to produce a black and white checkerboard pattern that is visible in all image channels.

Key-point coverage over spatial image domain:

The non-rigid registration approaches discussed in this work are based on curve fitting. A dense set of key-points spread over the entire sensor domain is required to capture potential local displacement variations. A dense sampling can be achieved by using a checkerboard pattern that is rotated. Not rotating the pattern would lead to a large amount of key-point locations with similar x coordinate that are redundant for displacement correction of a line sensor.

We determine the rotation angle such that the initial spacing between black checker patches with side length d is divided equally by the number of key-points extracted from each column of the checker pattern. It follows that for a checkerboard with g rows, the rotation angle α can be computed as

$$\alpha = sin^{-1}\left(\frac{1}{g-1}\right). \tag{4.3}$$

In practice, the checkerboard width is selected such that it spans the entire sensor domain after rotation. For our system, this can be achieved by constructing a checkerboard with $h = 30$ columns, $g = 11$ rows and patches of side length $d = 1\,\mathrm{cm}$. In this configuration, $\alpha = 5.74\,^{\circ}$ and the resulting key-point spacing in horizontal direction is $s = d\sin(\alpha) = 0.1\,\mathrm{cm}$. These geometrical relations are illustrated in Figure 4.3.

As mentioned before, the rotated target can be printed on white paper using a laser printer. A corresponding image cube can be acquired by aligning the printed sheet with the scan line.

4.2.2 Key-point extraction

Let $\mathcal{K}$ be a set of k key-points with correspondence over n images in cube $\mathcal{C}$. The locations of the key-points are $\{(x_i, y_i) \in \mathbb{R}^2 | 1 \leq x_i \leq c, 1 \leq y_i \leq r, i = 1, 2, ...k\}$. The key-point locations in the reference image are denoted as $(x_{0,i}, y_{0,i})$. Key-point extraction is the process of obtaining $\mathcal{K}$. Note that pixel locations in an image are integer values over the

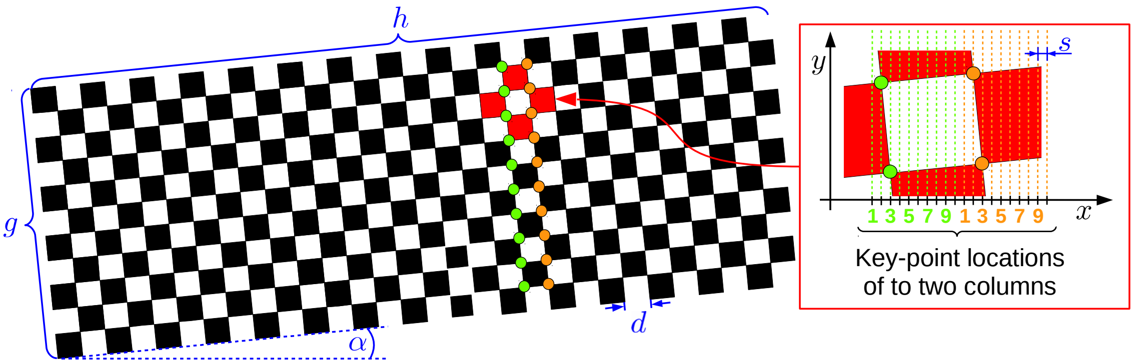


Figure 4.3: Geometrical relation of number of rows g and rotation angle α for the design of a calibration and test target using a checkerboard pattern. The red box illustrates the resulting horizontal key-point spacing s for a checkerboard pattern with g rows.

image domain Ω. Key-point coordinates on the other hand can be extracted in sub-pixel accuracy from image data and therefore can be real numbers.

There are many approaches to obtain key-points from multi-channel image data and automatic detection schemes exist in several application domains. The problem is generally well-studied and has application for instance in stereo image matching, motion tracking and fusion of multi-modal image data. A great overview of local invariant feature detection is given in [61]. The aforementioned applications are distinct from our problem, because key-points are extracted from image data with unknown image content. In geometrical calibration, we consider extracting key-points from the geometrical calibration target, which was designed to allow robust and automated key-point extraction. This can be achieved by using a two-step corner detection scheme:

Step 1: locate the minimum size quadrangular region enclosing all checker patches. Span a regular grid to estimate initial locations where the corners of intersecting black squares would be located if the imaging system would obey rectilinear projection. To do this, the number of patches in horizontal and vertical direction have to be known.

Step 2: detect the real location of the corners in a search window of size $\frac{2}{3}d \times \frac{2}{3}d$ around the initial estimate. The scalar d is the side-length in pixel units of one square checker pattern.

In practice, the imaging system does not obey rectilinear projection. Hence, the initial estimated locations are not accurate. Finding the exact locations of the corners is achieved with an algorithm based on Harris-corner finder. We use an implementation by Jean-Yves Bouguet, included in the *Camera Calibration Toolbox for Matlab* [62].

4.2.3 Numerical analysis of image channel misalignment

Multi-channel image data of the calibration target obtained with our laboratory LSMSC system in combination with the key-point extraction framework allow us to perform a numerical analysis of the channel misalignment inherent to the image data. Specifically, we can analyze the channel-wise horizontal and vertical displacement vectors $\mathbf{z}_u$ and $\mathbf{z}_v$, computed from k key-point locations in $\mathcal{K}$ as

$$\begin{aligned} \mathbf{z}_u &= [x_i - x_{0,i}]_{k\times 1},\ i = 1,2,...k \\ \mathbf{z}_v &= [y_i - y_{0,i}]_{k\times 1},\ i = 1,2,...k. \end{aligned} \tag{4.4}$$

We performed this analysis to get a better understanding about the nature of image displacement of our system and illustrate the corresponding vectors in Figure 4.4. Note that we use subscripts to denote the image channel for each displacement function, so $\mathbf{z}_{u,\lambda}$ and $\mathbf{z}_{v,\lambda}$ are horizontal and vertical displacement vectors for image channel λ and $\lambda = 1,2,...n$. Several observations can be made:

- Channels with scene content projected over the same lens (ie. channels from the same sub-image) show a similar displacement pattern with respect to the reference channel. The displacement pattern is distinct for each lens, which can be explained by lens distortion and viewpoint differences for individual lenses[2].
- The displacement between the R, G and B channel from the same lens show a similar pattern. However, some deviation can be observed. This can also be explained by lens distortion and the fact that the R, G and B channels are projected over different locations on the lens, and therefore undergo distinct distortions.

The most common form of lens distortion, barrel distortion, results in an approximate polynomial distortion pattern when only a line image is considered, as is the case for the LSMSC. Barrel distortion is caused by a decrease in image magnification at points off the optical axis [63]. If the image line is projected over the optical axis, only horizontal distortion occurs. If the projection is shifted vertically, the distortion consists of a vertical and horizontal component. These two types of distortion are illustrated in Figure 4.5.

It is important to note that the illustration of key-point displacement in Figure 4.4 does not directly reflect the displacement caused by lens distortion, as the displacement is computed with respect to a reference channel, which is also distorted. In a previous work related to hyper-spectral image registration, we have computed theoretical key-point coordinates in the reference image that would result from (undistorted) rectilinear projection and therefore allow image registration and distortion correction at once [60]. Obtaining theoretical locations of key-points in sub-pixel accuracy from image data of the LSMSC is much more difficult and not considered in this work.

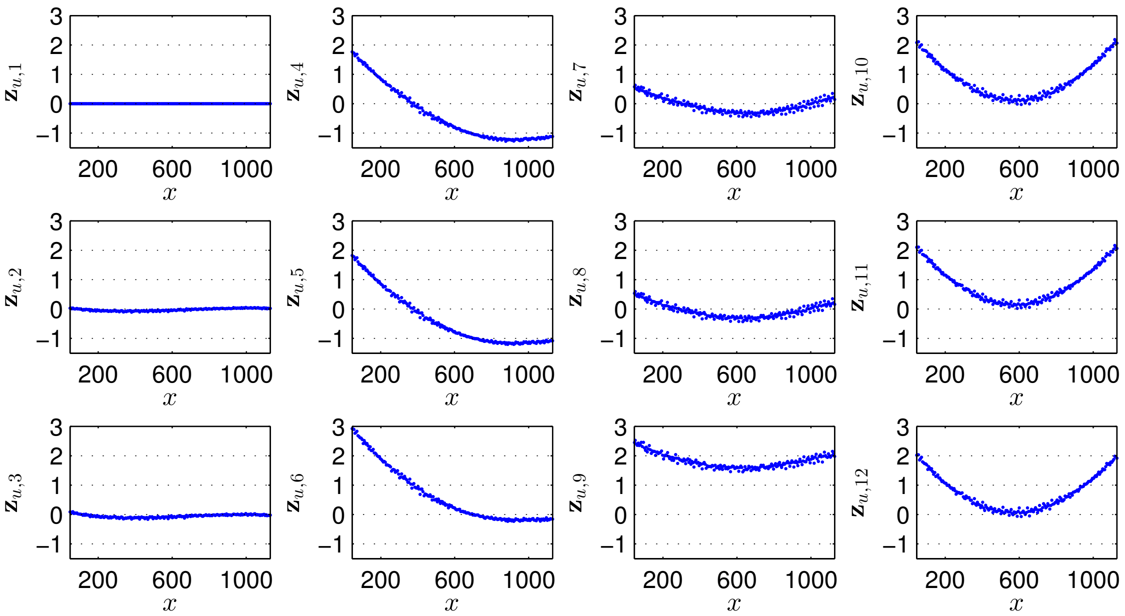
4.2.4 Polynomial curve model

The numerical analysis of displacement vectors performed in the previous subsection marks a good moment to return our focus to the initial geometrical calibration task. We noted that this task contains a curve fitting problem, specifically the problem of fitting image displacement functions to observation data extracted from an image cube of the geometrical calibration target. The observation data are the displacement vectors defined in Equation 4.4. Based on the analysis of image displacement, a polynomial model seems to be a good choice to fit the displacement vectors[3] $\mathbf{z}_u$ and $\mathbf{z}_v$.

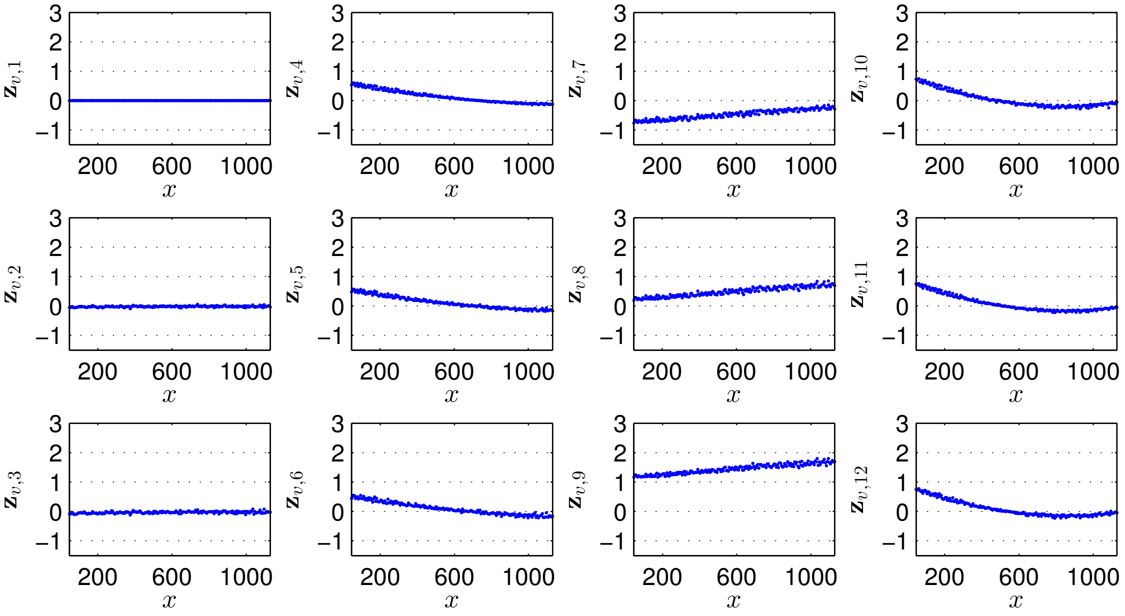
---

[2] See Figure 2.10a to recall the optical principle of the LSMSC system.

[3] We remind the reader that subscript λ, which refers to the channel dependence of this function, is omitted in this context.



(a) Horizontal displacement vectors $\mathbf{z}_{u,\lambda}$ for the 12 image channels of $\mathbf{SC2_{12C}}$.



(b) Vertical displacement vectors $\mathbf{z}_{v,\lambda}$ for the 12 image channels of $\mathbf{SC2_{12C}}$.

Figure 4.4: Illustration of displacement vectors and corresponding key-point x coordinates for image channel $\lambda = 1, 2, ...12$ of our laboratory LSMSC.

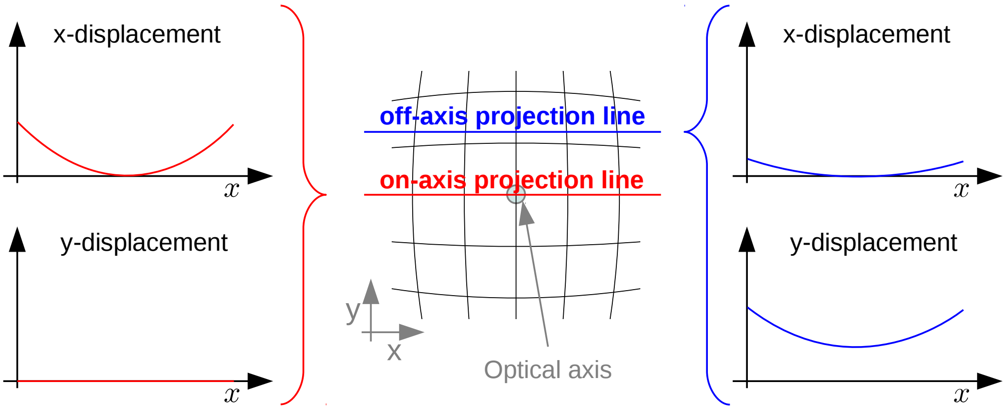


Figure 4.5: Polynomial-type distortion pattern of line-sensor caused by barrel-like lens distortion. The on-axis projection line is distorted only in the x-component, the off-axis projection line has x and y distortion components.

The corresponding model is of form

$$\begin{aligned} \tilde{z}_u(x) &= [1, x, x^2, ...x^{d-1}]\ \mathbf{a}_u \\ \tilde{z}_v(x) &= [1, x, x^2, ...x^{d-1}]\ \mathbf{a}_v, \end{aligned} \tag{4.5}$$

where $x \in \Pi$, scalar d denotes the number of polynomial terms (accordingly, $d-1$ is the degree of the polynomial), $\mathbf{a}_u = [a_{u,0}, a_{u,1}, ...a_{u,d-1}]^T \in \mathbb{R}^{d\times 1}$ and $\mathbf{a}_v = [a_{v,0}, a_{v,1}, ...a_{v,d-1}]^T \in \mathbb{R}^{d\times 1}$ are the vectors of polynomial coefficients and $\tilde{z}_u \in \mathbb{R}^{d\times 1}$ and $\tilde{z}_v \in \mathbb{R}^{d\times 1}$ the corresponding horizontal and vertical estimated displacements at sensor location x.

A Model fitting

To find coefficient vectors $\mathbf{a}_u$ and $\mathbf{a}_v$ from known key-point displacement vectors $\mathbf{z}_u$ and $\mathbf{z}_v$, as well as corresponding sensor locations $\mathbf{x} = [x_1, x_2, ...x_k]^T \in \mathcal{K}$, we have to solve the following minimization problems

$$\begin{aligned} &\arg\min_{\mathbf{a}_u} \|V\mathbf{a}_u - \mathbf{z}_u\| \\ &\arg\min_{\mathbf{a}_v} \|V\mathbf{a}_v - \mathbf{z}_v\|, \end{aligned} \tag{4.6}$$

where $V \in \mathbb{R}^{k\times d}$ is the Vandermonde matrix [46, p. 185] which contains polynomial terms of the elements in $\mathbf{x}$ and is defined as

$$V = \begin{bmatrix} 1 & x_1 & x_1^2 & \dots & x_1^{d-1} \\ 1 & x_2 & x_2^2 & \dots & x_2^{d-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_k & x_k^2 & \dots & x_k^{d-1} \end{bmatrix}. \tag{4.7}$$

The solution to this problem in least square sense [46, p. 230] is unique if $d < k$ and $\text{rank}(V) = d$. It is given by

$$\begin{aligned} \mathbf{a}_u &= (V^T V)^{-1} V^T \mathbf{z}_u \\ \mathbf{a}_v &= (V^T V)^{-1} V^T \mathbf{z}_v. \end{aligned} \tag{4.8}$$

To find displacement values $\tilde{z}_u(x)$ and $\tilde{z}_v(x)$ at arbitrary pixel location $x \in \Pi$, we evaluate Equation 4.5 at x using the fitted polynomial coefficient vectors $\mathbf{a}_u$ and $\mathbf{a}_v$.

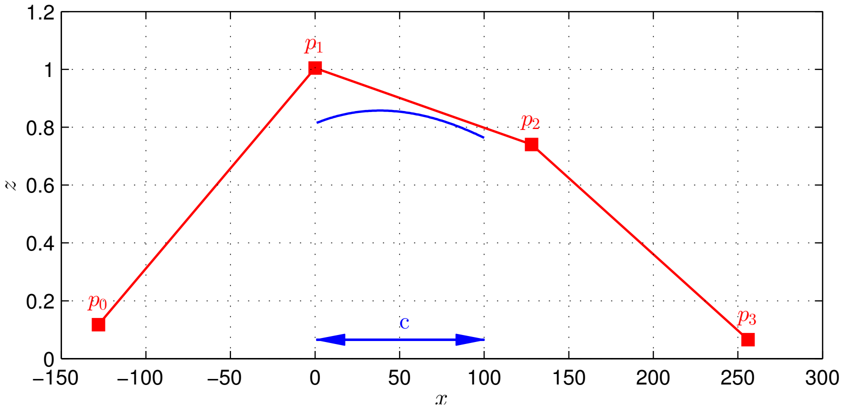


Figure 4.6: An example displacement curve segment (blue), controlled by a set of 4 control points. Note that the domain of control point x-coordinates exceeds the range of the curve segment and resulting from that is that these values can be negative.

4.2.5 Cubic B-spline curve model

If a polynomial function of high degree is used to fit a complex displacement pattern with low fitting error according to the previous model, a problem might occur due to the tendency of such models to oscillate at the extremes of the data domain. In the literature, this effect is referred to as Runge's phenomenon [64]. Piecewise polynomial models of low degree avoid this instability while yielding similar fitting results. In the following, we present a piecewise polynomial model based on cubic B-splines.

A Uniform cubic B-splines curves

The pixel displacement can be described by a parametric curve model of form $Q(s) = (x(s), z(s))$, where s is a curve parameter in the range $[0, 1]$ and z the displacement at pixel position x. Displacement in horizontal and vertical direction are modeled in a similar fashion. Hence, we present only the case of horizontal displacement without the loss of generality.

The sensor domain Π is divided into q segments, which are denoted $\{Q_\mu | \mu = 0, 1, ...q - 1\}$. Each curve segment is controlled by four so-called *control points*, and correspondingly, the entire curve is controlled by $q + 3$ control points $\{\mathbf{p}_i = (x_i, z_i) | x_i = (i - 1)\delta, i = 0, 1, ...q + 2\}$. In uniform cubic B-splines, the control points are spaced uniformly with distance δ and it shall be noted that their x-coordinates exceed the sensor domain. An example curve segment and corresponding control points are illustrated in Figure 4.6.

Each B-spline curve segment Q_μ can be represented as

$$\begin{aligned}\mathbf{Q}_\mu(s) &= \sum_{i=0}^{3} \mathbf{p}_{\mu+i} B_i(s) \\ &= \begin{pmatrix} \sum_{i=0}^{3} x_{\mu+i} B_i(s) \\ \sum_{i=0}^{3} z_{\mu+i} B_i(s) \end{pmatrix},\end{aligned} \tag{4.9}$$

where $\{\mathbf{p}_{\mu+i} = (x_{\mu+i}, z_{\mu+i}) | \mu = \lfloor \frac{x}{\delta} \rfloor, i = 0, 1, 2, 3\}$ are the 4 control points which control the shape of $\mathbf{Q}_\mu$, and $B_i(s)$ is the i-th uniform cubic B-spline basis function evaluated at

s. The basis functions are defined as [65, 66, 67]:

$$\begin{aligned}
&B_0(s) = (1-s)^3/6,\\
&B_1(s) = (3s^3 - 6s^2 + 4)/6,\\
&B_2(s) = (-3s^3 + 3s^2 + 3s + 1)/6,\\
&B_3(s) = s^3/6.
\end{aligned} \tag{4.10}$$

The modeling of the displacement vector as a parametric curve allows to predict displacement at arbitrary pixel location $x \in \Pi$ along the sensor. We can compute this predicted displacement $\widetilde{z}$ as

$$\widetilde{z} = \sum_{i=0}^{3} z_{\mu+i} B_i(s), \tag{4.11}$$

where $z_{\mu+i}$ is the second component of control point $\mathbf{p}_{\mu+i} = (x_{\mu+i}, z_{\mu+i})$ and $s = \frac{x}{\delta} - \lfloor \frac{x}{\delta} \rfloor$, where $\mu = \lfloor \frac{x}{\delta} \rfloor$.

A more compact representation for Equation 4.9 using matrix notation according to Catmull and Clark [68] is:

$$\begin{aligned}
\mathbf{Q}_\mu &= \mathbf{s} M \mathbf{P}_\mu\\
&= \begin{pmatrix} \mathbf{s} M X_\mu \\ \mathbf{s} M Z_\mu \end{pmatrix},
\end{aligned} \tag{4.12}$$

where $\mathbf{s} = [1, s, s^2, s^3]$, $\mathbf{P}_\mu = [\mathbf{p}_{\mu+i}]_{4\times1}$, $X_\mu = [x_{\mu+i}]_{4\times1}$, $Z_\mu = [z_{\mu+i}]_{4\times1}$, $i = 0, 1, 2, 3$ and

$$M = \frac{1}{6} \begin{bmatrix} 1 & 4 & 1 & 0 \\ -3 & 0 & 3 & 0 \\ 3 & -6 & 3 & 0 \\ -1 & 3 & -3 & 1 \end{bmatrix}. \tag{4.13}$$

It can be understood that the predicted displacement $\tilde{z}$ at location x is a weighted combination of the 4 control points in the local neighborhood of x.

An interesting property of cubic B-splines is *local control*, which means that changing one control point does not affect the whole curve but only a local curve segment [66, 69]. This is interesting from a computational point of view, as it allows efficient implementation of multi-level curve refinement with arbitrary spacing. We introduce equidistant multi-level curve refinement below, and arbitrary spacing was not considered in this work.

B Model fitting

The control point spacing determines the ability of the B-spline model to adjust to a complex displacement pattern. The smaller the spacing, the finer the oscillations that can be resolved by the model. This means that this parameter is specific to the complexity of the displacement pattern. By using an iterative multi-level refinement of the grid spacing in the fitting process, the B-spline curve is gradually fitted to the key-points until a termination condition is reached. This refinement process follows the steps depicted in Figure 4.7, which are described below. Note that for consistency reasons, this part of the process description is written in close accordance with our previous work [60].

- **Initialization**
  The initial curve contains one segment $\mathbf{Q}_0$ and is controlled by 4 control points

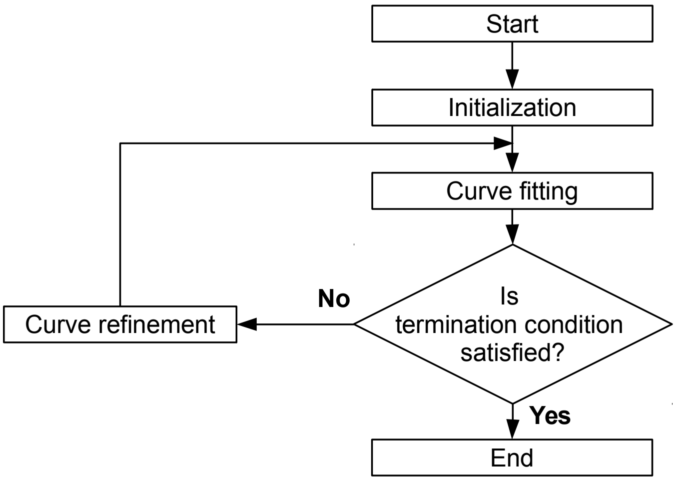


Figure 4.7: Flow-chart of the multi-level curve refinement of the uniform cubic B-spline fitting process, adopted from [60].

with zero displacement component z and spacing $\delta^{(0)} = 2^{\lceil \log_2 c \rceil}$. Its size is at least equivalent to the size of the sensor domain and by definition, we let the domain of the displacement vector be aligned with the left side of $\mathbf{Q}_0$.

- **Termination**
  Curve fitting ends under two conditions: Either a desired fitting accuracy is reached, which means that the residual displacement calculated from Equation 4.4 at any key-point location is smaller than a pre-defined threshold, or the maximum number of refinement steps $\log_2(\delta^{(0)})$ is reached.

- **Curve fitting**
  The objective of curve fitting is to find the z components of the 4 control points of each curve segment $\mathbf{Q}_\mu$. Each control point $\mathbf{p}_i$ ($i = 1, 2, 3, 4$) is influenced by key-points in a local 4δ neighborhood. At iteration l ($l = 1, 2, ...$), the set of these key-points is defined as the proximity set $\mathbf{K}_i^{(l)} = \{(x_c, z_c) \in \mathbf{K} | i - 3 \leq \frac{x_c}{\delta^{(l)}} < i + 1\}$. The residual displacements of key-points in $\mathbf{K}_i^{(l)}$ are used to update the z component of the control point [67]:

$$z_i^{(l)} = \tilde{z}_i^{(l-1)} + \frac{\sum_c W_c^2 \Delta z'_c}{\sum_c W_c^2}, \tag{4.14}$$

  where $W_c = B_a(s)$, $a = i + 1 - \left\lfloor \frac{x_c}{\delta^{(l)}} \right\rfloor$ and $s = \frac{x_c}{\delta^{(l)}} - \left\lfloor \frac{x_c}{\delta^{(l)}} \right\rfloor$, $(x_c, z_c) \in \mathbf{K}_i$. Further, $\Delta z'_c$ is defined as

$$\Delta z'_c = \frac{W_c \Delta z_c}{\sum_{l=0}^{3} B_k(s)^2}, \tag{4.15}$$

  and $\Delta z_c = \widetilde{z_c} - z_c$ is the residual displacement.

- **Curve refinement**
  In curve refinement we increase the number of control points, without altering the shape of the initial curve. This is achieved by dividing the parametric range s at its mid-point and inserting a new control point between each pair of adjacent control points. Accordingly, each curve segment is split into 2 curve segments and the control

point spacing halved in every iteration. The maximum number of refinement steps is therefore equal to $\log_2(\delta^{(0)})$ [70, 68, 71].

The splitting process of $\mathbf{Q}_\mu$ with control points $\mathbf{P}_\mu$ results in new curve segments $\{\mathbf{Q}_\mu^{(1)}, \mathbf{Q}_\mu^{(2)}\}$. The corresponding parametric cubic B-spline curve functions are

$$\begin{aligned}\mathbf{Q}_\mu^{(1)} &= \mathbf{s}M\mathbf{P}_\mu^{(1)} \\ \mathbf{Q}_\mu^{(2)} &= \mathbf{s}M\mathbf{P}_\mu^{(2)}.\end{aligned} \tag{4.16}$$

The new control points $\mathbf{P}_\mu^{(1)}$ and $\mathbf{P}_\mu^{(2)}$ are computed by evaluating

$$\begin{aligned}\mathbf{P}_\mu^{(1)} &= A^{(1)}\mathbf{P}_\mu \\ \mathbf{P}_\mu^{(2)} &= A^{(2)}\mathbf{P}_\mu,\end{aligned} \tag{4.17}$$

given

$$A^{(1)} = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ \frac{1}{8} & \frac{3}{4} & \frac{1}{8} & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & \frac{1}{8} & \frac{3}{4} & \frac{1}{8} \end{bmatrix}, \quad A^{(2)} = \begin{bmatrix} \frac{1}{8} & \frac{3}{4} & \frac{1}{8} & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & \frac{1}{8} & \frac{3}{4} & \frac{1}{8} \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} \end{bmatrix}. \tag{4.18}$$

We illustrate the curve refinement process for curve $\mathbf{Q}_0$ in Figure 4.8. The curve, initially controlled by 4 control points, is split into 2 curve segments, each controlled by 4 control points (3 of which both curve segments have in common). Accordingly, after curve refinement, the curve is controlled by 5 control points.

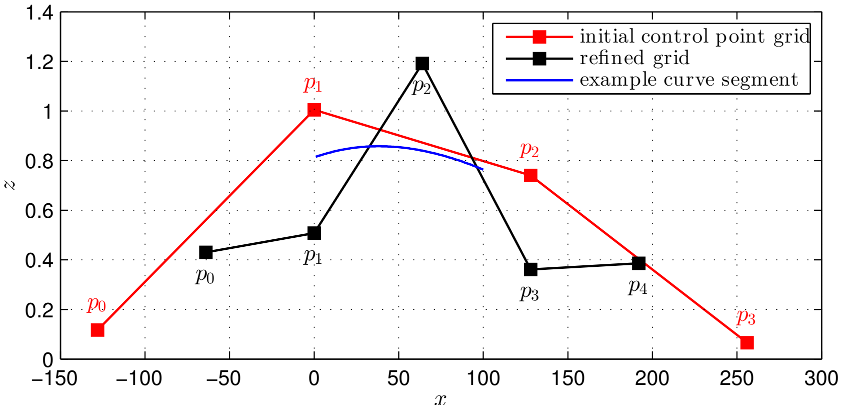


Figure 4.8: Curve refinement of example curve (in blue). The initial control points are illustrated in red, the refined control points in black. Note that the range of control point x-coordinates is decreased in the refinement.

4.2.6 Image Registration A

Channel-wise displacement for pixel locations $x \in \Pi$ is predicted by horizontal and vertical displacement functions $\tilde{z}_{\lambda,u}(x)$ and $\tilde{z}_{\lambda,v}(x)$ (obtained by the previous curve models) and same for each image row. The corresponding registration process was introduced as *Image registration A* in the block diagram of Figure 4.2 and is defined as follows:

Let $\widetilde{\mathbf{z}}_{\lambda,u}$ and $\widetilde{\mathbf{z}}_{\lambda,v}$ be the predicted displacement vectors for pixel locations $\mathbf{x} = [1, 2, ...c]^T \in \mathbb{R}^{c\times 1}$. By rewriting Equation 4.2, we obtain

$$I_\lambda\left(x + \widetilde{\mathbf{z}}_{\lambda,u}(x), y + \widetilde{\mathbf{z}}_{\lambda,v}(x)\right) = I_{\lambda,0}(x, y), \tag{4.19}$$

which describes the image transformation process that registers images in cube $\mathcal{C}$ to the reference $\mathcal{C}_0$.

Recall that the elements of image displacement vectors are real numbers, while image coordinates (ie. the discrete locations where pixel intensities can be displaced to) are integers. To overcome this problem and allow sub-pixel accurate image registration, image resampling has to be performed. Image resampling can then be described as the process of interpolating the image intensity profile values at discrete image locations [72]. We illustrate resampling Figure 4.9 for the one dimensional case with linear interpolation, which is used in this work.

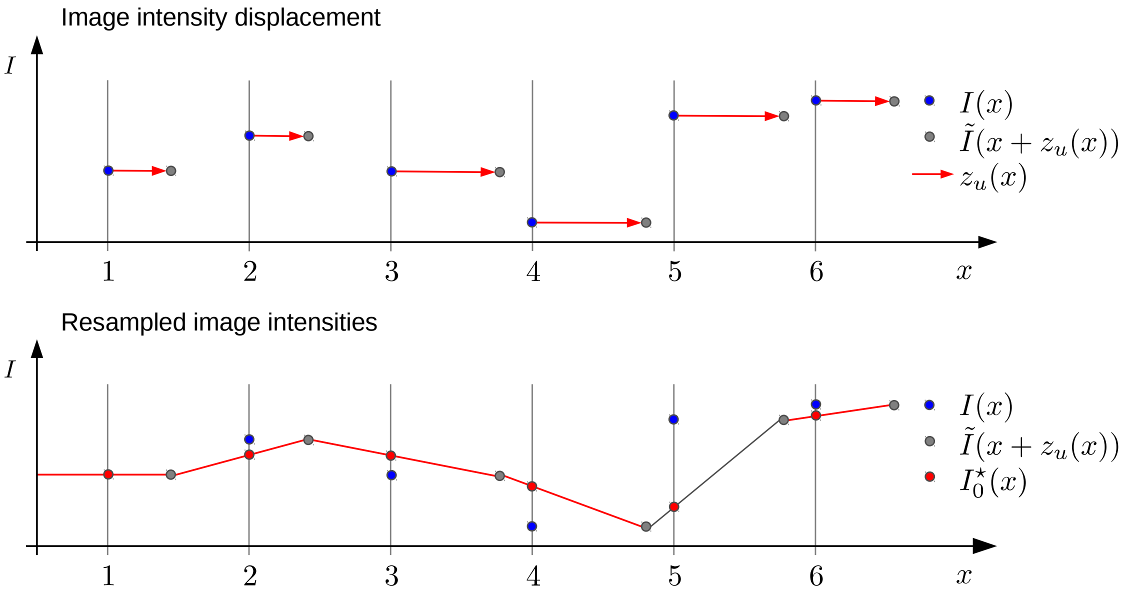


Figure 4.9: Illustration of image resampling: the blue circles indicate the initial image intensities I at integer image locations x. Red arrows indicate the real-valued displacement (in this example in horizontal direction). The displaced image intensities $\tilde{I}$ at non-integer locations (gray circles) are used as basis for linear interpolation (red lines between corresponding points) of image intensities $I_0^\star$ corresponding to the integer image locations (red circles).

4.3 Pixel correspondence based registration

As mentioned in the preface, simple geometrical calibration based image registration is insufficient when non-planar objects with three-dimensional surface structure are scanned. The registration fails because the displacement model fitted in the geometrical calibration process is only valid for registering image data with a similar geometrical structure as compared with the calibration target considered. The problem is more complex than geometrical calibration, because the three-dimensional surface structure is often unknown in practice. Without taking assumptions about the geometrical relations of camera and scanning object in three dimensional space, obtaining pixel displacement between individual image channels can be considered as a general correspondence problem. The task to be solved is then to find the pixel correspondence at every image location and image channel with respect to the reference image. Fortunately, for the case of LSMSC systems, we can make some assumptions and simplify the process:

- The displacements of image channels corresponding to the same sub-image (ie. acquired through the same lens) are constant, unlike the displacements with respect to channels of other sub-images. We therefore only need to establish correspondence between sub-images, and by that know the correspondence between all image channels.
- The geometrical calibration and subsequent image registration transform the image content in such a way that corresponding projected scene object points in each image channel are located in the same image row. The correspondence problem is then reduced from a search in two spatial image dimensions to a search only along the image dimension corresponding to the scanning line.
- The displacement between a pair of sub-images only occurs in one direction: either left, or right.

The second assumption is only valid to a certain degree. Because of the lens viewpoint variations illustrated in Figure 2.13b, there is an additional displacement in direction perpendicular to the scanning line that depends on the three-dimensional structure of the scanning object. We come back to this aspect in the experimental section (Section 4.4.3), where we quantify the corresponding displacement and also show that a violation of this assumption does not necessarily result in a failure of the image registration.

To solve the correspondence problem, we can adapt well-established methods from *computational stereo*, which is the problem of determining 3-dimensional structure of a scene from two or more images acquired by cameras with distinct viewpoints of the scene. In fact, following the problem classification scheme by Brown *et al.* [73], there are three major problems associated to computational stereo: *calibration*, *correspondence* and *reconstruction*.

In computational stereo, *calibration* is the process of determining the camera system's internal and external geometrical parameters (ie. the relative positions and orientations of each camera, optical centers, focal length and lens distortions), which allow relating the image coordinate system (with pixel units) to an external world coordinate system. Stated more generally as before, the *correspondence* problem consists of finding the locations in each camera image that are projections of the same physical location in space (ie. the world coordinate system). The displacement between a projected point in one image with respect to another image is called disparity and the set of all image disparities of a so-called stereo pair of images is called disparity map. The *reconstruction* problem consists of estimating three-dimensional scene object structure from a disparity map, based on known camera geometry.

For image registration of non-planar scanning objects using a LSMSC system, we do not need to solve all three computational stereo related problems. Calibration requirements are satisfied by the geometrical calibration process described in Section 4.2, which is a pre-processing step for pixel correspondence based registration. Accordingly, it is now assumed that corresponding object points are located in the same image row of each image channel. In most other stereo imaging applications, a transformation based on a calibration converts image pairs so that this is true as well (this process is then called image rectification). Further, we are only interested in the extraction of disparity maps for pairs of filtered RGB sub-images and not all image channels. For the case of $\mathbf{SC2_{12C}}$, there exist filtered RGB sub-images from 4 lenses and hence, 3 disparity maps are required to register all 12 channels with respect to a reference channel. Even though there are 6 possible stereo pairs, and correspondingly 20 disparity map combinations of three stereo pairs, we only consider the 4 combinations that share one RGB sub-image in common. This is because we want to reduce the risk of propagating errors from estimating pixel correspondence of one sub-image pair to another in the registration process. For convenience, we define this common sub-image as reference $L_0 \in \mathbb{R}^{r\times c\times 3}$ and denote the other sub-images as $L_1 \in \mathbb{R}^{r\times c\times 3}$, $L_2 \in \mathbb{R}^{r\times c\times 3}$ and $L_3 \in \mathbb{R}^{r\times c\times 3}$, where r and c are the image height and width respectively. Because the three disparity maps describe the pixel displacement of all 12 channels, we can stop after solving the correspondence problem and do not need to consider the reconstruction of the 3D scene[4].

There exists no general solution to the correspondence problem because pixel location matching can be ambiguous. For instance, the three-dimensional object structure might result in occlusion of object locations in one image view but not others. Brown *et al.* [73] summarize different approaches to cope with occlusion and distinguish correspondence algorithms that detect occlusion, algorithms that reduce the sensitivity to occlusion and algorithms that model occlusion geometrically. Because the viewpoint variations corresponding to sub-images of the LSMSC and the height variation of scanning objects considered in this dissertation are rather small, occlusion is not a problem and was therefore not considered particularly. Other examples where pixel location matching can be ambiguous are image specularities, image saturation, dark areas or regions that lack texture and therefore do not allow matching. For the case of the LSMSC, image pairs of sub-images are optically filtered differently, increasing the complexity of the matching. In fact, system optimality criteria used in filter selection often implicitly maximize the difference of the spectral information contained in individual sub-images, which is obviously contrary to the optimal condition for solving the correspondence problem[5].

To overcome the aforementioned problem, a color invariant feature matching for LSMSC systems was designed. Accordingly, filtered RGB sub-images of the 3 stereo pairs are transformed to so-called feature images that contain similar intensities for corresponding physical locations in the scanning scene (even though the image locations do not correspond). We denote the feature image corresponding to RGB sub-image $L_i \in \mathbb{R}^{r\times c\times 3}$ as $F_i \in \mathbb{R}^{r\times c}$, where $i = 0, 1, 2, 3$. Note that F_i is an intensity image and hence, the feature mapping is from $\mathbb{R}^3$ to $\mathbb{R}$.

Again, following the structure of Brown *et al.* [73], approaches to solve the correspondence problem can be categorized in local and global constraint based methods. Local methods constrain the matching based on a small number of pixels surrounding a pixel of interest, while global method constraints are related to the entire scanning line or the entire image.

We use a local constrained method that is based on block matching. Hence, we find disparity at a point in one image by comparing a local region around this point with regions in the other image. The comparison is based on the similarity of image intensities and we have found experimentally that feature mapping to a one-dimensional space is

---

[4] Note that the reconstruction problem can be considered as future work. We discuss this aspect further in Section 4.5

[5] The reader may refer to Section 3.2 to recall system optimality criteria in LSMSC system design.

adequate for correspondence estimation and reduces the computational burden.

For the LSMSC, we extract three disparity maps, denoted $\{D_i \in \mathbb{R}^{r\times c} | i = 1, 2, 3\}$, from three pairs of feature images $\{(F_0, F_i) | i = 1, 2, 3\}$. The block matching results in an estimate of pixel disparity in units of pixels. Sub-pixel accuracy can be achieved by incorporating an interpolation step in the block matching. This approach is described in Section 4.3.3.

For an illustration of the naming convention of variables in this section, the reader may refer to Figure 4.10.

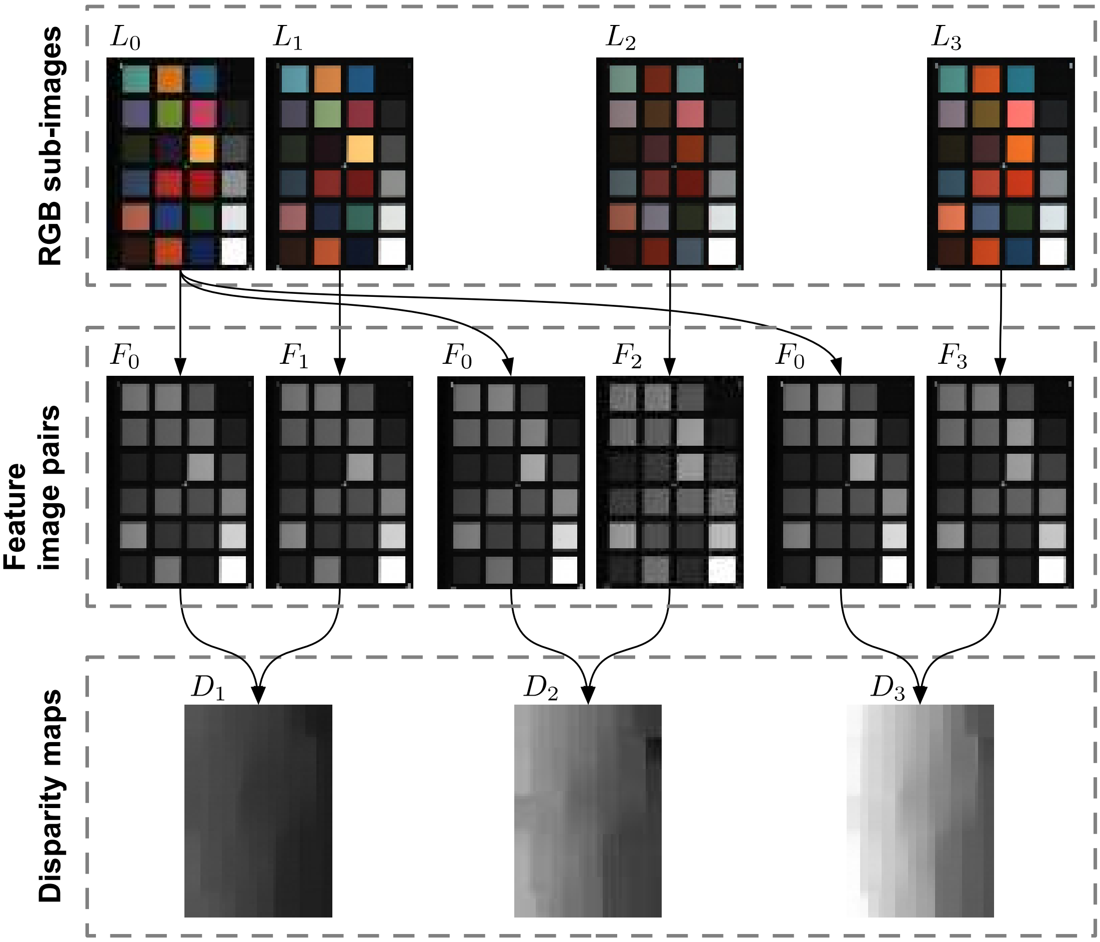


Figure 4.10: Example images of a scanning scene corresponding to a color chart on a ramp: The illustration contains RGB sub-images (L_0 to L_3), feature images (F_0 to F_3) and disparity maps (D_1 to D_3).

4.3.1 Color invariant feature mapping

The aim of color invariant feature mapping is to transform pairs of 3-channel RGB sub-images into 1-channel feature images that contain similar image intensities for corresponding physical locations in the scanning scene. The four filtered RGB sub-images of the LSMSC can be considered as images acquired by independent virtual camera systems. Without loss of generality, we consider only one pair of RGB sub-images, namely (L_0, L_1).

Suppose $\mathbf{x}_0$ and $\mathbf{x}_1$ are two vectors in $\mathbb{R}^3$, denoting camera responses of corresponding pixels in L_0 and L_1. Let $f_0 : \mathbb{R}^3 \mapsto \mathbb{R}$ and $f_1 : \mathbb{R}^3 \mapsto \mathbb{R}$ be functions that map RGB vectors $\mathbf{x}_0$ and $\mathbf{x}_1$ to scalar quantities x_0 and x_1. In mathematical terms, finding a color invariant feature mapping is the problem of minimizing $\|f_0(\mathbf{x}_0) - f_1(\mathbf{x}_1)\|$. This expression is satisfied by a mapping to zero, regardless the values of $\mathbf{x}_0$ and $\mathbf{x}_1$. Obviously, this is not a favorable solution as all image information is lost and hence, the problem has to be constrained further. In that sense, we seek for a solution that also retains as much image information of the RGB images as possible.

We propose two strategies for color invariant feature mapping. The first is based on a physical model of the acquisition process and the second is an empirically determined mapping function. In this dissertation, we only consider linear transformations as mapping functions. Preliminary results of experiments with other types of transformations are promising, but beyond the scope of this dissertation and hence motivate future research.

A Physical model based feature mapping

Let $\mathbf{x}_0 \in \mathbb{R}^{3\times 1}$ and $\mathbf{x}_1 \in \mathbb{R}^{3\times 1}$ be camera responses from sub-images L_0 and L_1 corresponding to the same physical location, captured by virtual camera systems with effective spectral responsivities $W_0 \in \mathbb{R}^{m\times 3}$ and $W_1 \in \mathbb{R}^{m\times 3}$ respectively. We note that the effective responsivities $W_0 = Y \operatorname{diag}(\mathbf{t}_0) \operatorname{diag}(\mathbf{l})$ and $W_1 = Y \operatorname{diag}(\mathbf{t}_1) \operatorname{diag}(\mathbf{l})$ combine the effect of spectral responsivity of the RGB sensor $Y \in \mathbb{R}^{m\times 3}$, scene illumination $\mathbf{l} \in \mathbb{R}^{m\times 1}$ and distinct spectral filter transmittances $\mathbf{t}_0 \in \mathbb{R}^{m\times 1}$ and $\mathbf{t}_1 \in \mathbb{R}^{m\times 1}$ for each sub-image respectively.

We start by defining the transformation $f_0 : \mathbb{R}^3 \mapsto \mathbb{R}$ for camera responses of the reference image L_0. Let $\mathbf{a}_{p,0} \in \mathbb{R}^{3\times 1}$ be the first principal component of matrix W_0 containing m effective sensor responsivities in rows. This component defines a linear transformation that maps camera response $\mathbf{x}_0$ to a one-dimensional feature space, such that the greatest variance of camera responses for W_0 is retained, assuming spectral signals acquired are uniformly distributed random variables. The problem is now to find an adequate transformation function for $\mathbf{x}_1$ with respect to W_1 to perform a mapping to the same feature space. The corresponding minimization problem is

$$\underset{\mathbf{a}_{p,1}}{\arg\min} \|W_1 \mathbf{a}_{p,1} - W_0 \ \mathbf{a}_{p,0}\|, \tag{4.20}$$

where $\mathbf{a}_{p,1} \in \mathbb{R}^{3\times 1}$. In most practical cases, there exists no solution such that $W_1 \mathbf{a}_{p,1} = W_0 \ \mathbf{a}_{p,0}$. The solution in least squares sense is

$$\mathbf{a}_{p,1} = (W_1^T W_1)^{-1} W_1^T W_0 \ \mathbf{a}_{p,0}. \tag{4.21}$$

Feature mappings of camera responses $\mathbf{x}_0$ and $\mathbf{x}_1$ are:

$$\begin{aligned} f_0(\mathbf{x}_0) &= \mathbf{a}_{p,0}^T \ \mathbf{x}_0 \\ f_1(\mathbf{x}_1) &= \mathbf{a}_{p,1}^T \ \mathbf{x}_1. \end{aligned} \tag{4.22}$$

B Empirically model based feature mapping

The physical model based method proposed in the previous section does not take assumptions about the statistical nature of the image scene. Instead of considering the acquisition process, the empirical model is based on observation data, ie. matrix $X_0 \in \mathbb{R}^{3\times rc}$ and $X_1 \in \mathbb{R}^{3\times rc}$ with camera response vectors from $L_0 \in \mathbb{R}^{r\times c\times 3}$ and $L_1 \in \mathbb{R}^{r\times c\times 3}$ in columns. We again start by defining the transformation $f_0 : \mathbb{R}^3 \mapsto \mathbb{R}$ for camera responses of the reference image L_0. Let $\mathbf{a}_{e,0} \in \mathbb{R}^{3\times 1}$ be the first principal component of the data matrix X_0. Likewise, the first principal component defines a linear transformation that transforms camera responses in X_0 to a one-dimensional feature space such that the greatest variance of the data is retained.

The transformation f_1 is found by solving the following least squares problem

$$\arg\min_{\mathbf{a}_{e,1}} \|\mathbf{a}_{e,1}^T X_1 - \mathbf{a}_{e,0}^T X_0\|, \tag{4.23}$$

where $\mathbf{a}_{e,1} \in \mathbb{R}^{3\times 1}$. The solution in least squares sense is

$$\mathbf{a}_{e,1} = (X_1 X_1^T)^{-1} X_1 X_0^T \ \mathbf{a}_{e,0}. \tag{4.24}$$

Feature mappings of camera responses $\mathbf{x}_0$ and $\mathbf{x}_0$ are

$$\begin{aligned} f_0(\mathbf{x}_0) &= \mathbf{a}_{e,0}^T \ \mathbf{x}_0 \\ f_1(\mathbf{x}_1) &= \mathbf{a}_{e,1}^T \ \mathbf{x}_1. \end{aligned} \tag{4.25}$$

4.3.2 Block matching

We use exhaustive search block matching to extract pixel-wise disparities from feature images corresponding to stereo pairs of RGB sub-images. Like before, we only consider disparity extraction for one stereo image pair, namely (F_0, F_1). We remind the reader that we regard geometrical calibration as a preprocessing step to block matching and because of this, sub-images of the LSMSC are assumed to be rectified (ie. points in F_0 are located in the same image row as corresponding points in F_1). Hence, the result of block matching is the disparity map $D_1 \in \mathbb{Z}^{r\times c}$, containing horizontal pixel displacements of F_1 with respect to F_0.

To find disparity $D_1(x, y)$ at pixel location (x, y), a macro block in F_0 is defined in the neighborhood of (x, y) and a matching block with similar image intensities is sought in the same row of F_1. The value of $D_1(x, y)$ is then the pixel displacement between pixel location (x, y) and the matching location in horizontal direction. To reduce the computational burden, we further define a relatively small search range (ie. the search window) rather than the entire image line. The block matching process is illustrated in Figure 4.11. Because F_0 and F_1 do not contain exactly the same intensities for pixels of the same physical location in the scanning scene, we consider block matching as a minimization problem and use the sum of squared difference (SSD) as similarity measure:

$$\arg\min_{D_1(x,y)\in\Psi} \sum_{i=l}^{r} \sum_{j=e}^{t} (F_0(i, j) - F_1(i + D(x, y), j))^2, \tag{4.26}$$

where $l = x - (b - 1)/2$, $r = x + (b - 1)/2$, $e = y - (b - 1)/2$ and $t = y + (b - 1)/2$ are the indexes of the left column, right column, bottom row and top row of the block centered at position (x, y) and $b \times b$ is the block size. The search window is defined in the range $\Psi = \{0, 1, ...w - 1\}$ for $D_1(x, y)$ in case F_0 is the left image of the stereo pair. If F_0 is the right image, $\Psi = \{-w+1, -w+2, ...0\}$. Scalar w denotes the width of the search window, and hence defines the number of times the reference block has to be compared in order to find a match. The scalars b and w are user defined block matching parameters and have to be odd integers. The optimal choice for b is related to the texture of image objects and the size of repetitive patterns in the images and w is usually selected slightly larger than

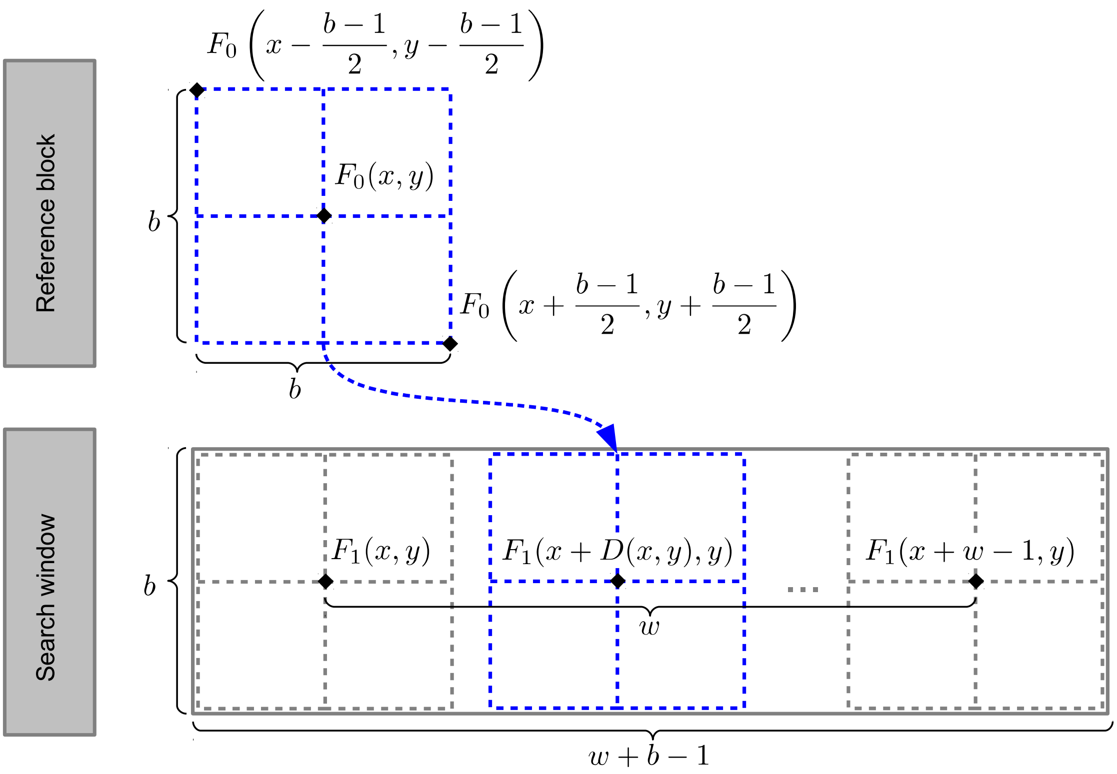


Figure 4.11: Block matching process in the case that F_0 is the left image and F_1 is the right image of a stereo pair.

the maximum absolute disparity expected to occur between F_0 and F_1. We implemented SSD based block matching using a fast Fourier transformation based approach that is computationally less expensive than computing SSD in the spatial domain [74].

There exists a severe limitation of SSD based block matching, namely *edge-fattening*, an effect that causes the center of a block to inherit disparity of the most contrasted pixels in this block [75]. We recognize that a solution to this problem has been proposed in the literature [75], but have not included it in our work because of time constraints.

4.3.3 Sub-pixel displacement estimation

The disparity map $D_1 \in \mathbb{Z}^{r\times c}$ obtained from the feature image pair $(F_0, F_1) \in \mathbb{R}^{r\times c}$ via SSD based block matching contains disparities in pixel units. Various extensions for sub-pixel accurate disparity extraction have been proposed (eg. [76, 77, 78]). We follow an approach that consists of fitting a quadratic curve to the correlation window of discrete disparities in block-matching, and evaluating this curve at its maximum [76]. Accordingly, let $d \in \mathbb{Z}$ be the integer displacement obtained by block matching corresponding to minimum SSD value $s \in \mathbb{R}$. A quadratic estimator (ie. 2nd order polynomial function) of the interpolated minimum location is:

$$\widetilde{d} = d + \frac{1}{2}\left(\frac{s_r - s_l}{2s - s_l - s_r}\right), \tag{4.27}$$

where sub-pixel displacement $\widetilde{d}$ takes values $0 \leq \widetilde{d} \leq w-1 \in \mathbb{R}$. Scalars $s_l \in \mathbb{R}$ and $s_r \in \mathbb{R}$ are the SSD values of discrete locations on the right and left of d respectively. The sub-pixel disparities are computed at every block-matching step and assigned to $\widetilde{D} \in \mathbb{R}^{r\times c}$.

The approach was chosen because it adds little additional cost to the block-matching. As noted by various researchers, computing disparities using a quadratic estimator can introduce a systematic bias called *pixel-locking*, which is a tendency of the estimated sub-pixel displacements to concentrate towards integer values. Various methods have been proposed to overcome this problem [79, 80, 77], which were not considered in this work and therefore motivate future work.

4.3.4 Image registration B

The registration of non-planar scanning objects (*Image registration B* in Figure 4.2) is based on geometrically calibrated images that are assumed to be void of vertical displacement. The horizontal channel-wise displacement for pixel location $(x, y) \in \Omega$ is derived from scene content of stereo pairs $\{(L_0, L_i) \in \mathbb{R}^{r\times c\times 3} \times \mathbb{R}^{r\times c\times 3} | i = 1, 2, 3\}$ of RGB sub-images and contained in disparity maps $\{\widetilde{D}_i \in \mathbb{R}^{r\times c} | i = 1, 2, 3\}$. Accordingly, displacements for image channels of the same sub-image are similar and image registration B only consists of registering sub-images. The respective image transformation is then

$$\widetilde{L}_i(x, y) = L_i(x + \widetilde{D}_i(x + y), y), \tag{4.28}$$

where $\widetilde{L}_i$ is the registered RGB sub-image and $L_i(x, y)$ is the pixel value at location $(x, y) \in \Omega$ in the unregistered image. Disparity $D_i(x + y)$ is the horizontal displacement of $L_i(x, y)$ with respect to $L_0(x, y)$, and $i = 1, 2, 3$. As before, image resampling has to be performed to allow sub-pixel accurate image registration and we again use linear interpolation for the resampling.

4.3.5 Limitations with respect to reflectance measurement

We have mentioned several potential limitations that might reduce the registration performance of non-planar scanning objects. Edge-fattening, pixel-locking, occlusion or specularities are some examples. These potential limitations are specific to correspondence estimation, but there exist other restrictions in LSMSC image acquisition of non-planar scanning objects with respect to reflectance measurements.

The first is related to the scene illumination considered for LSMSC systems, which is a directional line light source[6]. For this type of light source, the light is focused to a rather narrow region around the scanning line. If scanning objects with three dimensional structure are considered, the scene radiance sensed by the image sensor depends on the object height. This effect is usually undesired but preliminary experiments have shown that it can be corrected effectively with an additional calibration step if the object height is estimated correctly from pixel disparities and is not too large.

Apart from that, height dependence of scene illumination also influences spectral reflectance and color measurement. We mostly regard flat scanning objects in this dissertation, so the effect on spectral or color measurement performance was not evaluated numerically yet.

4.4 Evaluation

This section is devoted to the evaluation of the image registration framework proposed before. We have introduced scenarios of LSMSC image registration for planar and non-planar scanning objects. In the case of planar scanning objects, registration is based on

---

[6] See Section 2.3.2.A.

geometrical calibration. For non-planar scanning objects, color invariant feature mapping and block matching based correspondence estimation are required additionally.

The evaluation in this section is divided into subsections that individually assess these modules. All experiments are performed with image data acquired by our laboratory LSMSC corresponding to $\mathbf{SC2_{12C}}$.

4.4.1 Geometrical calibration

We compared registration performance of the polynomial and B-spline based models of the geometrical calibration scheme. Based on Zitová *et al.* [81], registration performance can be quantified in terms of *localization error*, *matching error* and *alignment error*. We additionally assessed *fitting error* and explain each measure as follows:

Localization error refers to the displacement of key-point coordinates due to inaccurate detection. This type of error is intrinsic to the key-point detection and therefore can not be measured directly from the image data.

Matching error is measured by counting false matches when establishing correspondence between key-points for all image channels. The key-point extraction scheme used in this work seems robust, as matching error was always found to be zero.

Alignment error refers to the residual displacement with respect to a reference image after image registration. This error is often quantified by computing the mean-square error over available key-point locations after registration. We add measures of first order statistics to our evaluation. We refer to the alignment error when evaluating registered image cubes other than the one used to extract key-points for model fitting. By that, we separate model fitting and model evaluation data, which allows us to identify whether or not a registration model is prone to over-fitting.

Fitting error is similar to alignment error and defined as the residual displacement of the cube used for model fitting after registration.

Because horizontal and vertical displacement are independent quantities, we proceed by evaluating each component separately. Obviously, fitting error was computed from a registered image cube of the geometrical calibration target. The alignment error was computed from another image cube that contains image data of a test target. For convenience, we designed the test target in a similar fashion as the calibration target, containing a printed checkerboard pattern (see Section 4.2.1). The rotation angle was modified, so instead of using $\alpha = 5.47\,^{\circ}$, we used $\alpha = -5.47\,^{\circ}$.

We denote errors in vector notation. Hence, for every image channel λ, we computed $\mathbf{e}_{\lambda,u}$ and $\mathbf{e}_{\lambda,v}$. These vectors contain the horizontal and vertical residual displacement of k key-point locations in $\mathcal{K}$ with respect to reference image channel $I_{\lambda,0}$ and are defined as

$$\begin{aligned}\mathbf{e}_{\lambda,u} &= [x_{\lambda,i} - x_{0,i}]_{k\times 1},\ i = 1,2,...k\\ \mathbf{e}_{\lambda,v} &= [y_{\lambda,i} - y_{0,i}]_{k\times 1},\ i = 1,2,...k.\end{aligned} \tag{4.29}$$

The polynomial and B-spline based model were fitted using key-points extracted from an image cube of the scanned calibration scene. For the B-spline model, a refinement threshold of 0.1 pixel average residual displacement was selected. Fitting error was computed for the calibration scene cube for both methods and is illustrated in Table 4.1, Table

4.2 and Figure 4.12. The alignment error was computed from the image cube of the test scene and is illustrated in Table 4.3, Table 4.4 and Figure 4.13. Note that channel 1 is generally void of displacement because it was used as reference. We therefore omit the corresponding numbers in tabulated data.

The following observation can be drawn from the results:

Displacement of unregistered calibration scene cube: Table 4.1 (a), 4.2 (a)

Horizontal displacement is on average larger than vertical displacement. The maximum displacement in horizontal direction is approximately 1.9 pixels and in vertical direction approximately 0.9 pixels.

The maximum displacement of key-points from channels belonging to the same subimage is less than 0.2 pixels for both horizontal and vertical direction.

Displacement of unregistered test scene cube: Table 4.3 (a), 4.4 (a)

1st oder statistics of horizontal and vertical displacement are in good agreement with the data from the unregistered calibration scene cube.

Fitting error: Table 4.1 (b,c), 4.2 (b,c)

Polynomial as well as B-spline model result in a residual displacement of less than 0.1 pixel on average for any channel, for both horizontal and vertical direction.

Judging from the maximum displacement, a larger residual remains in horizontal direction as compared with vertical direction. Generally, a smaller maximum displacement in both directions is achieved by the B-spline model.

Comparing the residual displacement graph in Figure 4.12, it can be observed that the data scatter appears smaller in some channels for B-spline as compared with the polynomial model. This finding also manifests in low corresponding channel-wise standard deviation of the residuals.

Alignment error: Table 4.3 (b,c), 4.4 (b,c)

Again, polynomial as well as B-spline model result in a residual displacement less than 0.1 pixel on average for any channel, for both horizontal and vertical direction, confirming that both models generalize well from calibration scene cube to test scene cube. The vertical displacement seems to be corrected slightly better than the horizontal displacement for both models, if average residuals are considered.

Contrary to the fitting error, the maximum error in horizontal and vertical direction is slightly larger for B-spline than the polynomial model. This can be an indicator for a slight over-fitting of the calibration scene data (because the maximum fitting error is very low for the B-spline model) which in turn leads to an over-compensation of the residuals in the test scene data. This effect can also be confirmed by observing the fitting and alignment error graphs in Figure 4.12(b) and 4.13(b). For instance in channel 4, residual displacement pattern of fitting and alignment error for the polynomial model are similar, whereas a clear difference can be observed for the B-spline model.

Table 4.1: Calibration cube: First order statistics of channel-wise and overall absolute **horizontal** key-point displacement and **fitting error** $|\mathbf{e}_{u,\lambda}|$ in pixel units.

(a) Unregistered

λ	Avg.	Std.	Max.
2	0.04	0.03	0.11
3	0.06	0.04	0.16
4	0.90	0.41	1.77
5	0.88	0.40	1.82
6	0.90	0.42	1.92
7	0.22	0.13	0.63
8	0.21	0.12	0.60
9	0.26	0.15	0.55
10	0.57	0.30	1.18
11	0.55	0.29	1.16
12	0.57	0.30	1.07
2-12	0.47	0.42	1.92

(b) Polynomial model

λ	Avg.	Std.	Max.
2	0.01	0.01	0.05
3	0.02	0.01	0.07
4	0.02	0.02	0.14
5	0.02	0.02	0.14
6	0.03	0.03	0.16
7	0.05	0.04	0.17
8	0.05	0.04	0.15
9	0.05	0.04	0.18
10	0.04	0.04	0.17
11	0.04	0.03	0.16
12	0.04	0.03	0.16
2-12	0.03	0.03	0.18

(c) B-spline model

λ	Avg.	Std.	Max.
2	0.01	0.01	0.04
3	0.01	0.01	0.03
4	0.02	0.01	0.05
5	0.01	0.01	0.06
6	0.02	0.01	0.05
7	0.02	0.01	0.04
8	0.02	0.01	0.05
9	0.02	0.01	0.05
10	0.02	0.01	0.07
11	0.02	0.01	0.07
12	0.02	0.01	0.07
2-12	0.02	0.01	0.07

Table 4.2: Calibration cube: First order statistics of channel-wise and overall absolute **vertical** key-point displacement and **fitting error** $|\mathbf{e}_{v,\lambda}|$ in pixel units.

(a) Unregistered

λ	Avg.	Std.	Max.
2	0.03	0.02	0.10
3	0.04	0.03	0.12
4	0.19	0.16	0.60
5	0.18	0.15	0.57
6	0.18	0.14	0.55
7	0.47	0.16	0.77
8	0.51	0.16	0.81
9	0.56	0.16	0.85
10	0.22	0.17	0.74
11	0.21	0.19	0.76
12	0.21	0.19	0.77
2-12	0.25	0.23	0.85

(b) Polynomial model

λ	Avg.	Std.	Max.
2	0.02	0.01	0.08
3	0.02	0.02	0.11
4	0.02	0.01	0.05
5	0.02	0.02	0.08
6	0.03	0.02	0.11
7	0.02	0.02	0.10
8	0.03	0.02	0.11
9	0.03	0.02	0.12
10	0.02	0.02	0.13
11	0.02	0.01	0.05
12	0.02	0.02	0.09
2-12	0.02	0.02	0.13

(c) B-spline model

λ	Avg.	Std.	Max.
2	0.01	0.01	0.05
3	0.02	0.02	0.09
4	0.02	0.01	0.08
5	0.02	0.02	0.08
6	0.03	0.02	0.11
7	0.03	0.02	0.09
8	0.03	0.02	0.12
9	0.03	0.03	0.12
10	0.03	0.02	0.10
11	0.02	0.02	0.09
12	0.02	0.02	0.09
2-12	0.02	0.02	0.12

Table 4.3: Test cube: First order statistics of channel-wise and overall absolute **horizontal** key-point displacement and **alignment error** $|\mathbf{e}_{u,\lambda}|$ in pixel units.

(a) Unregistered

λ	Avg.	Std.	Max.
2	0.04	0.02	0.12
3	0.06	0.04	0.18
4	0.91	0.43	1.86
5	0.89	0.42	1.94
6	0.92	0.44	2.07
7	0.25	0.15	0.61
8	0.24	0.14	0.58
9	0.31	0.15	0.61
10	0.60	0.32	1.18
11	0.59	0.31	1.19
12	0.61	0.33	1.20
2-12	0.49	0.43	2.07

(b) Polynomial model

λ	Avg.	Std.	Max.
2	0.01	0.01	0.05
3	0.02	0.01	0.06
4	0.05	0.03	0.15
5	0.04	0.03	0.15
6	0.05	0.03	0.15
7	0.08	0.05	0.21
8	0.08	0.05	0.20
9	0.08	0.05	0.20
10	0.10	0.05	0.28
11	0.10	0.05	0.28
12	0.10	0.05	0.28
2-12	0.06	0.05	0.28

(c) B-spline model

λ	Avg.	Std.	Max.
2	0.03	0.02	0.10
3	0.04	0.03	0.16
4	0.06	0.04	0.27
5	0.05	0.04	0.30
6	0.06	0.04	0.32
7	0.07	0.05	0.21
8	0.08	0.05	0.20
9	0.08	0.05	0.21
10	0.10	0.05	0.27
11	0.10	0.05	0.27
12	0.10	0.05	0.27
2-12	0.07	0.05	0.32

Table 4.4: Test cube: First order statistics of channel-wise and overall **absolute** vertical key-point displacement and **alignment error** $|\mathbf{e}_{v,\lambda}|$ in pixel units.

(a) Unregistered

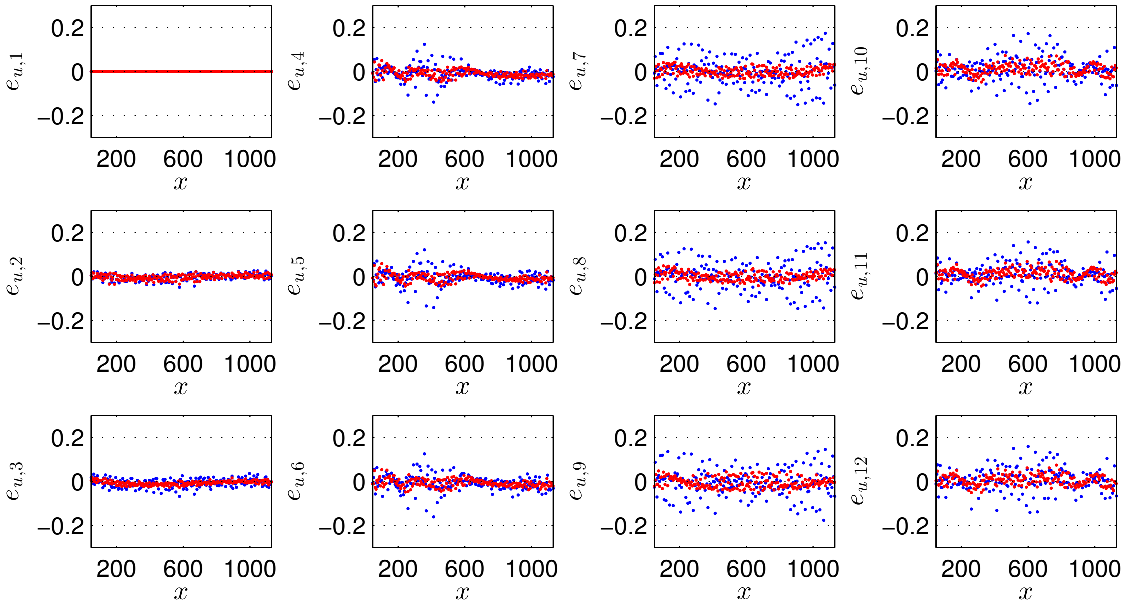
λ	Avg.	Std.	Max.
2	0.02	0.01	0.07
3	0.03	0.02	0.12
4	0.19	0.16	0.60
5	0.18	0.15	0.58
6	0.18	0.14	0.58
7	0.46	0.16	0.75
8	0.50	0.16	0.81
9	0.55	0.17	0.85
10	0.21	0.18	0.76
11	0.21	0.20	0.78
12	0.20	0.21	0.79
2-12	0.25	0.23	0.85

(b) Polynomial model

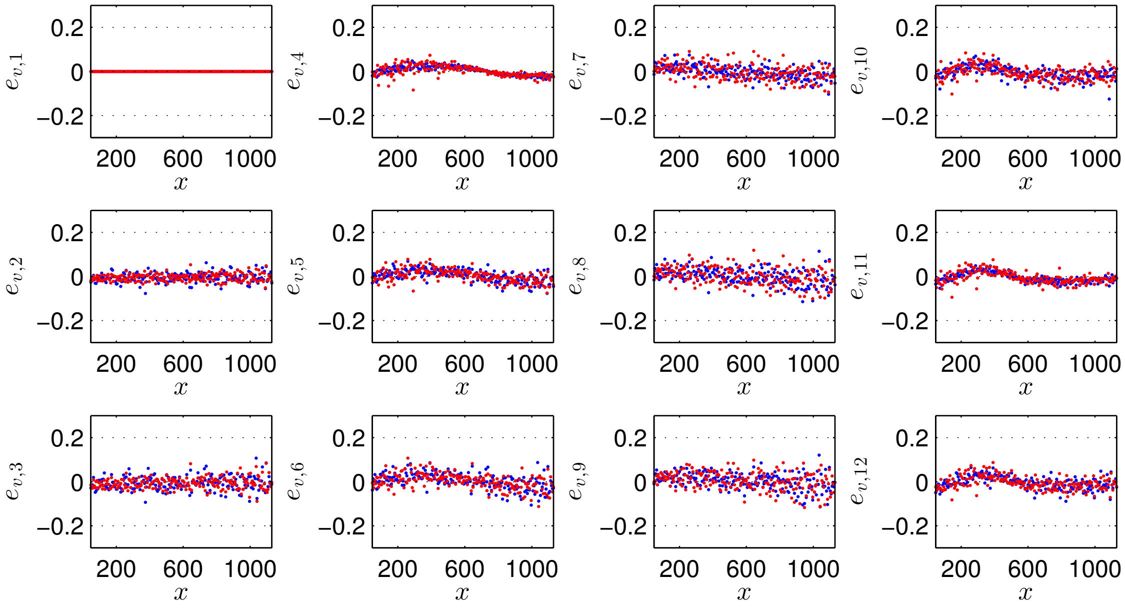
λ	Avg.	Std.	Max.
2	0.01	0.01	0.04
3	0.02	0.01	0.08
4	0.01	0.01	0.05
5	0.02	0.01	0.06
6	0.02	0.01	0.08
7	0.02	0.01	0.07
8	0.02	0.02	0.08
9	0.02	0.02	0.08
10	0.02	0.02	0.08
11	0.02	0.02	0.07
12	0.02	0.02	0.08
2-12	0.02	0.02	0.08

(c) B-spline model

λ	Avg.	Std.	Max.
2	0.01	0.01	0.05
3	0.02	0.02	0.08
4	0.04	0.03	0.17
5	0.04	0.03	0.18
6	0.04	0.04	0.19
7	0.02	0.02	0.07
8	0.03	0.02	0.11
9	0.03	0.03	0.13
10	0.03	0.03	0.15
11	0.02	0.03	0.14
12	0.03	0.03	0.15
2-12	0.03	0.03	0.19

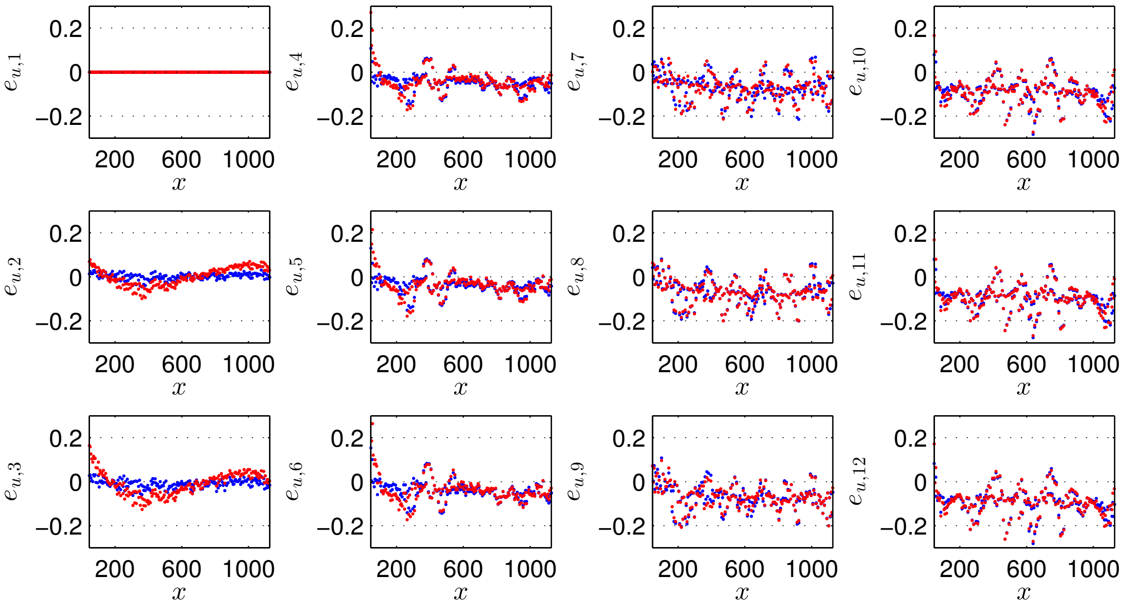


(a) Horizontal fitting error $e_{u,\lambda}$.

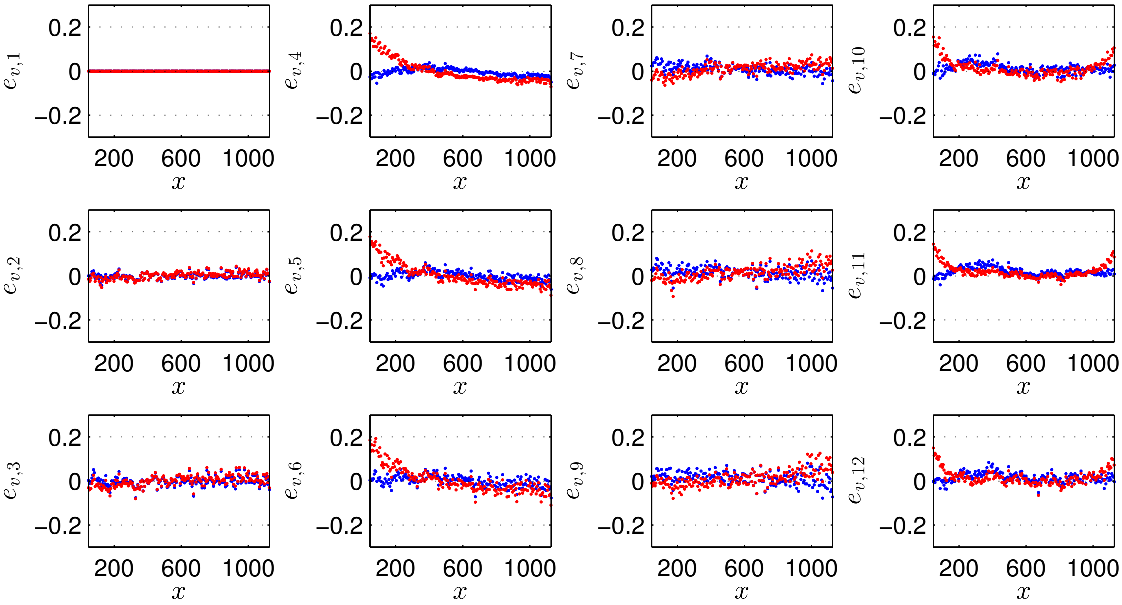


(b) Vertical fitting error $e_{v,\lambda}$

Figure 4.12: Fitting error and corresponding key-point x coordinates for image channel $\lambda = \{1, 2, ...12\}$. Illustrated in blue are the residuals for the polynomial model, in red for the B-spline model.



(a) Horizontal alignment error $e_{u,\lambda}$.



(b) Vertical alignment error $e_{v,\lambda}$

Figure 4.13: Alignment error and corresponding key-point x coordinates for image channel $\lambda = \{1, 2, ...12\}$. Illustrated in blue are the residuals for the polynomial model, in red for the B-spline model.

4.4.2 Color invariant feature mapping

Feature mapping is required to transform pairs of RGB sub-images to feature images that contain similar image intensities for corresponding physical locations in the scanning scene. We quantified feature mapping performance by evaluating the $RMSE$ between feature images in normalized feature space[7]. In this experiment, we considered data from **DS1** and **DS4** and compared $\mathbf{SC1_{12C}}$ and $\mathbf{SC2_{12C}}$. The physical model based feature mapping functions corresponding to Equation 4.21 were obtained from the system's effective spectral responsivities. The empirical model based mapping functions were extracted from acquired camera response data of **DS1** and respective mapping functions are in accordance with Equation 4.24.

To select a reference lens for the two systems, we empirically evaluated all four options and selected the one with highest average feature mapping performance for the empirical model. Accordingly, for $\mathbf{SC1_{12C}}$, lens 2 was selected as reference. We denote the corresponding RGB sub-image as L_0. The other sub-images are then L_1, L_2 and L_3 for lens 1, 3 and 4 respectively. For $\mathbf{SC2_{12C}}$, lens 1 was selected as reference, so the corresponding sub-image is denoted L_0. The other sub-images L_1, L_2 and L_3 correspond to lens 2, 3 and 4 respectively.

We show the numerical results of the experiment in Table 4.5. Further, we illustrate the mappings for $\mathbf{SC1_{12C}}$ in Figure 4.14 and 4.15, and for $\mathbf{SC2_{12C}}$ in Figure 4.16 and 4.17. We conclude the following from the experiment:

Comparing $\mathbf{SC1_{12C}}$ and $\mathbf{SC2_{12C}}$: for the physical model, the average $RMSE$ is considerably higher for $\mathbf{SC2_{12C}}$ than for $\mathbf{SC1_{12C}}$. As channels of $\mathbf{SC2_{12C}}$ were specially optimized to capture distinct information (see Section 3.4.4), mapping from different RGB sub-images to a common feature space is more difficult. This can explain the lower feature mapping performance for $\mathbf{SC2_{12C}}$. For the empirical model we do not see this effect.

Comparing DS1 and DS4: for the physical model, performance of **DS4** is generally higher than for **DS1**. The physical modeling process is data independent, so the reason for the higher overall performance might be that **DS1** contains a larger amount of color patches for which the performance is low, as compared with **DS4** which contains approximately eight times less samples. For the empirical model, performance is in most cases and on average higher for **DS1**. This does not come as a surprise, as this dataset was used for model fitting.

Comparing physical and empirical model: for both system configurations, the performance of the empirical model is significantly higher than that of the physical model. For the physical model, we made an assumption about the statistical distribution of camera responses, namely that any physically possible camera response is equally likely to occur. The empirical model takes the distribution of a set of camera responses from a color chart into account and the higher performance indicates that this is useful in practice.

Another possible explanation for the lower performance is the potential shortcoming in the acquisition model that was mentioned in the last chapter (Section 3.4.4)[8].

---

[7] Intensity values over all feature maps were scaled to be in the range $[0, 1]$.

[8] We note that a direct comparison of feature mapping based on simulated and measured camera response data could clarify this aspect. This was not done yet because of time constraints.

Table 4.5: Empirical evaluation of the performance of physical and empirical feature mapping for $\mathbf{SC1_{12C}}$ and $\mathbf{SC2_{12C}}$. Dataset **DS1** was considered for fitting the empirical model, **DS4** is an independent dataset for testing. Numerical values correspond to the $RMSE$ between pairs of feature images. The last column shows the average $RMSE$ over all sub-image pairs.

$\mathbf{SC1_{12C}}$	(F_0, F_1)	(F_0, F_2)	(F_0, F_3)	Avg.
Phys. model **DS1**	0.077	0.040	0.085	0.068
Phys. model **DS4**	0.091	0.039	0.071	0.067
Emp. model **DS1**	0.050	0.010	0.022	0.027
Emp. model **DS4**	0.071	0.013	0.024	0.036

$\mathbf{SC2_{12C}}$	(F_0, F_1)	(F_0, F_2)	(F_0, F_3)	Avg.
Phys. model **DS1**	0.215	0.153	0.083	0.150
Phys. model **DS4**	0.206	0.142	0.077	0.142
Emp. model **DS1**	0.023	0.042	0.026	0.030
Emp. model **DS4**	0.014	0.060	0.040	0.038

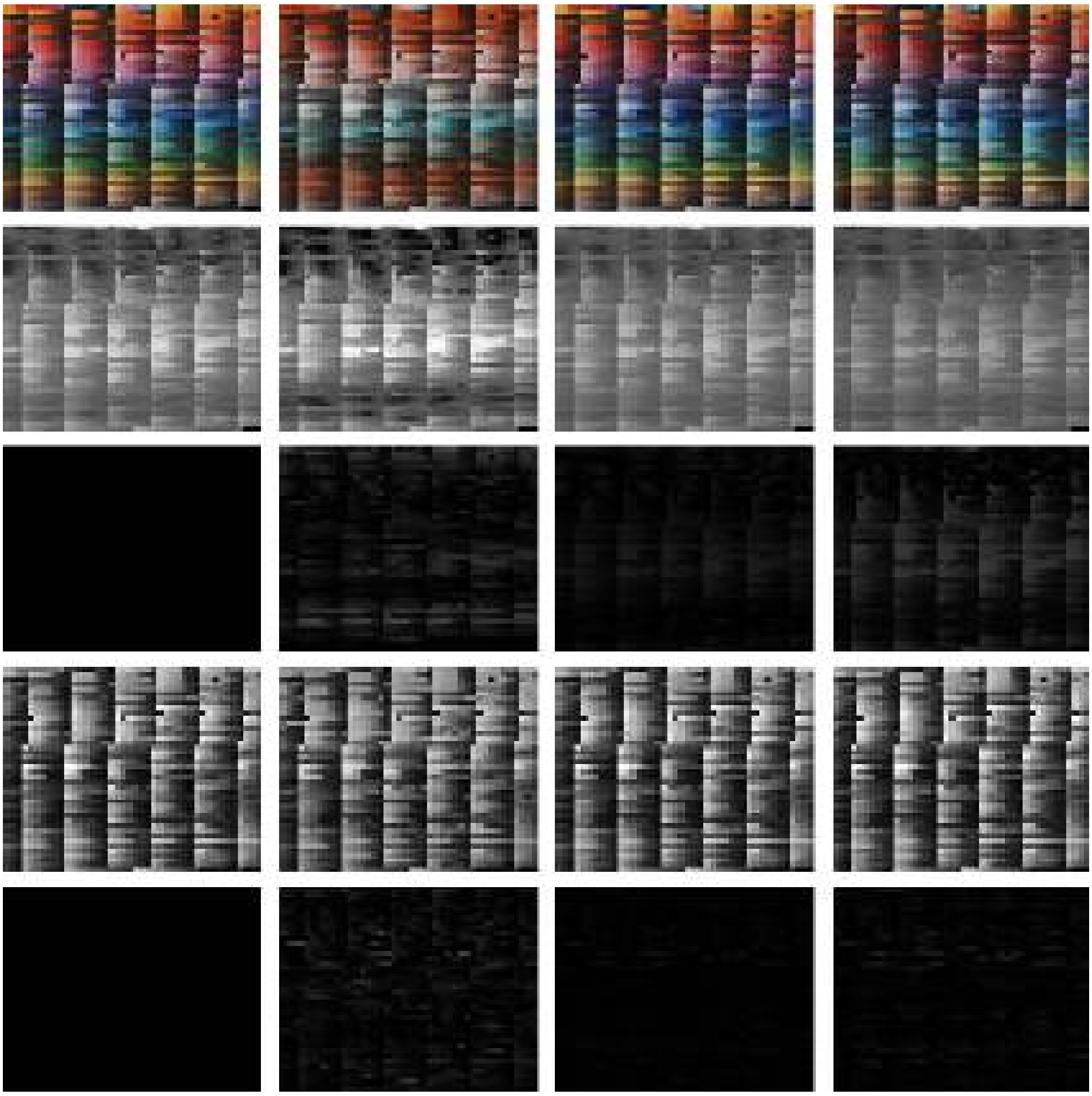


Figure 4.14: Color invariant feature mapping for camera responses from LSMSC $\mathbf{SC1_{12C}}$ - **DS1**: color patches from RGB sub-images L_0 to L_3 (row 1); physical model based feature images F_0 to F_3 (row 2); corresponding residuals with respect to F_0 (row 3); empirical model based feature images (row 4); corresponding residuals with respect to F_0 (row 5).

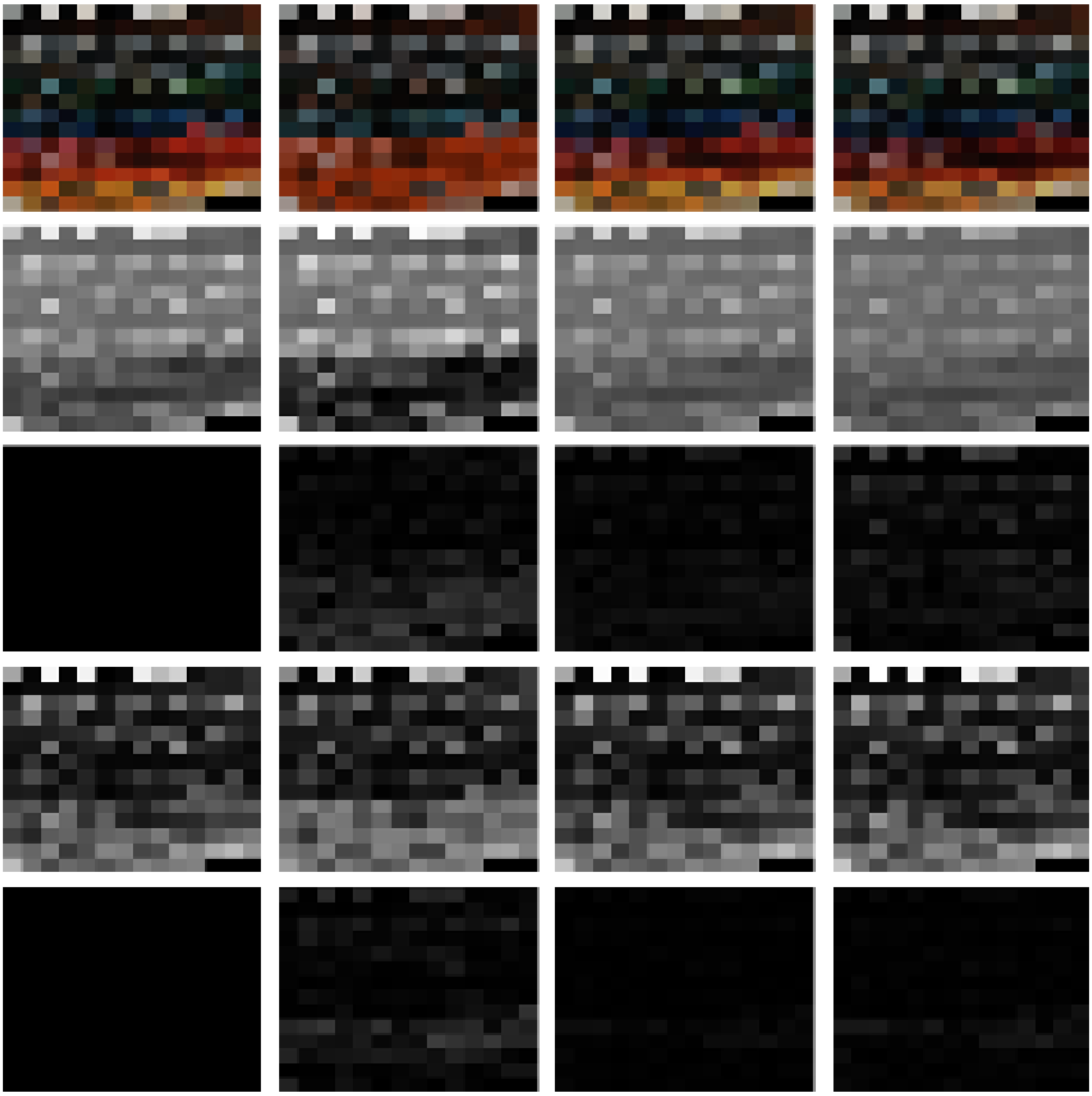


Figure 4.15: Color invariant feature mapping for camera responses from LSMSC $\mathbf{SC1_{12C}}$ - **DS4**: color patches from RGB sub-images L_0 to L_3 (row 1); physical model based feature images F_0 to F_3 (row 2); corresponding residuals with respect to F_0 (row 3); empirical model based feature images (row 4); corresponding residuals with respect to F_0 (row 5).

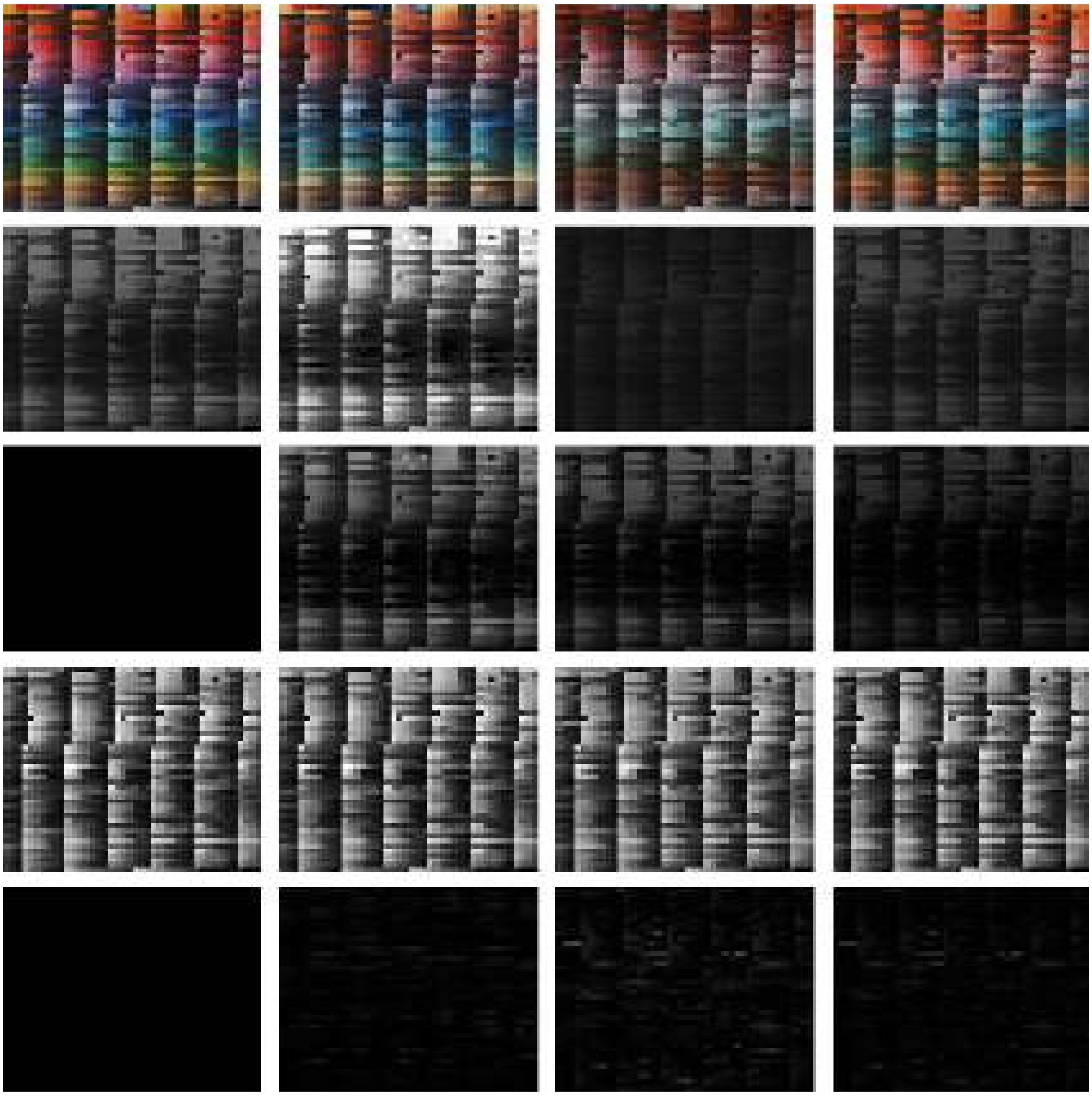


Figure 4.16: Color invariant feature mapping for camera responses from LSMSC $\mathbf{SC2_{12C}}$ - **DS1**: color patches from RGB sub-images L_0 to L_3 (row 1); physical model based feature images F_0 to F_3 (row 2); corresponding residuals with respect to F_0 (row 3); empirical model based feature images (row 4); corresponding residuals with respect to F_0 (row 5).

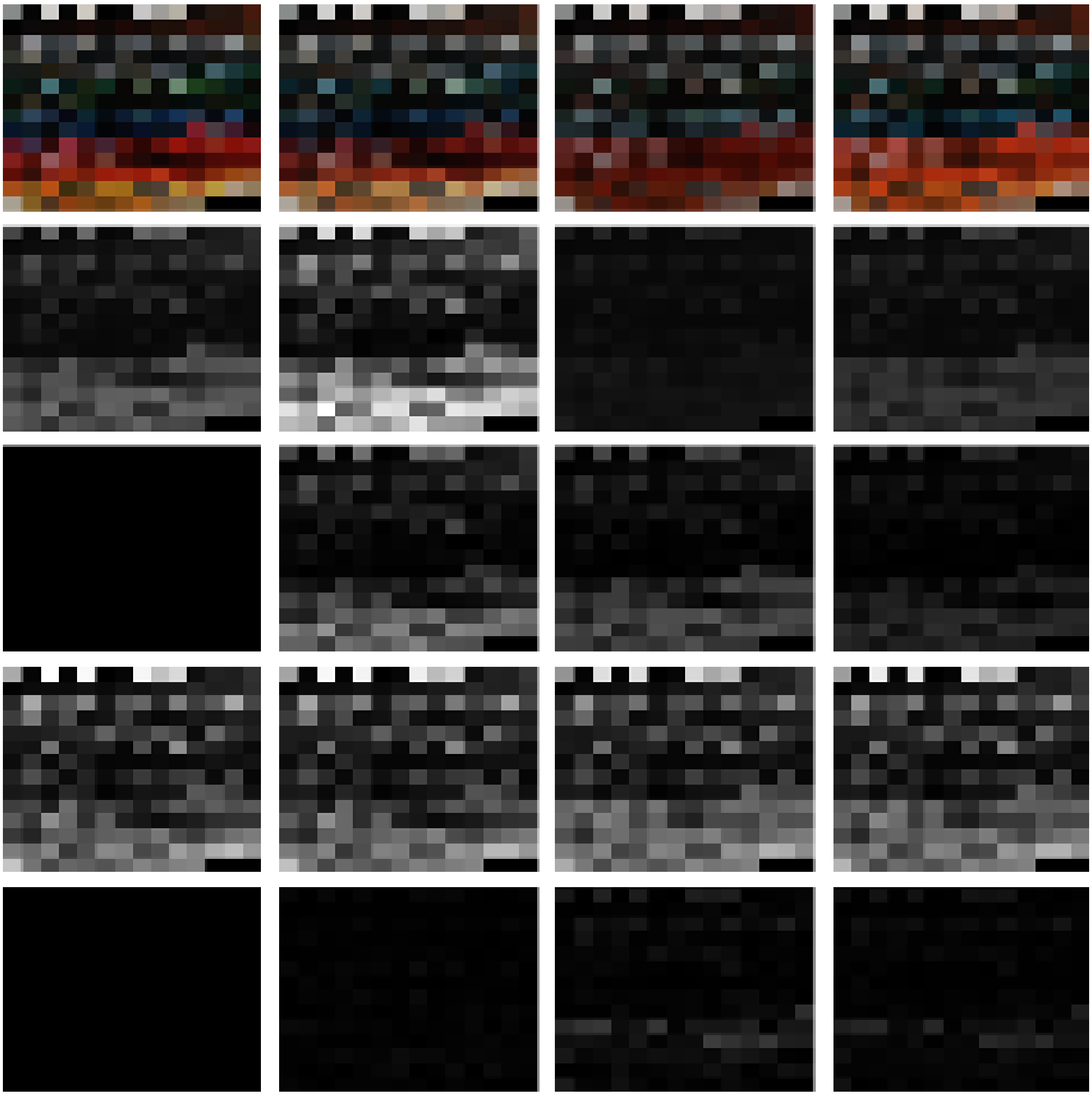
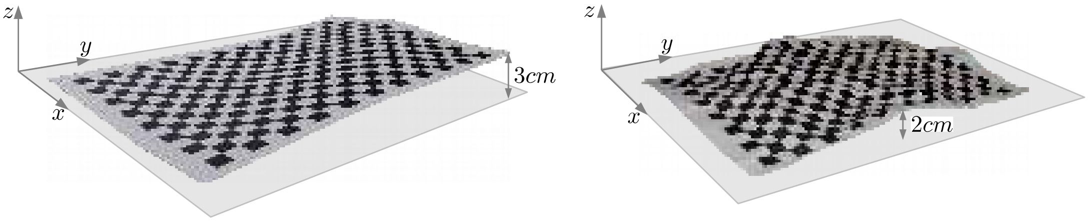


Figure 4.17: Color invariant feature mapping for camera responses from LSMSC **SC2$_{\mathbf{12C}}$ - DS4**: color patches from RGB sub-images L_0 to L_3 (row 1); physical model based feature images F_0 to F_3 (row 2); corresponding residuals with respect to L_0 (row 3); empirical model based feature images (row 4); corresponding residuals with respect to L_0 (row 5).

4.4.3 Block matching and sub-pixel displacement estimation

To experimentally evaluate the performance of block matching and sub-pixel displacement estimation, we have designed two checkerboard test scenes that were used as non-flat scanning surfaces. Photographic illustrations of the test scenes are given in Figure 4.18. The matching process used feature images obtained with the empirical model based color invariant feature mapping described in Section 4.3.1.B. We considered LSMSC $\mathbf{SC2_{12C}}$, so feature mapping model parameters were taken from the corresponding experiment in Section 4.4.2. Because the checkerboard contains only black and white patches and feature mapping for these colors generally performs well, we assumed that feature-mapping did not bias the experimental evaluation of block matching in this subsection.

The parameters for block-matching were empirically selected as $w = 20$, $b = 81$.



(a) Test scene 1 **(b)** Test scene 2

Figure 4.18: Illustration of the irregularly bended and the buckled checkerboard image scenes used to evaluate pixel correspondence estimation.

The result of block-matching and sub-pixel displacement estimation are disparity maps $\widetilde{D}_1$ to $\widetilde{D}_3$, computed from three feature image pairs (ie. (F_0, F_1), (F_0, F_2) and (F_0, F_3)). By reusing the automatic key-point extraction scheme described in Section 4.2.2, we can determine the sub-pixel accurate locations where black checker patterns intersect in the feature images F_0 to F_3. We define $\mathcal{K}_0$ to $\mathcal{K}_3$ as the sets of k key-points with locations $\{(x_i, y_i) \in \mathbb{R}^2 | 1 \leq x_i \leq c, 1 \leq y_i \leq r, i = 1, 2, ...k\}$, where r and c are the image height and width respectively. The checkerboard used in the test scenes spans a regular grid of 8×11 black patches before bending or buckling, from which we extracted $k = 280$ key-points. So we computed vectors of key-point displacement for feature image pairs from elements in $\mathcal{K}_0$ to $\mathcal{K}_3$. For instance, $\mathbf{d}_{r,u}(F_0, F_1) \in \mathbb{R}^{k \times 1}$ contains horizontal displacement values of key-point in $\mathcal{K}_0$ with respect to $\mathcal{K}_1$, and $\mathbf{d}_{r,v}(F_0, F_1) \in \mathbb{R}^{k \times 1}$ the corresponding vertical displacement.

We further resampled the disparity maps $\widetilde{D}_1$ to $\widetilde{D}_3$ at key-point pixel locations in $\mathcal{K}_0$ to obtain similar displacement vectors for the horizontal direction of the three pairs of feature images. We denote these vectors as $\mathbf{d}_{b,u}$ and note that an implication of the assumption that geometrically calibrated images are void of vertical displacement is $\mathbf{d}_{b,v} = 0$. We assume that key-point extraction is exact and hence consider the residual between the displacement obtained by correspondence estimation and the reference displacement obtained with the corner finder as a performance measure for block matching and sub-pixel displacement estimation. This error is $\mathbf{e}_u = \mathbf{d}_{r,u} - \mathbf{d}_{b,u}$ for horizontal direction and $\mathbf{e}_v = \mathbf{d}_{r,v} - \mathbf{d}_{b,v} = \mathbf{d}_{r,v}$ for vertical direction. For reference, we summarize the displacement and error terms used in Table 4.6.

The first order statistics of $\mathbf{e}_u$ and $\mathbf{e}_v$ for the two test scenes are given in Table 4.7. For the horizontal case, we illustrate in Figure 4.19 surface plots of spatially arranged

Table 4.6: Displacement and error terms used in the experimental evaluation of block-matching and sub-pixel displacement estimation.

Term		Description
$\mathbf{d}_{r,u}$		Key-point displacement in horizontal direction (computed with corner-finder).
$\mathbf{d}_{r,v}$		Same as $\mathbf{d}_{r,u}$, but for vertical direction.
$\mathbf{d}_{b,u}$		Key-point displacement in horizontal direction (computed from correspondence estimation / block matching).
$\mathbf{d}_{b,v}$		Same as $\mathbf{d}_{b,u}$, but for vertical direction.
$\mathbf{e}_u$	$= \mathbf{d}_{r,u} - \mathbf{d}_{b,u}$	Difference between key-point displacement computed from block matching and corner finder in horizontal direction.
$\mathbf{e}_v$	$= \mathbf{d}_{r,v} - \mathbf{d}_{b,v}$	Same as $\mathbf{e}_u$, but for vertical direction.

displacement $\mathbf{d}_{r,u}$, $\mathbf{d}_{b,u}$ and residuals $\mathbf{e}_u$ according to corresponding locations in $\mathcal{K}_0$ for **Test scene 1**. For **Test scene 2**, corresponding plots are shown in Figure 4.20.

We summarize the findings from this experiment as follows:

Overall performance: On average, sub-pixel accuracy is reached in correspondence estimation for all feature image pairs and both test scenes. The maximum errors are at most slightly larger than 1.1 pixels. The corresponding surface plots confirm that the low residual error is approximately randomly distributed over the image scene.

Comparison of image pairs: From surface plot of $\mathbf{d}_{r,u}$ and $\mathbf{d}_{b,u}$, we can see bigger displacement for image pairs for which the physical distance between sub-images is larger. The average errors seem to follow the same trend, so for (F_0, F_1) corresponding to lens 1 and lens 2, the average errors are always smaller as compared with (F_0, F_3) corresponding to lens 1 and lens 4.

Comparison of Test scene 1 and 2: The correspondence estimation performance is systematically lower for image pairs of **Test scene 2** than for **Test scene 1**. The block size in our experiment was selected rather large as compared with the local height variation of the buckled paper, which might have reduced the block matching accuracy. Using a smaller block size was not an option, as the block size must be bigger than individual checker pattern to allow block matching for this type of image scene. Obviously, this is a limitation related to our test scene design rather than the block matching procedure.

Comparison of horizontal and vertical displacement: The existence of error for vertical displacement is a strong indicator that the previous assumption that geometrically calibrated images are void of vertical displacement does not hold entirely. As discussed previously, one possible explanation for this residual is the lens viewpoint variation mentioned and illustrated in Figure 2.13b. For our particular system, we are not much concerned about that, as the overall displacement error is still in sub-pixel range.

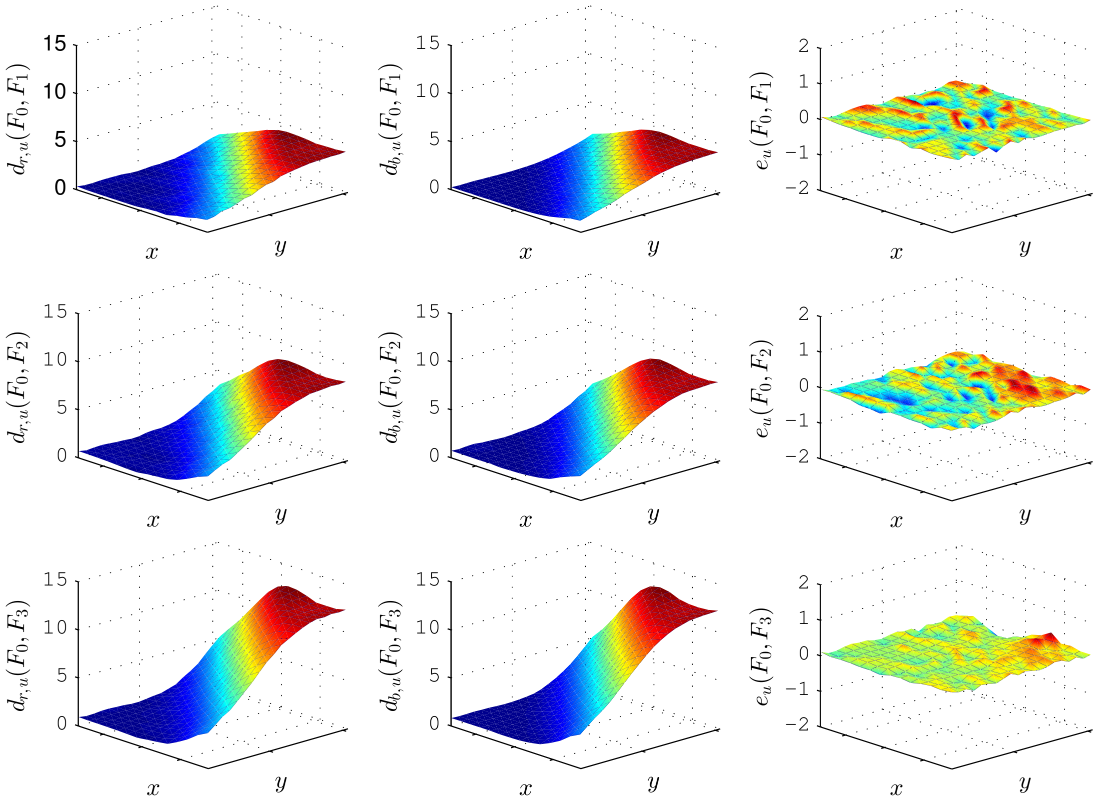


Figure 4.19: Surface plots of displacements $\mathbf{d}_{r,u}$ extracted via corner finder, and displacements $\mathbf{d}_{b,u}$ extracted via block-matching from **Test scene 1**. The last column corresponds to residuals $\mathbf{e}_u$, and the rows correspond to individual feature image pairs. The data points are spatially arranged according to their relative key-point locations in $\mathcal{K}$ and the surfaces are interpolated based on that.

Table 4.7: First order statistics of displacement error $\mathbf{e}_u$ and $\mathbf{e}_v$ for Test scenes 1 and 2.

(a) Test scene 1

$\mathbf{e}_u$	Avg.	Min.	Max.	Std.
(F_0, F_1)	0.037	0.000	0.127	0.030
(F_0, F_2)	0.075	0.000	0.245	0.052
(F_0, F_3)	0.094	0.001	0.270	0.048

$\mathbf{e}_v$	Avg.	Min.	Max.	Std.
(F_0, F_1)	0.074	0.024	0.153	0.025
(F_0, F_2)	0.136	0.000	0.391	0.112
(F_0, F_3)	0.137	0.034	0.309	0.061

(b) Test scene 2

$\mathbf{e}_u$	Avg.	Min.	Max.	Std.
(F_0, F_1)	0.131	0.001	0.673	0.137
(F_0, F_2)	0.150	0.001	0.682	0.129
(F_0, F_3)	0.204	0.001	0.982	0.183

$\mathbf{e}_v$	Avg.	Min.	Max.	Std.
(F_0, F_1)	0.055	0.001	0.940	0.071
(F_0, F_2)	0.117	0.001	1.107	0.107
(F_0, F_3)	0.105	0.001	0.630	0.063

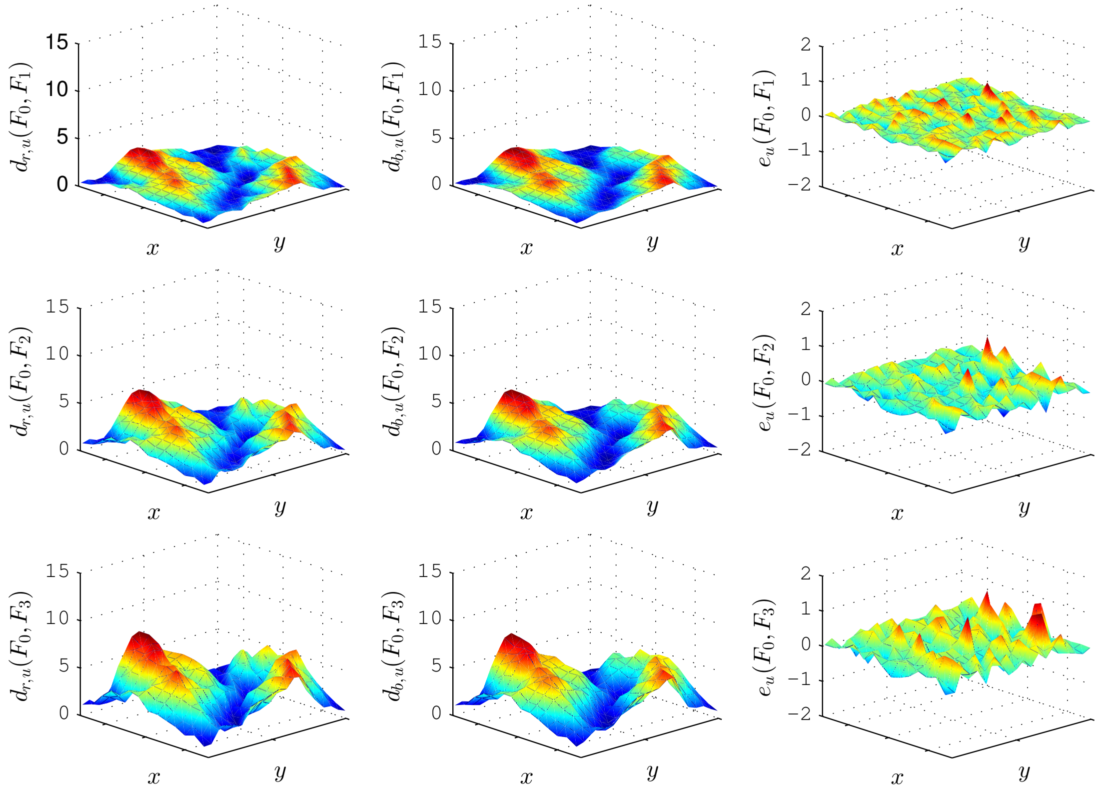


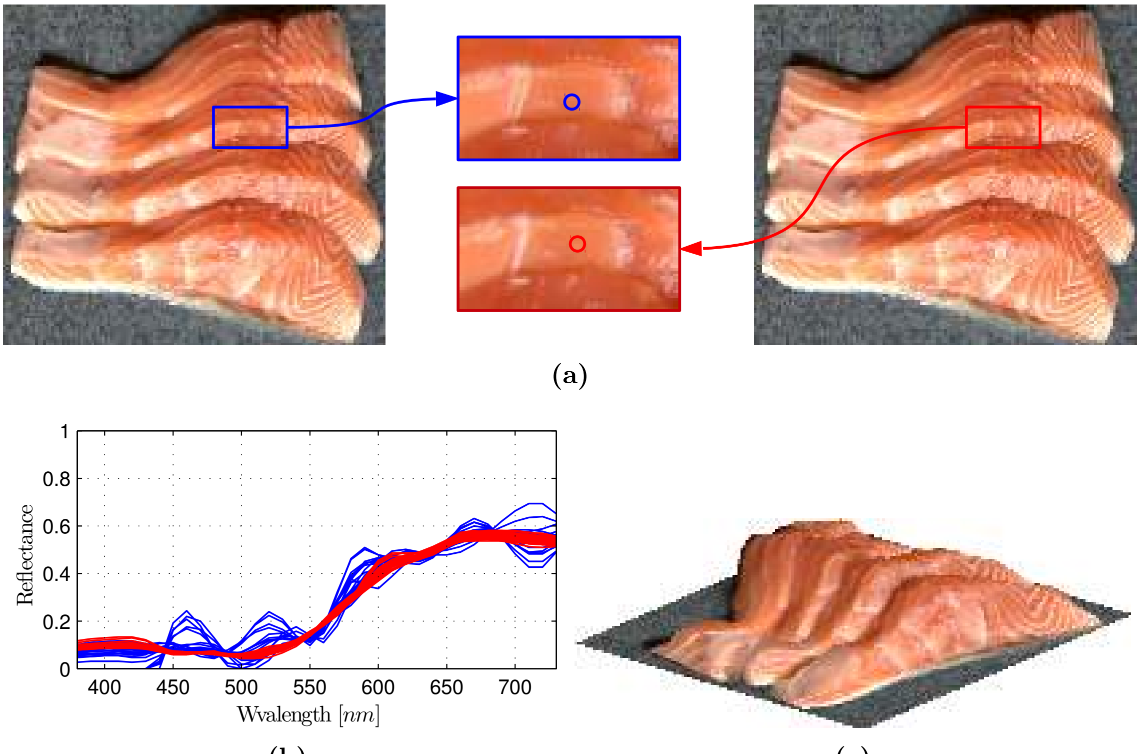
Figure 4.20: Surface plots of displacements $\mathbf{d}_{r,u}$ extracted via corner finder, and displacements $\mathbf{d}_{b,u}$ extracted via block-matching from **Test scene 2**. The last column corresponds to residuals $\mathbf{e}_u$, and the rows correspond to individual feature image pairs. The data points are spatially arranged according to their relative key-point locations in $\mathcal{K}$ and the surfaces are interpolated based on that.

4.4.4 Illustration of image registration for non-planar scanning objects

In the following, we consider examples of non-planar scanning scenes to demonstrate the proposed registration framework. RGB images were rendered from estimated reflectance data[9] (assuming the CIE-10°standard observer and CIE-D65 standard illuminant) and we compare geometrically calibrated images before and after scene-adaptive registration (Image registration B in the block diagram of Figure 4.2).

The first scanning scene is illustrated in Figure 4.21a. The zoomed region of the uncorrected image (marked in blue) shows severe color fringes at image locations where colors change abruptly, while this defect is not visible in the corrected image. The reflectance data corresponding to the area marked by the blue and red circles are plotted in Figure 4.21b. Even though these areas appear homogeneous in the rendered RGB images, we can see from the spectral data that reflectances estimated from uncorrected multi-channel image data (illustrated in blue) are oscillating much more than those estimated from corrected data (illustrated in red). We show in Figure 4.21c that our framework also allows reconstructing the three dimensional object structure from LSMSC image data. This illustration is obtained by mapping the rendered RGB image on a three-dimensional plot of the disparity map.

The second example refers to the color chart on a ramp that was used in Figure 4.10. For this example, the color fringe effect is very apparent.



(a)

(b) **(c)**

Figure 4.21: Test image scene salmon: RGB images rendered from estimated spectral data (a), spectral reflectances corresponding to the red and blue circles (b) and illustration of the three dimensional object surface structure recovered from LSMSC image data.

---

[9] Spectral estimation was performed using a linear least squares regression model with linear feature mapping. We note that spectral estimation will be considered in Chapter 5.

(a) Photography of image scene **(b)** Uncorrected image **(c)** Corrected image

Figure 4.22: Test image scene color chart on a ramp: the images in the middle and right are RGB images rendered from estimated spectral image data.

4.5 Summary, conclusions and future work

Image registration is a vitally important processing step in LSMSC image acquisition and is required to bring pixels corresponding to the same physical locations of the scanning scene in all image channels into alignment. We have distinguished registration procedures for two application cases, namely planar and non-planar scanning objects.

Registering images of planar scanning objects requires a geometrical calibration, which involves modeling horizontal and vertical pixel displacement with respect to a reference image channel, using image data from a predefined calibration scene target. Once obtained, the model can be used to determine the displacement required to align any other scanned images with sub-pixel accuracy.

We have developed a geometrical calibration target that is specifically useful for line-scan imaging systems, and proposed a semi-automatic scheme for dense key-point extraction over the entire sensor domain. This scheme was used to empirically analyze channel displacement for our laboratory LSMSC system ($\mathbf{SC2_{12C}}$), and allowed us to identify specific characteristics. It was found that the displacement can be up to 2 pixels and it is approximately polynomial shaped and distinct for each image channel. Based on these observations, we developed a registration framework for LSMSC systems and implemented two distinct models for displacement fitting: a polynomial model and a uniform cubic B-spline model. Sub-pixel accuracy was achieved by incorporating image resampling.

We evaluated the image registration performance of the two curve models with respect to alignment and fitting error and found that the residuals after registration are on average below 0.1 pixel for any channel in both horizontal and vertical directions. For the case of B-spline model based registration, we identified slightly larger maximum residual displacements when evaluating an independent test scene. We suspect that this model has a slight tendency to over-fit. Because the overall performance of both models is high, we did not analyze this aspect further.

When non-planar scanning objects are considered, geometrical calibration alone is not adequate when LSMSC systems with more than one lens are considered. The three-dimensional structure of non-planar surfaces results in scene dependent channel misalignment between RGB sub-images. This can be corrected in an adaptive registration process based on pixel correspondence estimation. We have discussed specific aspects and conditions of the correspondence problem with respect to LSMSC systems, which allowed predefining a search range for pixel correspondence among all image channels. A particularly challenging problem in finding pixel correspondence between RGB sub-images is that the image intensities relate to different spectral content due to the RGB sub-image filtering. We address this problem by transforming each RGB sub-image into a so-called color invariant feature image. Desired transformations produce feature images in which corresponding image object pixels share similar image intensities, and which retain most information of the RGB sub-images. We have proposed two approaches: the first is based on a physical model of the acquisition process, and the second is based on empirical data from a color chart. In experiments with two color test charts, it was found that the physical model based feature mapping is outperformed by the empirical model, which takes the distribution of camera responses into account.

After feature mapping, pixel-wise correspondence estimation is performed on feature images by block-matching with sub-pixel accuracy. In experiments with two test targets, we found average errors below 0.3 pixels and a maximum error only slightly larger than 1.1 pixel. The last step of LSMSC image registration for non-planar scanning objects is the pixel correspondence based image registration with linear interpolation image resampling.

We consider research related to LSMSC image registration of planar scanning objects to be adequately addressed. However, there exist various opportunities for future research with respect to non-planar scanning objects.

One aspect is related to the empirical method for color invariant feature mapping proposed in Section 4.3.1. This model takes the distribution of a set of camera responses from a color chart into account and we saw that this is useful in practice. If the actual image content of a scanning scene for a particular application is constrained to objects with certain colors, it might be possible to find a better transformation model than the one obtained from the color chart with many colors. We have conducted a preliminary study that involves unsupervised correspondence estimation of few image regions in RGB sub-images based on the so-called *Speeded-up robust features (SURF)* [82]. The camera responses of these image regions can be used to compute a color invariant feature transformation that is scene-adaptive and therefore likely to achieve better mapping than the non-adaptive transformation. Experimental results support this assumption, however, a systematic evaluation with different scene types is pending.

Further, we also experimented with non-linear feature mapping algorithms based on ridge regression in reproducing kernel Hilbert space [83, pp.232-234]. We used a Gaussian kernel and achieved results which indicate that the non-linear mapping performance is higher than linear feature mapping.

Another aspect that was not considered in this dissertation is 3D scene reconstruction from LSMSC image data. We used the disparity maps of RGB sub-images for registration, but as mentioned earlier, an additional geometrical calibration step would allow transforming pixel disparities in the image coordinate system to a height profile in the object coordinate system and hence allow spatial measurements on 3D objects.

The last aspect to mention is that image registration of non-planar scanning objects is also a fundamental requirement for spectral reflectance measurements on 3D objects. For complex object surface structure, measurement conditions will deviate from what is required for norm conform spectral reflectance factor and device independent color measurements. The influence on measurement performance has to be evaluated systematically.

"We have to remember that what we observe is not nature herself, but nature exposed to our method of questioning."

Werner Heisenberg (1901 – 1976)

5

Spectral reflectance recovery and estimation

Contents

5.1 Preface

The spectral reflectance estimation problem can be stated as follows: Given a camera response vector $\mathbf{x} = [x_1, ...x_n]^T \in \mathbb{R}^{n\times 1}$, and corresponding spectral reflectance vector $\mathbf{r} = [r_1, ...r_m]^T \in \mathbb{R}^{m\times 1}$, find a mapping $f : \mathbb{R}^{n\times 1} \mapsto \mathbb{R}^{m\times 1}$ that minimizes $\mathcal{L}(f(\mathbf{x}), \mathbf{r})$, where $\mathcal{L}$ is a loss function that quantifies the difference between $f(\mathbf{x})$ and $\mathbf{r}$, and typically $n < m$.

The task of finding the mapping function f can be solved by several fundamentally different strategies. One strategy is based on the physical process that relates spectral reflectances with camera responses, which is the image acquisition. If the forward mapping from reflectance space to camera response space is known, we can attempt to invert this mapping to find a f that relates camera responses with spectral reflectances. As this forward model maps from a typically higher dimensional space to a lower dimensional space, the inverse function is not unique, which is a major limitation in practice. Methods that follow this strategy are often referred to as *direct recovery* or *physical model based estimation* approaches.

Another strategy to find f does not make any assumptions on the image acquisition model used, but rather aims at modeling the relation of camera response and reflectance data empirically. Finding f is then considered as a pattern recognition or machine learning problem. The mapping from camera response to reflectance space is still from a lower dimensional to a higher dimensional space. But the learning process allows to include certain prior knowledge in the model, which is typically inferred from a known set of training data and/or physical constraints. Methods that follow this strategy are often referred to as *indirect recovery* or *empirical model based estimation* approaches. The two strategies are schematically illustrated in Figure 5.1.

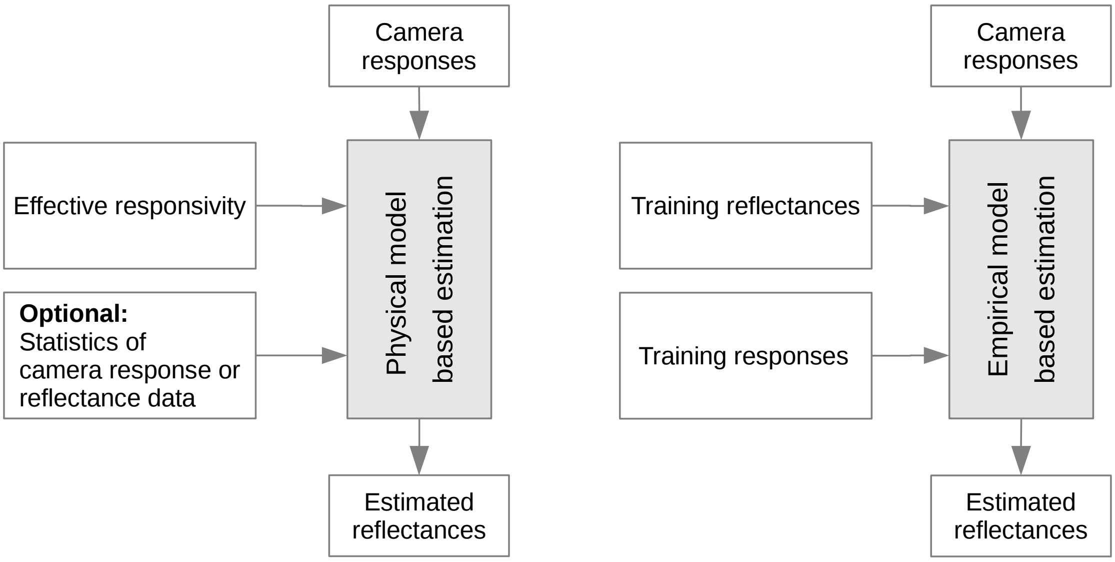


Figure 5.1: Block diagram of physical model and empirical model based estimation principles.

The active research in the spectral estimation domain in the past 20 years has led to the development of many approaches that follow either of the two strategies depicted above, or a mixture of both. We note that there exists many works related to theoretical aspects of spectral estimation and even more studies that apply spectral estimation to practical problems. Summarizing all of them is beyond the scope of this thesis and we therefore focus our attention to a general introduction of the underlying mathematical models in Sections 5.2 to 5.4. In Section 5.5, we compare the most prominent spectral estimation methods experimentally based on simulated and measured camera responses for various system configurations and various levels of simulated camera response noise. We conclude this chapter with a summary in Section 5.6.

Before we continue, we clarify that the terms *spectral reconstruction*, *spectral recovery* and *spectral estimation* refer to the same procedure, which is to obtain spectral reflectances from camera responses. To our knowledge, there exists no consensus in the scientific community about which term to use preferably. In this dissertation, we use the term spectral reconstruction exclusively when referring to the process of reconstructing reflectances from corresponding coordinates of an orthogonalized spectral reflectance subspace (see Section 2.2.4). The usage of the other terms is not specified.

5.2 Physical model based spectral estimation

We have introduced the general image transfer function in Section 2.3.3, which describes the integral process involved in image acquisition. We recall Equation 2.20:

$$\mathbf{x} = W^T\mathbf{r} + \boldsymbol{\epsilon}, \tag{5.1}$$

where $\mathbf{x} = [x_1, ...x_n]^T \in \mathbb{R}^{n\times 1}$ is a vector of camera responses, corresponding to surface reflectance ratio vector $\mathbf{r} = [r_1, ...r_m]^T \in \mathbb{R}^{m\times 1}$ and $W = \text{diag}(\mathbf{l})Y \in \mathbb{R}^{m\times n}$ is the effective camera responsivity, describing the combined effect of scene illumination $\mathbf{l} \in \mathbb{R}^{m\times 1}$ and camera responsivity $Y \in \mathbb{R}^{m\times n}$. Vector $\boldsymbol{\epsilon} \in \mathbb{R}^{n\times 1}$ denotes an additive noise term.

We define physical model based spectral estimation as the problem of finding a reconstructed reflectance $\widetilde{\mathbf{r}} \in \mathbb{R}^{m\times 1}$ corresponding to camera response $\mathbf{x} \in \mathbb{R}^{n\times 1}$ that satisfy Equation 5.1. If W is known and we assume the absence of noise (ie. $\boldsymbol{\epsilon} = \mathbf{0} \in \mathbb{R}^{n\times 1}$), $\widetilde{\mathbf{r}}$ can be reconstructed directly from $\mathbf{x}$ by inverting W. The corresponding system equation is

$$\mathbf{x} = W^T\mathbf{r}. \tag{5.2}$$

Because W is seldom a square matrix with full rank, the direct inverse usually does not exist. In fact, in most practical applications of multi-spectral imaging, the number of camera spectral channels is smaller than the number of discrete wavelengths of the spectral data, i.e. $n < m$. This means that the system $\mathbf{x} = W^T\mathbf{r}$ is underdetermined and there exist infinite possible $\widetilde{\mathbf{r}}$ vectors that satisfy Equation 5.2 (ie. which are metamers in camera response space). In the following, we discuss some particular solutions that are often considered in practice.

5.2.1 Minimum norm solution

As mentioned before, there exist infinite solutions that satisfy Equation 5.2 and a specific one is the one with minimum norm. Accordingly, the problem can be stated as:

$$\begin{gathered}\arg\min_{\mathbf{r}} \|\widetilde{\mathbf{r}}\|^2 \\ \text{s.t. } \mathbf{x} = W^T\widetilde{\mathbf{r}}\end{gathered}$$

It can be solved by the method of Lagrange multipliers. If the column vectors of W are linearly independent (ie. rank(W) = n), the solution takes the form

$$\widetilde{\mathbf{r}} = W(W^TW)^{-1}\mathbf{x} = (W^T)^+\mathbf{x}, \tag{5.3}$$

where $(W^T)^+$ is called the pseudoinverse of W^T. This solution is based on the Moore-Penrose pseudoinverse, which generalizes the notion of matrix inverse to non-square matrices [84, p.142]. In the following, we refer to this method as **P-MIN**.

Related literature
Maloney and Wandell [17] have used the minimum norm solution of the physical model based spectral estimation. Instead of estimating spectral reflectances, they estimated the coordinates of reflectance data transformed in a truncated orthogonalized reflectance space and subsequently reconstructed the reflectances from these coordinates. The orthogonal basis of the reflectance space contained in matrix $U \in \mathbb{R}^{m\times m}$ can be derived from a set of known reflectances by SVD or PCA (Section 2.2.4), and the truncated orthogonal space is spanned by the first b vectors in U, denoted by $\hat{U} \in \mathbb{R}^{m\times b}$, which carry the most variance of the reflectance data. By replacing $\mathbf{r} \in \mathbb{R}^{m\times 1}$ in Equation 5.2 by the reconstructed spectral reflectance, ie. the corresponding orthogonal subspace representation $\hat{\mathbf{r}} = \hat{U}\hat{\mathbf{c}} \in \mathbb{R}^{m\times 1}$ (2.9), we obtain

$$\mathbf{x} \approx \hat{\mathbf{x}} = W^T\hat{U}\hat{\mathbf{c}}, \tag{5.4}$$

where $\mathbf{x} \in \mathbb{R}^{n\times 1}$ is the camera response corresponding to $\mathbf{r}$, $\hat{\mathbf{x}} \in \mathbb{R}^{n\times 1}$ is the camera response corresponding to the reconstructed spectral reflectance $\hat{\mathbf{r}}$, and $\hat{\mathbf{c}} \in \mathbb{R}^{b\times 1}$ is the truncated coefficient vector corresponding to $\hat{\mathbf{r}}$. The scalar b is a parameter to be optimized in the model and can take values from 1 to m. Note that if $b = m$, the approximation sign becomes equality and the expression equivalent to Equation 5.2.

The solution for Equation 5.4 is given by the pseudoinverse

$$\hat{\hat{\mathbf{c}}} = (W^T\hat{U})^+\mathbf{x}, \tag{5.5}$$

where $\hat{\hat{\mathbf{c}}} \in \mathbb{R}^{b\times 1}$ is an estimated coefficient vector and the other terms are as before. Thus, we can reconstruct an estimated spectral reflectance $\hat{\hat{\mathbf{r}}} \in \mathbb{R}^{m\times 1}$ from $\hat{\hat{\mathbf{c}}}$ as

$$\hat{\hat{\mathbf{r}}} = \hat{U}\,\hat{\hat{\mathbf{c}}} = \hat{U}(W^T\hat{U})^+\mathbf{x}. \tag{5.6}$$

In this method, the reconstruction error is defined as $RMSE(\mathbf{r}, \hat{\mathbf{r}})$, while the estimation error is defined as $RMSE(\mathbf{r}, \hat{\hat{\mathbf{r}}})$. In the following, we refer to this method as **P-MW**.

5.2.2 Wiener estimation

The minimum norm solution to the physical model based estimation problem is not necessarily the desired one. This solution results in zero residuals in camera response space, ie. the estimated reflectance $\widetilde{\mathbf{r}}$ has the same camera response as its true counterpart $\mathbf{r}$. However, what we are really interested in is to minimize the error $\|\widetilde{\mathbf{r}} - \mathbf{r}\|^2$. Further, the solutions developed previously are based on the assumption of absence of noise, which is generally not true in image acquisition.

In Wiener estimation, we consider a linear estimation model $\widetilde{\mathbf{r}} = A\mathbf{x}$. The problem can be defined as [85]

$$\arg\min_{A} E[\|A\mathbf{x} - \mathbf{r}\|^2], \tag{5.7}$$

where $A \in \mathbb{R}^{m\times n}$ is the estimation matrix, $\mathbf{x} \in \mathbb{R}^{n\times 1}$ is a multivariate random variable of camera response and $\mathbf{r} \in \mathbb{R}^{m\times 1}$ the corresponding spectral reflectance variable. $E[\cdot]$ denotes the expectation. Obviously, $\mathbf{r}$ is unknown in practice.

Assuming that reflectances and noise are normal distributed multivariate random variables that are statistical independent, the optimal linear estimation model is based on the so-called Wiener filter [86]. The solution is then given by the matrix

$$A_{WN} = K_{\mathbf{r}} W (W^T K_{\mathbf{r}} W + K_{\mathbf{e}})^{-1}, \tag{5.8}$$

where $A_{WN} \in \mathbb{R}^{m \times n}$, and $W \in \mathbb{R}^{m \times n}$ is as before, $K_{\mathbf{r}} \in \mathbb{R}^{m \times m}$ is the covariance matrix of reflectances, and $K_{\mathbf{e}} \in \mathbb{R}^{n \times n}$ the covariance matrix of the camera response noise.

A reflectance vector $\widetilde{\mathbf{r}} \in \mathbb{R}^{m \times 1}$ can now be estimated from $\mathbf{x}$ by evaluating

$$\widetilde{\mathbf{r}} = A_{WN} \mathbf{x}. \tag{5.9}$$

We refer to this method as **P-WN**.

For various reasons, this model is prevented from having optimal performance [87]. Firstly, natural reflectances do not follow a normal distribution [47] and covariance matrix $K_{\mathbf{r}}$ is usually unknown. In practice, $K_{\mathbf{r}}$ can be estimated. The unbiased estimate of the covariance matrix of a random variable is described in Appendix A.4.

Furthermore, camera response noise is not statistically independent from reflectances (for instance due to signal dependent shot noise. See Section 2.3.1.B.), nor is it normally distributed. In practice, second order statistics of camera response noise can either be assumed or approximated from measurements. Assuming it is an uncorrelated random variable, we have $K_{\mathbf{e}} = \mathrm{diag}(\boldsymbol{\sigma})$, where $\boldsymbol{\sigma} = [\sigma_1, \sigma_2, ...\sigma_n]^T \in \mathbb{R}^{n \times 1}$ is a vector of n channel dependent noise variances.

Related literature

To our knowledge, Pratt *et al.* [86] were the first who used the Wiener estimation model for spectral estimation in the imaging domain. Their work can be considered trailblazing in this field, as it appeared already in 1976 and therefore much earlier than most other works related to spectral estimation.

The aforementioned practical limitation of the Wiener model due to the signal dependence of camera noise was recognized by Haneishi *et al.* [49]. They proposed to divide the camera response space into sub-blocks for which they precomputed various $K_{\mathbf{e}}$ matrices. The performance obtained with this approach was found to be higher than for assuming signal independent noise.

Shimano [88] developed a model for estimating the noise variance of a multi-spectral imaging system and showed that this model agrees fairly well with the optimal variance with respect to the Wiener filter model.

5.3 Empirical model based spectral estimation

Empirical estimation is a form of supervised learning, which means that the mathematical relation between reflectances and the corresponding camera responses generated by a multi-spectral acquisition system are obtained from a set of observation data. We refer to this set as *training set* and denote it as $\mathcal{S}_{tr} = \{(\mathbf{x}_1, \mathbf{r}_1), (\mathbf{x}_2, \mathbf{r}_2), ...(\mathbf{x}_l, \mathbf{r}_l)\} \subset \mathbb{R}^{n \times 1} \times \mathbb{R}^{m \times 1}$, where $(\mathbf{x}_i, \mathbf{r}_i)$ is a corresponding pair of camera response vector and surface reflectance, and $i = 1, 2, ...l$.

We can express empirical estimation in form of the following minimization problem:

$$\arg\min_f \left(\frac{1}{2} \sum_{i=1}^{l} \mathcal{L}(f(\mathbf{x}_i) - \mathbf{r}_i) \right), \tag{5.10}$$

which can be interpreted as the problem of finding a function $f : \mathbb{R}^n \mapsto \mathbb{R}^m$ that minimizes the cumulative estimation loss $\mathcal{L}$ over all elements in $\mathcal{S}_{tr}$. We note that the factor $\frac{1}{2}$ is inserted only for convenience of calculations.

The following solutions to this minimization problem are based on linear systems. Let's start by considering the scalar case

$$\widetilde{r}_\lambda = f_\lambda(\mathbf{x}, \mathbf{a}_\lambda) = \mathbf{a}_\lambda \mathbf{x} = a_{\lambda,0} + a_{\lambda,1}x_1 + ... + a_{\lambda,n}x_n, \tag{5.11}$$

where $\mathbf{x} = [x_1, x_2, ...x_n]^T \in \mathbb{R}^{n\times 1}$ is a vector of camera responses, $\mathbf{a}_\lambda = [a_{\lambda,0}, a_{\lambda,1}, ...a_{\lambda,n}] \in \mathbb{R}^{1\times(n+1)}$ is a vector containing coefficients of the linear system, $f_\lambda(\mathbf{x}, \mathbf{a}_\lambda)$ is the prediction function for a single wavelength component $\widetilde{r}_\lambda$ of the estimated reflectance vector $\widetilde{\mathbf{r}} = [\widetilde{r}_1, \widetilde{r}_2, ...\widetilde{r}_m]^T \in \mathbb{R}^{m\times 1}$, and $\lambda = 1, 2, ...m$.

This model is characterized by the property that the prediction is a linear function of the input variables $x_1, ...x_n$, which in some cases might restrict the model severely. However, we can extend it by introducing a potentially non-linear mapping $\boldsymbol{\phi}$ that maps camera response vector $\mathbf{x}$ to a p-dimensional feature space [84, p. 138]. Accordingly, the prediction function is

$$\widetilde{r}_\lambda = f_\lambda(\mathbf{x}, \mathbf{a}_\lambda) = \mathbf{a}_\lambda \boldsymbol{\phi}(\mathbf{x}), \tag{5.12}$$

where $\mathbf{a}_\lambda \in \mathbb{R}^{1\times p}$. Feature mapping $\boldsymbol{\phi}$ can be considered as an array of p basis functions $[\phi_0, \phi_1, ...\phi_{p-1}]$, with each function mapping the camera response to a 1 dimensional space. We define $\phi_0 = 1$ to allow for a fixed bias, like in Equation 5.11.

This prediction function is able to model a non-linear relation between r_λ and $\mathbf{x}$, while the relation between r_λ and $\boldsymbol{\phi}(\mathbf{x})$ remains linear. This allows us to use linear techniques to solve the problem.

So far, we have considered prediction of a single wavelength component $\widetilde{r}_\lambda$ (scalar case), but at last we want to estimate each wavelength component comprising reflectance vector $\widetilde{\mathbf{r}}$. We can extend Equation 5.12 to the vectorial case

$$\widetilde{\mathbf{r}} = f(\mathbf{x}, \mathbf{A}) = A\boldsymbol{\phi}(\mathbf{x}), \tag{5.13}$$

where matrix $A = [\mathbf{a}_1^T, \mathbf{a}_2^T, ...\mathbf{a}_m^T]^T \in \mathbb{R}^{m\times p}$ contains m coefficient vectors for the linear system.

5.3.1 Least-squares solution

Before we can develop a solution for the minimization problem in Equation 5.10, we need to define the function $\mathcal{L}$ that quantifies the estimation loss. Recall the discussion in Section 2.4.1 related to spectral estimation error. Let us assume that the estimation error for each wavelength component in $\widetilde{\mathbf{r}}$ is equally important. A corresponding measure of spectral residuals is *RMSE* and the related loss function can therefore be defined as the Euclidean norm $\|\widetilde{\mathbf{r}} - \mathbf{r}\|$. Now we can rewrite Equation 5.10 and obtain

$$\arg\min_A \left(\frac{1}{2} \sum_{i=1}^{l} \|A\ \boldsymbol{\phi}(\mathbf{x}_i) - \mathbf{r}_i\|^2 \right), \tag{5.14}$$

which is equivalent to

$$\arg\min_A \left(\frac{1}{2} \operatorname{Tr} \left\{ (A\Phi(X) - R)^T (A\Phi(X) - R) \right\} \right), \tag{5.15}$$

where $A \in \mathbb{R}^{m \times p}$ is as before and matrix $\Phi(X) \in \mathbb{R}^{p \times l}$ contains the evaluation of p basis functions for l camera responses from $\mathcal{S}_{tr}$. Further, matrix $X \in \mathbb{R}^{n \times l}$ and $R \in \mathbb{R}^{m \times l}$ contain these camera responses and the corresponding reflectances in columns. Accordingly

$$\Phi(X) = \begin{pmatrix} \phi_0(\mathbf{x}_1) & \phi_0(\mathbf{x}_2) & \cdots & \phi_0(\mathbf{x}_l) \\ \phi_1(\mathbf{x}_1) & \phi_1(\mathbf{x}_2) & \cdots & \phi_1(\mathbf{x}_l) \\ \vdots & \vdots & \ddots & \vdots \\ \phi_{p-1}(\mathbf{x}_1) & \phi_{p-1}(\mathbf{x}_2) & \cdots & \phi_{p-1}(\mathbf{x}_l) \end{pmatrix}. \tag{5.16}$$

In the following, we omit X from $\Phi(X)$ for the sake of simplicity. A closed form solution for the minimization problem in Equation 5.15 can be obtained by setting the derivative of the cost function with respect to A equal to zero and then solve for A to obtain[1]

$$A = R\Phi^T(\Phi\Phi^T)^{-1} = R\Phi^+. \tag{5.17}$$

The dimensions of Φ, namely p and l characterize the type of this linear system. An inverse of $\Phi\Phi^T$ exists, if and only if rank$(\Phi) = p$ and $p \leq l$, which means that $\Phi\Phi^T$ is fully determined. In spectral estimation, typically $l > p$, which means that there are more reflectance samples and corresponding camera response samples to describe the system, than the number of dimensions of the feature space that embeds the camera responses.

At this point, we remind the reader that the quantity $\Phi^+ = \Phi^T(\Phi\Phi^T)^{-1}$ is known as the *Moore-Penrose pseudoinverse* of matrix Φ. This is why the linear least squares solution to the minimization problem in Equation 5.15 is often referred to as the pseudoinverse solution. We note that the pseudoinverse here has a different form from Equation 5.3 in the physical model based estimation problem. In that case, the pseudoinverse gives a solution $\widetilde{\mathbf{r}}$ with minimum norm that satisfies Equation 5.2. In case of the empirical estimation problem with $l > p$, the linear system $A\Phi = R$ is overdetermined, and the pseudoinverse gives the solution A that minimizes the estimation loss (sum of squared errors) over training samples in $\mathcal{S}_{tr}$.

Instead of using the closed form solution in Equation 5.17 to compute the pseudoinverse, we use SVD [89]. Assuming that $\Phi = U\Sigma V^T$ is the singular value decomposition of $\Phi \in \mathbb{R}^{p \times l}$, then

$$\Phi^+_{SVD} = V\Sigma^+U^T, \tag{5.18}$$

where $\Phi^+_{SVD} \in \mathbb{R}^{l \times p}$, the i-th column of $U \in \mathbb{R}^{p \times p}$ and $V \in \mathbb{R}^{l \times l}$ are the left and right singular vectors of Φ, corresponding to the singular value σ_i in matrix $\Sigma \in \mathbb{R}^{p \times l}$, and

$$\Sigma^+ = \text{diag}(\boldsymbol{\sigma}^+),\ \boldsymbol{\sigma}^+ = [\sigma_1^+, \sigma_2^+, ...\sigma_p^+]^T, \text{where } \sigma_i^+ = \begin{cases} 1/\sigma_i & \text{if } \sigma_i \neq 0 \\ 0 & \text{otherwise} \end{cases}, i = 1, 2, ...p. \tag{5.19}$$

A Linear feature mapping

In the simplest case, we assume the relation between camera responses and reflectance components to be linear. Hence, let $\boldsymbol{\phi}(\mathbf{x}) = \mathbf{x} \in \mathbb{R}^{n \times 1}$ and equivalently, $\Phi(X) = X \in \mathbb{R}^{n \times l}$. Equation 5.17 becomes

$$A_{LIN} = RX^T(XX^T)^{-1} = RX^+, \tag{5.20}$$

where $A_{LIN} \in \mathbb{R}^{m \times n}$, matrix $X \in \mathbb{R}^{n \times l}$ contains camera responses of $\mathcal{S}_{tr}$, matrix $R \in \mathbb{R}^{m \times l}$ contains the corresponding spectral reflectances and $X^+ \in \mathbb{R}^{l \times n}$ is the pseudoinverse of

---

[1] We derive this expression in Appendix A.2.

X. We can estimate reflectance $\widetilde{\mathbf{r}} \in \mathbb{R}^{m\times 1}$ corresponding to camera response $\mathbf{x} \in \mathbb{R}^{n\times 1}$ by computing

$$\widetilde{\mathbf{r}} = A_{LIN}\mathbf{x}. \tag{5.21}$$

In the spectral estimation literature, this model is also referred to as *Pseudoinverse*, *Linear Regression* or *Linear Least-Square*. We use the SVD approach to compute the pseudoinverse and therefore refer to this method as **E-PINV-SVD**.

Related literature

The least squares solution with linear feature mapping has been used for spectral reflectance estimation by various researchers, either as a solution to a specific application, or to compare with other spectral estimation approaches (eg. in [85, 90, 91, 92, 23, 93, 94, 95]). Sometimes it has been termed *Wiener estimation* [96, 97, 98].

A related approach to spectral reflectance estimation using least squares regression with linear feature mapping is a modified approach that estimates coordinates in an orthogonalized spectral reflectance space (see Section 2.2.4). Instead of finding matrix A_{LIN} that minimizes the Euclidean distance between $A\mathbf{x}$ and $\mathbf{r}$, we find a matrix A_{PCA} such that

$$\arg\min_{A_{PCA}} \left(\frac{1}{2} \sum_{i=1}^{l} \| A_{PCA}\mathbf{x}_i - \mathbf{c}_i \|^2 \right), \tag{5.22}$$

where $A_{PCA} \in \mathbb{R}^{m\times n}$ is an estimation matrix, vector $\mathbf{c}_i = U\mathbf{r}_i \in \mathbb{R}^{m\times 1}$ is the i-th coefficient vector corresponding to reflectance $\mathbf{r}_i \in \mathbb{R}^{m\times 1}$ and $i = 1, 2, ...l$. Vector $\mathbf{c}_i$ and the orthogonal matrix $U \in \mathbb{R}^{m\times m}$ can be obtained by performing PCA or SVD on a matrix of training reflectances from $\mathcal{S}_{tr}$.

The solution to this minimization in least squares sense is

$$A_{PCA} = CX^{+}, \tag{5.23}$$

where matrix $X^{+} \in \mathbb{R}^{l\times n}$ is the pseudoinverse of the matrix of training camera responses $X \in \mathbb{R}^{l\times n}$, and $C \in \mathbb{R}^{m\times l}$ is a matrix of coefficient vectors corresponding to reflectances in matrix $R \in \mathbb{R}^{m\times l}$.

We can now use A_{PCA} to estimate a coefficient vector $\widetilde{\mathbf{c}} = A_{PCA}\ \mathbf{x} \in \mathbb{R}^{m\times 1}$ corresponding to camera response $\mathbf{x} \in \mathbb{R}^{n\times 1}$ in the orthogonal reflectance space. According to Equation 2.7, reflectance $\widetilde{\mathbf{r}} \in \mathbb{R}^{m\times 1}$ can be reconstructed by evaluating $U\ \widetilde{\mathbf{c}}$, so the estimated reflectance can be computed as

$$\widetilde{\mathbf{r}} = UA_{PCA}\ \mathbf{x}. \tag{5.24}$$

Estimation of coefficients in the orthogonal reflectance space and subsequent back-projection to reflectance space is equivalent to the direct least squares estimation in reflectance space. However, instead of considering all basis components of the orthogonalized reflectance space, a truncated version can be used, which was found by various researchers to be more robust with respect to camera response noise [18, 99].

B Nonlinear feature mapping

Instead of considering a linear mapping of the input vector as in the previous case, researchers in the color science field have also explored nonlinear feature mappings. A popular choice is a polynomial expansion of the input vectors using various basis functions.

An example of a polynomial expansion for camera response vector $\mathbf{x} = [x_1, x_2]^T \in \mathbb{R}^{2\times 1}$ is

$$\begin{aligned}\boldsymbol{\phi}(\mathbf{x}) &= [\phi_0(\mathbf{x}), \phi_1(\mathbf{x}), \phi_2(\mathbf{x}), \phi_3(\mathbf{x}), \phi_4(\mathbf{x}), \phi_5(\mathbf{x})]^T \\ &= [1, x_1, x_2, x_1x_2, x_1^2, x_2^2]^T,\end{aligned} \tag{5.25}$$

where $\boldsymbol{\phi}(\mathbf{x}) \in \mathbb{R}^{6\times 1}$ is a vector containing the evaluation of basis functions $\phi_0(\mathbf{x})$ to $\phi_5(\mathbf{x})$.

In related literature, this approach is often referred to as polynomial, or ordinary polynomial regression [100, 101, 102]. The number of basis functions in a polynomial expansion should be much smaller than the number of observations (ie. the elements in the training set). In this case, the system is overdetermined and a solution given in Equation 5.17. Using non-linear feature mapping with a polynomial expansion similar to the example above can be computationally expensive for imaging systems with many channels, because the expansion is modeled by many basis functions.

Related literature

Hong, Luo and Rhodes noted that the generalization of the linear model using polynomial expansion improved with increasing number of training samples [100]. Heikkinen *et al.* reported that for cases with few training data, the resulting polynomial model tended to overfit and consequently could lead to physically unreasonable estimates and oscillation. They motivated additional regularization of the solution to account for this limitation [103].

5.3.2 Regularized solutions

The linear model based spectral estimation problem, ie. finding matrix A to satisfy the overdetermined system defined by $A\Phi = R$ (see Equation 5.15), is an ill-posed problem. Regularization is needed to reformulate the least squares problem in Equation 5.14 for numerical treatment. *Tikhonov regularization* is one of the commonly used approaches. It gives preference to solutions with smaller norms by modifying the minimization problem to the following form

$$\arg\min_{A} \left(\frac{1}{2} \sum_{i=1}^{l} \|A\boldsymbol{\phi}(\mathbf{x}_i) - \mathbf{r}_i\|^2 + \frac{1}{2}\lambda \|A\|_F^2 \right), \tag{5.26}$$

where matrix $A \in \mathbb{R}^{m\times p}$, function $\boldsymbol{\phi} : \mathbb{R}^n \mapsto \mathbb{R}^p$ defines the feature mapping, $\lambda \in \mathbb{R}^+$ is a regularization parameter and $\|A\|_F$ denotes the Frobenius norm. An explicit solution[2] is given by

$$A_{REG} = R\Phi^T(\Phi\Phi^T + \lambda I_p)^{-1}, \tag{5.27}$$

where $\Phi = \Phi(X) \in \mathbb{R}^{p\times l}$ is as before and $I_p \in \mathbb{R}^{p\times p}$ is the identity matrix. This particular case with quadratic regularizer is called *ridge regression* [84].

Another form of regularization is based on the SVD approach. According to Neumaier [89], ill-conditioned matrices are characterized by tiny singular values, which can lead to magnification of measurement errors. An improvement of the situation can be achieved by replacing the matrix Σ^+ in Equation 5.18 by a truncated version

$$\Sigma_h = \mathrm{diag}(\widetilde{\boldsymbol{\sigma}}),\ \widetilde{\boldsymbol{\sigma}} = [\widetilde{\sigma}_1, \widetilde{\sigma}_2, ...\widetilde{\sigma}_n]^T, \text{where } \widetilde{\sigma}_i = \begin{cases} 1/\sigma_i & \text{for } \sigma_i \geq h \\ 0 & \text{otherwise,} \end{cases}, i = 1, 2, ...p \tag{5.28}$$

---

[2] Note that the solution can be derived in a similar manner as for Equation 5.17, see Appendix A.2.

where h is a threshold parameter. The corresponding regularized SVD solution to the least squares problem is then

$$A_{tSVD} = RV\Sigma_h U^T, \tag{5.29}$$

where $A_{tSVD} \in \mathbb{R}^{m \times p}$, the i-th column of $U \in \mathbb{R}^{p \times p}$ and $V \in \mathbb{R}^{l \times l}$ are the left and right singular vectors of Φ, corresponding to the singular value σ_i in matrix $\Sigma \in \mathbb{R}^{p \times l}$.

Note that this approach is similar with the aforementioned estimation of coordinates in a truncated orthogonalized spectral reflectance space.

Remarks on the implementation

The truncated SVD solution for the pseudoinverse of a matrix X is readily available in Matlab when using the $pinv(X, h)$ function, where h refers to the threshold parameter described in Equation 5.28. Matlab also provides a heuristic to determine h automatically, which is used when calling $pinv(X)$ without additional input arguments apart from the matrix to be inverted. This heuristic excludes small singular values that are close to the numerical accuracy of floating point computations for a given variable precision.

In the practical case of spectral estimation with noise contaminated camera response data, we might be interested in regularizing the solution beyond the level required to stabilize the numerical computation. Hence, h can be considered as a free model parameter that can be optimized with respect to actual application data to increase estimation performance.

5.3.3 Reproducing kernel Hilbert space based solution

Recall that the solution to the regularized least squares regression problem in Equation 5.27 requires the explicit computation of the feature mapping $\boldsymbol{\phi}(\mathbf{x}) = [\phi_1(\mathbf{x}), \phi_2(\mathbf{x}), ...\phi_p(\mathbf{x})]^T \in \mathbb{R}^{p \times 1}$ of camera response $\mathbf{x} \in \mathbb{R}^{n \times 1}$. This feature mapping is often nonlinear and can be infinite dimensional. The explicit mapping can be avoided by using the kernel trick, which refers to modifying the regression algorithm such that only pairwise inner-products of the observation data are required [83].

According to the works of Heikkinen *et al.* [91, 103, 104] and by assuming that the feature mapping is induced by a positive definite kernel function $k(\cdot, \cdot)$, the minimization problem in Equation 5.26 is equivalent to

$$\arg\min_{A} \left(\sum_{i=1}^{l} \|A\mathbf{k}(\mathbf{x}_i) - \mathbf{r}_i\|^2 + \lambda \operatorname{Tr}(AKA^T) \right), \tag{5.30}$$

where $A \in \mathbb{R}^{m \times l}$ is a matrix of weight vectors and operator $\operatorname{Tr}(\cdot)$ denotes matrix trace. The vector $\mathbf{k}(\mathbf{x}_i) = [k(\mathbf{x}_1, \mathbf{x}_i), ...k(\mathbf{x}_l, \mathbf{x}_i)]^T \in \mathbb{R}^{l \times 1}$ contains kernel evaluations between sample $\mathbf{x}_i \in \mathbb{R}^{n \times 1}$ and the camera responses in the training set, and scalar $\lambda \in \mathbb{R}^+$ is a regularization parameter. The kernel matrix of training data $K \in \mathbb{R}^{l \times l}$ is defined as

$$K = \begin{pmatrix} k(\mathbf{x}_1, \mathbf{x}_1) & k(\mathbf{x}_1, \mathbf{x}_2) & \cdots & k(\mathbf{x}_1, \mathbf{x}_l) \\ k(\mathbf{x}_2, \mathbf{x}_1) & k(\mathbf{x}_2, \mathbf{x}_2) & \cdots & k(\mathbf{x}_2, \mathbf{x}_l) \\ \vdots & \vdots & \ddots & \vdots \\ k(\mathbf{x}_l, \mathbf{x}_1) & k(\mathbf{x}_l, \mathbf{x}_2) & \cdots & k(\mathbf{x}_l, \mathbf{x}_l) \end{pmatrix}. \tag{5.31}$$

The minimization problem can be solved by setting the derivative of Equation 5.30 with respect to A equal to zero to obtain

$$A_{KER} = R(K + \lambda I_l)^{-1}, \tag{5.32}$$

where $A_{KER} \in \mathbb{R}^{m\times l}$, matrix $I_l \in \mathbb{R}^{l\times l}$ is the identity matrix and $R = [\mathbf{r}_1, ...\mathbf{r}_l] \in \mathbb{R}^{m\times l}$ the matrix of training reflectances. We can estimate reflectance $\widetilde{\mathbf{r}} \in \mathbb{R}^{m\times 1}$ corresponding to camera response $\mathbf{x}$ by computing

$$\widetilde{\mathbf{r}} = A_{KER}\ \mathbf{k}(\mathbf{x}). \tag{5.33}$$

In the following, we refer to the reproducing kernel Hilbert space (RKHS) ridge-regression estimation as **E-KER**.

Positive definite kernels in ridge regression

The definition of a positive definite (PD) function is as follows [105, p.67]: Let X be a closed and bounded subset of $\mathbb{R}^{n\times 1}$. A function k is called positive definite if $\sum_{i,j=1}^{l} \alpha_i \alpha_j k(\mathbf{x}_i, \mathbf{x}_j) \geq 0$, for all $l \in \mathbb{N}$, $\mathbf{x}_i, \mathbf{x}_j \in X$ and $\alpha_i, \alpha_j \in \mathbb{R}$.

The simplest kernel function is the linear kernel [106] defined as

$$k(\mathbf{x}_i, \mathbf{x}_j) = \mathbf{x}_i^T \mathbf{x}_j, \tag{5.34}$$

where $\mathbf{x}_i, \mathbf{x}_j \in \mathbb{R}^{n\times 1}$. Using the linear kernel induces a linear feature mapping and the resulting model is the same as the regularized least squares solution with a linear feature map (see Equation 5.27).

The probably most widely used kernel function in spectral estimation is the Gaussian kernel [103, 91, 104, 107, 108][P1], which is defined as

$$k(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{2\sigma^2}\right), \tag{5.35}$$

where $\mathbf{x}_i, \mathbf{x}_j \in \mathbb{R}^{n\times 1}$ and $\sigma \in \mathbb{R}^+$.

The selection of parameter σ is data dependent. Heikkinen *et al.* discussed that too small σ can lead to over-fitting of the training data and therefore a risk of poor generalization, whereas too large σ can lead to oversimplification [91]. In [P1], we have used the Gaussian kernel and empirically found that for our data, the width of the Gaussian kernel σ and the regularization parameter λ are dependent variables and that the parameter selection has a great influence on the estimation performance (see Section 5.3.4).

Conditionally positive definite kernels in ridge regression

The previously defined framework for kernel ridge regression is valid for PD kernel functions. By imposing certain constraints on the regression model, a larger class of kernel functions can be used, namely conditionally positive definite (CPD) functions, which are defined as [106, p.49]: Let X be a closed and bounded subset of $\mathbb{R}^{n\times 1}$. A function k is called conditionally positive definite if $\sum_{i,j=1}^{l} \alpha_i \alpha_j k(\mathbf{x}_i, \mathbf{x}_j) \geq 0$, for all $l \in \mathbb{N}$, $\mathbf{x}_i, \mathbf{x}_j \in X$ and $\alpha_i, \alpha_j \in \mathbb{R}$ with $\sum_{i=1}^{l} \alpha_i = 0$.

Following the works of Schöllkopf *et al.* [106] and Heikkinen *et al.* [104], a semi-parametric model of ridge regression with CPD kernels and d-conditionally positive kernels is formulated by introducing a polynomial expansion of feature vectors in Equation 5.30. In [P1], we proposed a conditionally positive kernel function for spectral reflectance estimation, which we refer to as the *logarithmic kernel*. It is defined as

$$k(\mathbf{x}_i, \mathbf{x}_j) = -\log(1 + \|\mathbf{x}_i - \mathbf{x}_j\|^{\beta}), \text{with } 0 < \beta \leq 2, \tag{5.36}$$

where $\mathbf{x}_i, \mathbf{x}_j \in \mathbb{R}^{n\times 1}$ and $\log(\cdot)$ is the natural logarithm. The CPD-ness of the logarithmic kernel was demonstrated in [P1]. The corresponding semi-parametric model for CPD kernel functions is formed by adding a constant term $\boldsymbol{a} \in \mathbb{R}^{m\times 1}$ to Equation 5.30. The minimization problem becomes

$$\begin{aligned} &\underset{\{A,\mathbf{a}\}}{\arg\min} \left(\sum_{i=1}^{l} \|\mathbf{r}_i - A\mathbf{k}(\mathbf{x}_i) - \boldsymbol{a}\|^2 + \lambda \operatorname{Tr}(AKA^T) \right) \\ &\text{s.t.} \;\; A\,\mathbf{1} = \mathbf{0} \end{aligned} \tag{5.37}$$

where $A \in \mathbb{R}^{m\times l}$, $\mathbf{k}(\mathbf{x}_i) \in \mathbb{R}^{l\times 1}$ and $K \in \mathbb{R}^{l\times l}$ are the same as in the PD case, and $\mathbf{1} \in \mathbb{R}^{l\times 1}$ and $\mathbf{0} \in \mathbb{R}^{m\times 1}$ are the all-one and all-zero vectors respectively.

The block matrix notation corresponding to the solution of the minimization problem in Equation 5.37 is

$$\begin{bmatrix} A & \mathbf{a} \end{bmatrix} \begin{bmatrix} K + \lambda I_l & \mathbf{1} \\ \mathbf{1}^T & 0 \end{bmatrix} = \begin{bmatrix} R & \mathbf{0} \end{bmatrix} \tag{5.38}$$

Using this model, a reflectance $\widetilde{\mathbf{r}}$ can be estimated from camera response $\mathbf{x}$ as

$$\widetilde{\mathbf{r}} = A\mathbf{k}(\mathbf{x}) + \mathbf{a}. \tag{5.39}$$

5.3.4 Summary of author publication [P1]

In [P1], we have compared the performance of the logarithmic kernel function with the linear and Gaussian kernel function. We were specifically interested in factors such as model parametrization, training set size and the number of camera spectral channels. The experiments were performed using simulated as well as measured camera response data of a LSMSC system similar to $\mathbf{SC1_{12C}}$.

In this subsection, we focus on a summary of aspects related to the first two factors, as the findings for the other aspects evaluated in [P1] are quite similar to those presented in Section 5.5 below.

Selection of model parameters:

For the case of logarithmic and Gaussian kernel functions, we have evaluated parameter search grid with respect to spectral $RMSE$. The logarithmic kernel contains parameter β (see Equation 5.36), and the Gaussian kernel contains the parameter σ (see Equation 5.35). The corresponding estimation models with both kernel functions further contain the regularization parameter λ.

For the Gaussian kernel, it was identified that the selection of λ influenced the optimal choice of σ with respect to average $RMSE$ and vice versa, while for the logarithmic kernel the selection of the optimal β was independent of λ. This is a particularly interesting observation, as it allows performing parameter selection for the logarithmic kernel sequentially for each parameter, instead of using grid search. For illustration purposes, we show in Figure 5.2 examples of $RMSE$ surface plots similar to those in [P1]. The $RMSE$ values are average values over 10 folds of parameter evaluation on training data.

We further evaluated parameter selection for the logarithmic kernel in a similar manner for a varying number of training samples (ie. 200, 400, 600 and 800). It was found that the shape of the $RMSE$ surface was in a wide area invariant to the number of training samples. Further, it was identified that $\beta = 2$ led to the lowest estimation error in all experiments, regardless of the number of training samples, number of camera channels, using simulated or measured camera response data. It was pointed out that the logarithmic kernel with $\beta = 2$ is a special form, which directly incorporates the negative squared distance kernel. Fixing the β parameter reduces the parameter selection in E-KER to an optimization problem in a one-dimensional search space, which can simplify parameter selection in practice.

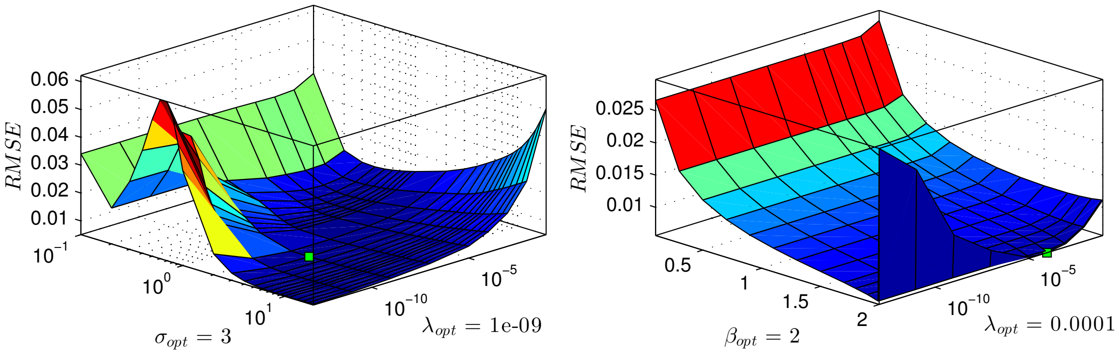
Influence of the number of training samples:

The last aspect of the previous item was concerned with the influence of the number of training samples on parameter selection for the case of the logarithmic kernel. What we summarize next is the influence of the number of randomly selected training samples on the overall estimation performance. For this experiment, 18 training sets with increasing number of samples from 50 to 1314 were considered.

For simulated camera response data, it was found that the estimation performance increased with the size of training set for all the three kernel functions evaluated. Particularly, the performance of logarithmic kernel showed the strongest dependence on the number of training samples, manifesting in low performance as compared with Gaussian and linear kernel when only few training samples were used. With large number of training samples, the performance of the logarithmic kernel approached that of the Gaussian kernel. The performance of linear kernel function was found to be least dependent on the number of training samples, but it was also the one with the lowest average performance.

For measured camera response data, the general trend of increasing performance with the number of training samples was also identified. However, the relation between average $RMSE$ and number of training samples was rather similar for all kernel functions. The average $RMSE$ of linear kernel was higher than that of logarithmic and Gaussian kernel for any number of training samples.

These results indicated a discrepancy between simulated and measured camera response data with respect to spectral estimation performance, a finding that is supported by various experiments in this dissertation.



(a) Gaussian kernel **(b)** Logarithmic kernel

Figure 5.2: Average $RMSE$ (over 10 folds of parameter optimization) over the parameter search space with measured camera responses of Pantone dataset. The optimal parameter selection is illustrated by a green square. Illustration is adopted from [P1].

5.4 Constrained estimation

We mentioned earlier the fact that from a statistical point of view, most information of natural reflectances is contained in a space that is of lower dimension than $\mathbb{R}^m$. There are other general properties that further constrain the spectral reflectances, for instance smoothness or physical realizability. The previously introduced spectral estimation approaches do not impose such constraints explicitly on the solution, and it was reported by several researchers that some approaches result in spiky or not smooth estimates or in estimates that exceed the range of values that are physically meaningful (eg. [103, 90]). This limitation motivated the development of various constrained estimation approaches and we present in the following some ideas that have been explored by others.

5.4.1 Physical realizability

Physical realizability refers to the fact that the numerical range of spectral reflectances is bounded in the interval $[0, 1]$ (Note that this is not the case for fluorescent samples, which are not considered in this work). This means that no less than no light and no more than all light can be reflected from a surface [109]. The numerical bound of reflectances can be considered as a hard constraint of physical realizability. A relaxed constraint of physical realizability is positivity, which might be easier to impose on a solution.

Li *et al.* [110] considered a physical model based spectral estimation of coordinates in an orthogonalized reflectance space. They defined a minimization problem that constrains the estimated reflectance to be in range $[0, 1]$, but did not further elaborate on the method used to solve it. Park *et al.* [111] used quadratic programming with a positivity constraint to solve the problem. Both authors also introduced a smoothness constraint, which we mention below.

Morovič and Finlayson [9] enforced the $[0, 1]$ range in their constrained metamer set solution to the spectral estimation problem by defining a set of inequality equations among other constraints (a more detailed description can be found in [109]). They solved the resulting minimization problem subject to linear inequalities using quadratic programming.

Heikkinen *et al.* [91, 104, 107] proposed various pairs of forward and backward transformation functions that enforce physical realizability. These functions were referred to as *link* functions, and they can be used in the RKHS based empirical estimation model (see Section 5.3.3). Examples of these function pairs are the $\operatorname{arctanh}(\cdot)$ and $\tanh(\cdot)$ [91], the $\operatorname{logit}(\cdot)$ and $\operatorname{logistic}(\cdot)$ [107] and the Gaussian copula and the corresponding inverse function [107]. They bound the estimated reflectances to the range $[0, 1]$. Further, the square root and square function pair [107] enforces positivity.

5.4.2 Smoothness constraint

Surface reflectances are typically smooth functions [8]. This condition can be used in spectral estimation to constrain a solution.

Li *et al.* [110] introduced a smooth operator in form of a $m \times m$ matrix to their definition of the minimization problem, where m is the number of discrete wavelengths of spectral reflectance data. This matrix is a finite approximation for computing the second derivative of the estimated reflectance vector and the desired solution minimizes the squared Euclidean norm of this second derivative.

Park *et al.* [111] enforced smoothness of the estimated reflectances by an additional additive weighted term in the minimization functional, which penalizes large values of the second derivative of the estimated spectral reflectance.

There are other ways to enforce smoothness on the solution. Morovič and Finlayson [9] defined a *naturalness* constraint, which states that an estimated reflectance is only natural, if it can be written as a convex combination of a set of representative spectral reflectances. If this set contains smooth reflectances, a convex combination is smooth as well and smoothness is therefore enforced indirectly.

5.4.3 Summary of author publication [P2]

In some applications, negativity of estimated reflectances is very problematic. One example is density measurement from estimated reflectance data in print quality control.

Optical print density measurement is traditionally used in multicolor printing presses as a measure of ink layer thickness to control the stability of the printing process. The measurement principle to obtain so-called color density is similar to that of a spectrophotometer, in the sense that specific color filters and a photosensitive sensor are incorporated. The color filters are typically narrow-band and specific to the ink to be measured. Accordingly, the peak transmittance of the filter corresponds to a spectral region of minimum reflectance (ie. the absorption maximum) of the ink on the printing substrate. In a certain range of reflectance magnitude, the sensor response can then be related to the ink-layer thickness.

Obviously, color density measurement can be obtained from spectral reflectance data directly, if the spectral transmittance function of the specific color filters are known. In fact, a less restrictive definition for print density measurements considers a filter whose peak wavelength is located at the minimum reflectance region of any kind of reflectance. This so-called spectral density $D_{mf}(\mathbf{r})$ is computed as

$$D_{mf}(\mathbf{r}) = -\log_{10}\left(\frac{\mathbf{r}^T\mathbf{a}_{\lambda_{min}}}{\sum_{i=1}^{m} a_i}\right), \tag{5.40}$$

where $\mathbf{a}_{\lambda_{min}} = [a_1, a_1, ...a_m]^T \in \mathbb{R}^{m\times 1}$ denotes a discretized narrow-band Gaussian shaped filter function, with peak-wavelength λ_{min} corresponding to the index of the minimum value in reflectance vector $\mathbf{r} \in \mathbb{R}^{m\times 1}$. It becomes clear that spectral density measurement fails when $\mathbf{r}^T\mathbf{a}_{\lambda_{min}}$ is negative or can be considered erroneous when λ_{min} corresponds to a negative reflectance value.

In [P2], we showed that E-KER-Log with adequate constraints can be used to overcome this limitation and hence, improve spectral density measurement performance. As mentioned before, constrained estimation with E-KER can be achieved by incorporating a pair of transformation functions in the estimation model. A forward transformation $\mathcal{T}$ is applied to the reflectances $\mathbf{r}$ in the model training phase, and a backward transformation $\mathcal{T}^{-1}$ is applied to the estimate in order to obtain the recovered reflectance $\widetilde{\mathbf{r}}$. We have evaluated the transformation function pairs reported in Table 5.1, which were proposed by Heikkinen *et al.* [107].

Table 5.1: *Square root* and *logit* transformation functions.

Link function	$\mathcal{T}(\mathbf{x})$	$\mathcal{T}^{-1}(\mathbf{y})$	Data range
Square root	$\mathbf{y} = \sqrt{\mathbf{x}}$	$\mathbf{x} = \mathbf{y}^2$	$\mathbf{y} \in [0, +\infty), \mathbf{x} \in [0, +\infty)$
Logit	$\mathbf{y} = \ln\left(\frac{\mathbf{x}}{1-\mathbf{x}}\right)$	$\mathbf{x} = \frac{\exp(\mathbf{y})}{1+\exp(\mathbf{y})}$	$\mathbf{y} \in (-\infty, +\infty), \mathbf{x} \in (0, 1)$

The backward transformation $\mathcal{T}^{-1}$ of the *square root* function pair is a quadratic function, which restricts the range of estimated reflectances to positive values. As the upper bound is not constrained, values of estimated reflectances can be larger than one.

For spectral density measurements, this is not critical. The backward transformation $\mathcal{T}^{-1}$ of the *logit* function pair imposes the numerical bound $[0, 1]$ to the solution.

We have evaluated two LSMSC system configurations, one with 12 camera channels similar to $\mathbf{SC1_{12C}}$, and the other with 3 camera channels. A set of 2698 color patches printed with a 7 ink wide gamut inkjet printer were considered as dataset. When using estimation with a least squares regression with linear feature mapping (E-PINV-SVD), 101 estimated reflectances contained negative values for the 12-channel system, and 193 for the 3-channel system. When using E-KER estimation, only samples from the 3-channel LSMSC resulted in estimated reflectances with negative values. However, it was concluded that for spectral density measurements, non-negativity should be enforced to avoid the risk of density measurement failure.

Experimental evaluation was divided into two parts: one related to spectral estimation, and the other related to density measurement. For spectral estimation, $RMSE$ and ΔE_{00} were considered as performance measures and the evaluation was based on 10-fold cross-validation. For spectral density measurement, $RMSE$ was computed between the spectral density $D_{mf}(\mathbf{r})$ computed from measured reflectance data, and the spectral density $D_{mf}(\widetilde{\mathbf{r}})$ computed from estimated reflectance data.

The most important experimental results can be summarized as follows:

Spectral estimation performance

The use of link functions guarantees physical realizability of the solutions while slightly increasing the estimation error. The performance of E-KER without link function was roughly twice as high as that of E-PINV-SVD. Using the square root link function reduced the average colorimetric performance, but the average spectral estimation performance was only influenced a little (3-channel system) or not at all (12-channel system). Using the *logit* link function reduced the average colorimetric and spectral estimation performance. The average performance of the 3-channel system when E-KER was used was approximately a factor of two lower than the 12-channel system. When using E-PINV-SVD, the performance was approximately 8 times lower than the 12-channel system.

Density measurement performance

For a reduced set of samples that do not result in negative estimates for any method, it was found that E-KER outperformed E-PINV-SVD on average roughly by a factor of 3 or more for both 3 and 12-channel system. When using link functions with E-KER, only slight differences in spectral density $RMSE$ were identified as compared with not using a link function.

Overall, it was found that E-KER with link functions results in a significantly higher spectral density measurement performance than E-PINV-SVD. Constraining estimates to positivity via link functions in E-KER decreased the colorimetric and spectral estimation performance slightly on average, as compared with using the same model without link function. This decrease in performance when using link functions was not found in case of spectral density measurement. Apart from that, the most important aspect to mention is that the constrained estimation is a necessary requirement to guarantee spectral density measurement to be meaningful.

5.5 Experiments

We have classified spectral estimation algorithms into two main groups, namely physical model based and empirical model based approaches. The methods presented form the basis of most spectral reflectance estimation algorithms that can be found in related literature, while various derivatives or hybrid methods exist. We compared P-MIN (Section 5.2.1), P-MW (Section 5.2.1), P-WN (Section 5.2.2), E-PINV-SVD (Section 5.3.1.A) and E-KER (Section 5.3.3)[3], and focus specifically on three aspects:

- Robustness with respect to simulated camera response noise
- Number of camera channels of the acquisition system
- Estimation with simulated as compared with measured camera response data

The last item of this list relates exclusively to the experimental evaluation of our laboratory LSMSC, ie. system configuration $\mathbf{SC2_{12C}}$. Corresponding experiments were performed by empirically quantifying spectral and colorimetric errors in reflectance estimation for a specific dataset, and we considered $RMSE$ and ΔE_{00} as error metrics. The configurations of experiments in this section is summarized in Table 5.2.

For the case of E-KER, reflectance data was centered by subtracting the mean of the training data prior to estimation. Consequently, the mean of the training data had to be added to the recovered spectral data after estimation. Further, for E-PINV-SVD and E-KER, camera response data was normalized to the range $[0, 1]$ by dividing each camera response channel-wise by 255, which is the maximum value corresponding to the 8-bit data range.

Table 5.2: Conditions and parameter settings of the experiments.

Systems:	Section 5.5.1: $\mathbf{SC2_{12C}}$ Section 5.5.2: $\mathbf{SC4_{3C}}$, $\mathbf{SC3_{6C}}$ and $\mathbf{SC2_{12C}}$ Section 5.5.3: $\mathbf{SC2_{12C}}$
Dataset:	**DS1**: 1761 Pantone solid chips
Parameter grid for E-KER: (before refinement)	$\lambda = \{10^{-14}, 10^{-12}, 10^{-10}, 10^{-8}, 10^{-6}, 10^{-4}, 0.01, 0.1, 1\}$ $\beta = \{0.001, 0.01, 0.1, 1, 1.5, 2\}$ $\sigma = \{0.01, 0.05, 0.1, 0.5, 1, 3, 5\}$
Camera response data:	Section 5.5.1: simulated (**Noise Scenario 1**) $\sigma_i = \sigma_d = \{0, 0.001, 0.005, 0.01, 0.05, 0.1, 0.2, 0.5, 1, 3, 5\}$ Section 5.5.2: simulated (**Noise Scenario 2**) $\mathbf{SC4_{3C}}$: average SNR 64 dB $\mathbf{SC3_{6C}}$: average SNR 62 dB $\mathbf{SC2_{12C}}$: average SNR 58 dB Section 5.5.3: measured

---

[3] We note that the first letter of each method's abbreviation indicates whether the method belongs to the class of physical models (**P**) or empirical models (**E**).

5.5.1 Robustness with respect to simulated camera response noise

The aim of this experiment is to compare the estimation performances at various levels of simulated camera noise. We considered system configuration $\mathbf{SC2_{12C}}$, ie. the LSMSC system for which optimal color filters were selected as part of this dissertation (Section 3.4.4). The camera response simulations were performed under **Noise Scenario 1** (which allows simulating camera responses in accordance with Equation 2.21). The average SNR for simulated camera response data ranged from approximately 30 dB to infinity (ie. noiseless case).

This experiment is divided into three parts. At first, we compared direct recovery methods, namely the minimum norm solution P-MIN and the regularized version P-MW. After that, we compared P-MW with Wiener estimation P-WN. The last part contains the comparison of empirical estimation methods and we focused on the least squares solution with linear feature mapping E-PINV-SVD and compared it with the kernel based solution E-KER with Gaussian kernel and the logarithmic kernel proposed by us in [P1].

A Physical model based estimation: Minimum norm vs. regularized solution (P-PINV-SVD vs. P-MW)

It is widely known that the minimum norm solution to the direct recovery problem does not perform well in the presence of noise, because the system to be inverted is heavily underdetermined. In this experiment, we compared the minimum norm solution P-MIN with the P-MW method, which estimates coefficient vectors of a truncated orthogonalized spectral reflectance space and subsequently reconstructs reflectances from the coefficient vectors (see Section. 5.2.1).

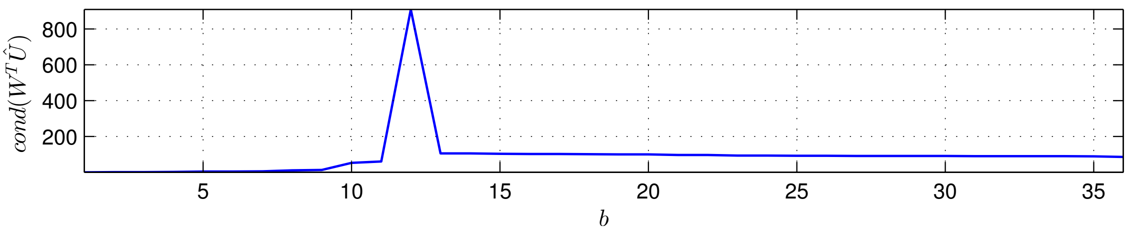
First we discuss the parametrization of P-MW, ie. the selection of dimensionality b of the truncated orthogonal reflectance space. From Equation 5.4 we know that the system equation of P-MW is $\hat{\mathbf{x}} = W^T\hat{U}\hat{\mathbf{c}}$, and correspondingly the system matrix is $W^T\hat{U}$. The condition number of this matrix measures the stability of the linear system with respect to the presence of noise, and gives a bound on how inaccurate the solution vector $\hat{\hat{\mathbf{c}}}$ will be. Experiments with $\mathbf{SC2_{12C}}$ and **DS1** were conducted to analyze the relation between condition number, estimation error and reconstruction error. Results are illustrated in Figure 5.3, where we plot the condition number against b (Figure 5.3a), the average estimation error ($RMSE$) as a function of b for various levels of additive camera response noise (Figure 5.3b), and the average reconstruction error ($RMSE$) as a function of b (Figure 5.3c). The reconstruction error quantifies the information loss when using the truncated orthogonalized spectral reflectance space. The plots motivate some discussions:

Condition number: (Figure 5.3a)

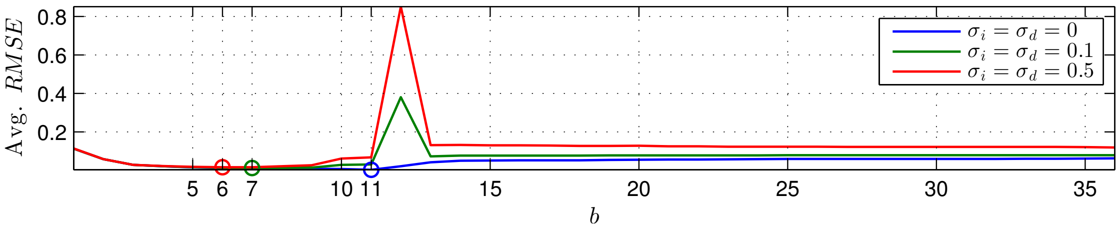
The condition number of $W^T\hat{U}$ reaches its maximum at $b = n$, with n being the number of camera channels, and decreases towards both sides. We note that the type of problem we solve in P-MW also depends on b. If $b < n$, the system is over determined and there exists no exact solution. The one computed by the pseudoinverse minimizes the sum of squared errors. If $b > n$, the system is underdetermined and the pseudoinverse gives the solution with minimum norm. For $b = n$, the system is fully determined, but it is severely ill-posed judging from the high condition number. Therefore even though a unique solution exists in this case, it is not a preferable one.

Estimation error: (Figure 5.3b)

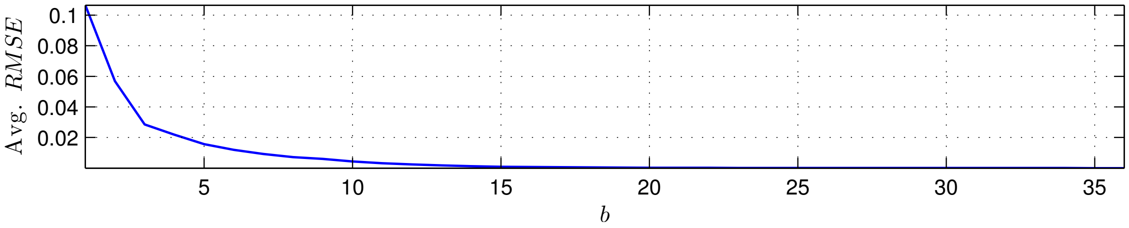
The average estimation error ($RMSE$) increases with σ (ie. the level of additive



(a) Condition number



(b) Estimation error



(c) Reconstruction error

Figure 5.3: Analysis of the influence of parameter b on condition number, estimation error and reconstruction error for P-MW. The b values corresponding to minimum average estimation error ($RMSE$) are marked by circles.

noise). The performance plunges drastically with increasing noise when $b = n$, which was expected from previous analysis of the condition number. In fact, the shape of the condition number curve coincides with the average $RMSE$ at different noise levels except when b is small, which we explain below.

Reconstruction error: (Figure 5.3c)

When $b < m$, reconstruction error occurs. We can see that this error decreases rapidly with increasing values of b. When comparing this curve with the average $RMSE$ plots in Figure 5.3b, we find that for low values of b, the reconstruction error dominates the overall estimation error. This explains why the average estimation error ($RMSE$) increases with decreasing condition number when b is small.

We can conclude that the selection of an optimal b for P-MW with respect to average spectral estimation performance depends on the level of noise of camera response data, the number of camera response channels n and the statistical structure of the data. In Figure 5.3b, we have marked optimal values of b for our data and $\mathbf{SC2_{12C}}$ with circles. For other system configurations, the optimal b for a particular level of noise will be different. But the analysis of condition number, estimation and reconstruction error as a function of b can be performed in a similar manner to select b adequately.

We now proceed to analyze the influence of camera response noise on estimation performance for P-MIN and P-MW. Parameter b was selected heuristically to minimize average estimation error ($RMSE$). We evaluated average $RMSE$ and ΔE_{00} as functions of signal dependent and independent noise variances. The results of the experiment are illustrated graphically in Figure 5.4. It can be seen that P-MW significantly outperforms P-MIN with respect to average colorimetric and spectral error for all levels of noise evaluated. Both models seem to be influenced in a similar way by signal dependent and signal independent noise. We further note that the performance of P-MIN on data with a high level of noise is so low that this method is unlikely to be usable in practical applications.

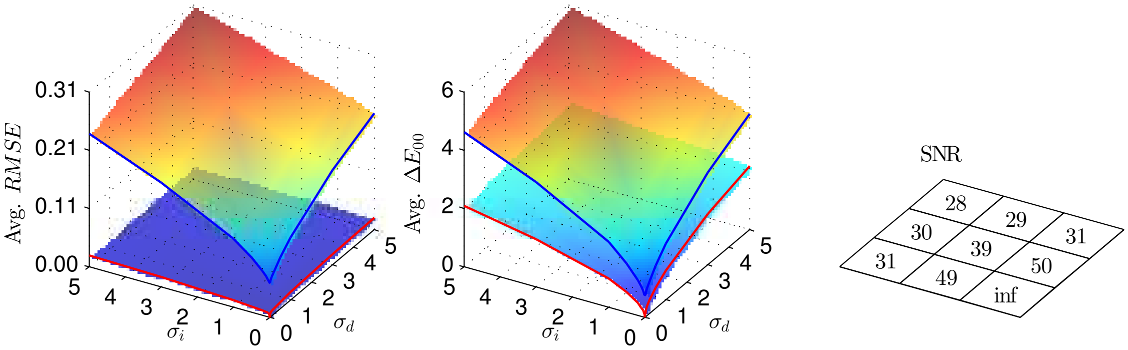


Figure 5.4: Error surfaces for direct recovery methods: P-MIN (blue frame) and P-MW (red frame). The plot on the right illustrates approximate average SNRs (in dB) for corresponding regions in the surface plots on the left side.

B Physical model based estimation: Wiener vs. regularized solution (P-WN vs. P-MW)

The Wiener model P-WN (see Section 5.2.2) takes into account statistical properties of reflectances and camera noise and is probably the most widely used model for direct recovery in reflectance estimation. To analyze model performance with respect to noise stability, we performed a comparison with P-MW, which is the better performing physical model based estimation method from the previous experiment. Noise variances evaluated correspond again to those reported in Table 5.2.

For P-WN, we modeled the statistical properties of reflectances by matrix $K_{\mathbf{r}} = \text{cov}(R) \in \mathbb{R}^{m \times m}$, where $R \in \mathbb{R}^{m \times l}$ is a matrix of l reflectance vectors from **DS1** stacked in columns, and $\text{cov}(\cdot)$ denotes the unbiased estimator of the covariance matrix (see Appendix A.4). Further, $K_{\mathbf{e}} = \text{cov}(\Xi) \in \mathbb{R}^{n \times n}$, where $\Xi \in \mathbb{R}^{n \times l}$ is a matrix of l noise vectors in columns, which correspond to simulated camera responses from **DS1**. Note that for this experiment, both matrices were assumed to be known.

The corresponding surface plots are shown in Figure 5.5.

We can see that the overall spectral and colorimetric performance of both methods is quite similar, regardless of the level of signal dependent or signal independent noise. When $\sigma_i > 0$ and $\sigma_d > 0$, P-MW outperforms P-WN slightly in terms of average $RMSE$, and vice versa in terms of average ΔE_{00}. When $\sigma_i = \sigma_d = 0$, the two methods have similar performance.

As mentioned earlier, the Wiener model P-WN is the optimal linear estimator if reflectances and noise are statistically independent normally distributed random variables. However, in this experiment this is not the case, as $\sigma_d > 0$ and camera response noise is therefore signal dependent. It was found that there is no clear advantage of using P-WN over P-MW under the conditions assumed in our experiment.

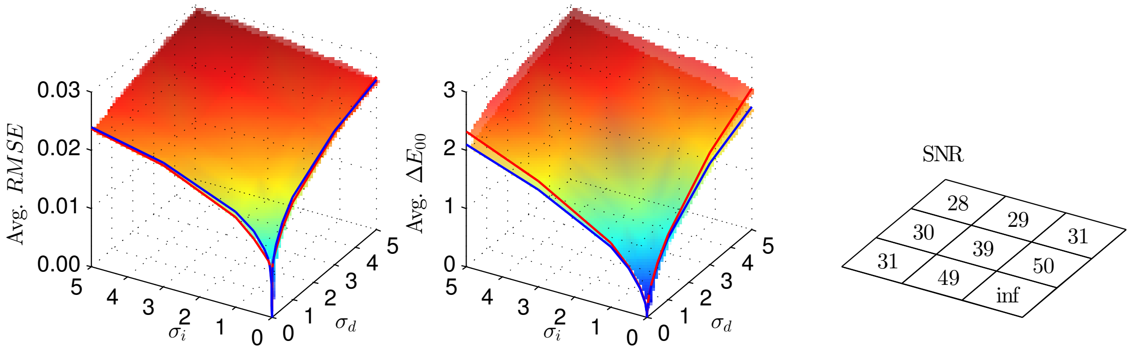


Figure 5.5: Error surfaces for direct recovery: P-MW (red frame) and P-WN (blue frame). The plot on the right illustrates approximate average SNRs (in dB) for corresponding regions in the surface plots on the left side.

C Empirical estimation: least squares vs. kernel based solution (E-PINV-SVD vs. E-KER)

We compared least squares regression E-PINV-SVD with linear feature mapping (see Section 5.3.1.A) and kernel based regularized least squares regression E-KER (see Section 5.3.3). For E-KER, we evaluated the logarithmic kernel function (E-KER-Log) and the Gaussian kernel function (E-KER-Gauss). Because the corresponding models are parametric, we used the refined grid search parameter selection scheme described in Section 2.4.3. Cross-validation technique was used to minimize the risk of biasing the evaluation by over-fitting.

The corresponding error surfaces for the three methods are shown in Figure 5.6. The results of this experiment indicate that for all the noise conditions considered, E-KER performs significantly better than E-PINV-SVD, and between the two kernel based methods, E-KER-Log outperforms E-KER-Gauss. For all models, performance drops with increasing the level of signal dependent and signal independent noise.

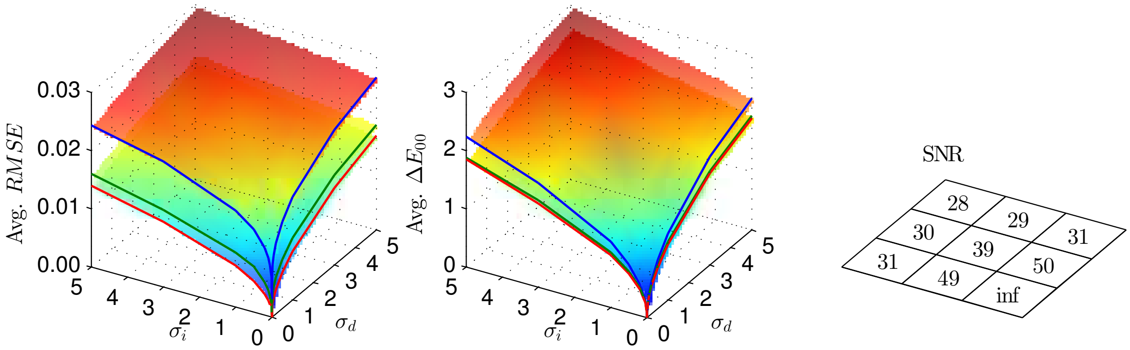


Figure 5.6: Error surfaces for empirical estimation methods: E-PINV-SVD (blue frame), E-KER-Log (red frame), E-KER-Gauss (green frame). The plot on the right illustrates approximate average SNRs (in dB) for corresponding regions in the surface plots on the left side.

D Physical model vs. empirical estimation: Minimum norm vs. least squares solution (P-MW vs. E-PINV-SVD)

Physical and empirical model based estimation were considered individually before. To compare approaches from both categories, we show the corresponding error surface graphs for P-MW (one of the best physical model based methods) and E-PINV-SVD (the worst empirical estimation method) in the same figure (see Figure 5.7).

It was found that the average spectral and colorimetric performance of P-MW is consistently lower than that of E-PINV-SVD at any level of noise. The relative difference increases with decreasing SNR.

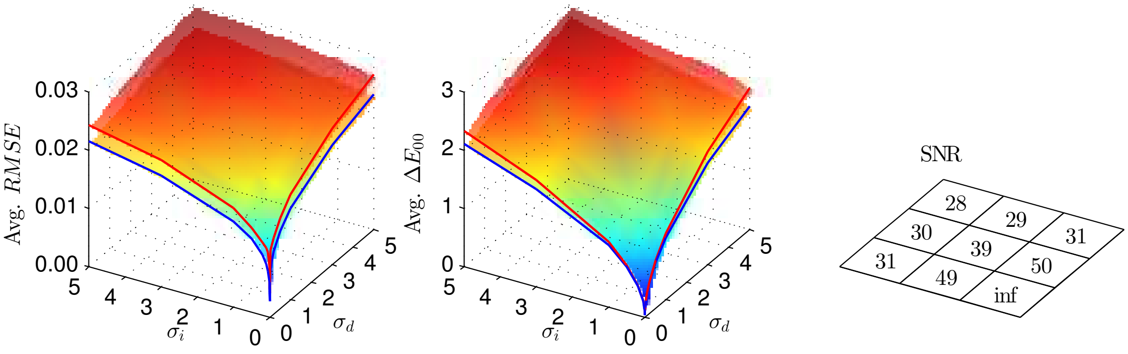


Figure 5.7: Error surfaces for P-MW (red frame) and E-PINV-SVD (blue frame). The plot on the right illustrates approximate average SNRs (in dB) for corresponding regions in the surface plots on the left side.

5.5.2 Number of spectral channels of the acquisition system

Beyond doubt, the number of camera spectral channels of the acquisition system influences the estimation performance. In this experiment, we considered system configurations $\mathbf{SC4_{3C}}$, $\mathbf{SC3_{6C}}$ and $\mathbf{SC2_{12C}}$. The corresponding effective responsivities are shown in Figure 5.8.

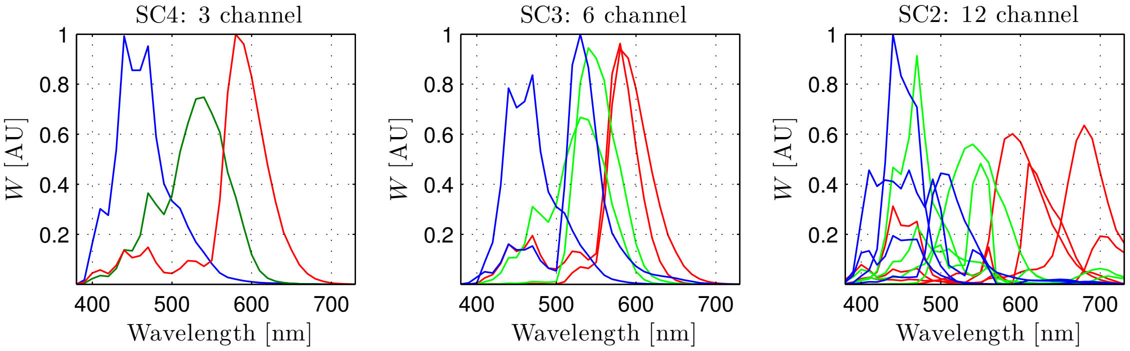


Figure 5.8: Spectral responsivity plots of 3 system configurations with different number of camera spectral channels (in 10 nm sampling).

A conclusion of the previous experiment was that the performance of all estimation algorithms is influenced by the level of camera response noise. A thorough analysis of the influence of the number of camera spectral channels on the estimation performance of different algorithms would require considerations of different camera response noise levels, as done in the previous experiment. However, such analysis is very time consuming, we therefore focused on a practical case, namely the evaluation under realistic camera response noise conditions, ie. **Noise Scenario 2**[4].

Numerical results of average $RMSE$ and ΔE_{00} errors are given in Table 5.3 and we summarize the most important findings as follows:

Number of channels:

For all methods compared, the three system configurations can be sorted in ascending order with respect to colorimetric and spectral estimation performance: $\mathbf{SC4_{3C}}$, $\mathbf{SC3_{6C}}$, $\mathbf{SC2_{12C}}$.

This finding seems obvious, but we remind the reader that it is not directly the number of channels that influences estimation performance, but the amount of relevant information captured by the system.

Physical model vs. empirical model based estimation:

The average performance of the best physical model based estimation algorithm is lower than that of the worst empirical model based estimation algorithm.

P-MIN:

Despite for $\mathbf{SC2_{12C}}$, the performance of this method is found to be so low that it is of no use for most practical applications in which spectral or color measurement is performed. For $\mathbf{SC2_{12C}}$, we illustrate in Figure 5.9 a plot of 5 estimated reflectances corresponding to the largest $RMSE$s. It can be seen that those reflectances do not satisfy physical realizability constraints (ie. smoothness, boundedness or positivity).

P-MW and P-WN:

There exists no significant difference between the performance of the two approaches. Estimated reflectance spectra are smooth, but in both cases, some reflectances do not satisfy positivity and boundedness for $\mathbf{SC4_{3C}}$ and $\mathbf{SC3_{6C}}$.

E-PINV-SVD and E-KER:

The performance of E-PINV-SVD is markedly lower than that of E-KER. Estimated reflectances of all methods are rather smooth, but for E-PINV-SVD with $\mathbf{SC4_{3C}}$ and $\mathbf{SC3_{6C}}$, boundedness and positivity are not satisfied.

E-KER-Log and E-KER-Gauss:

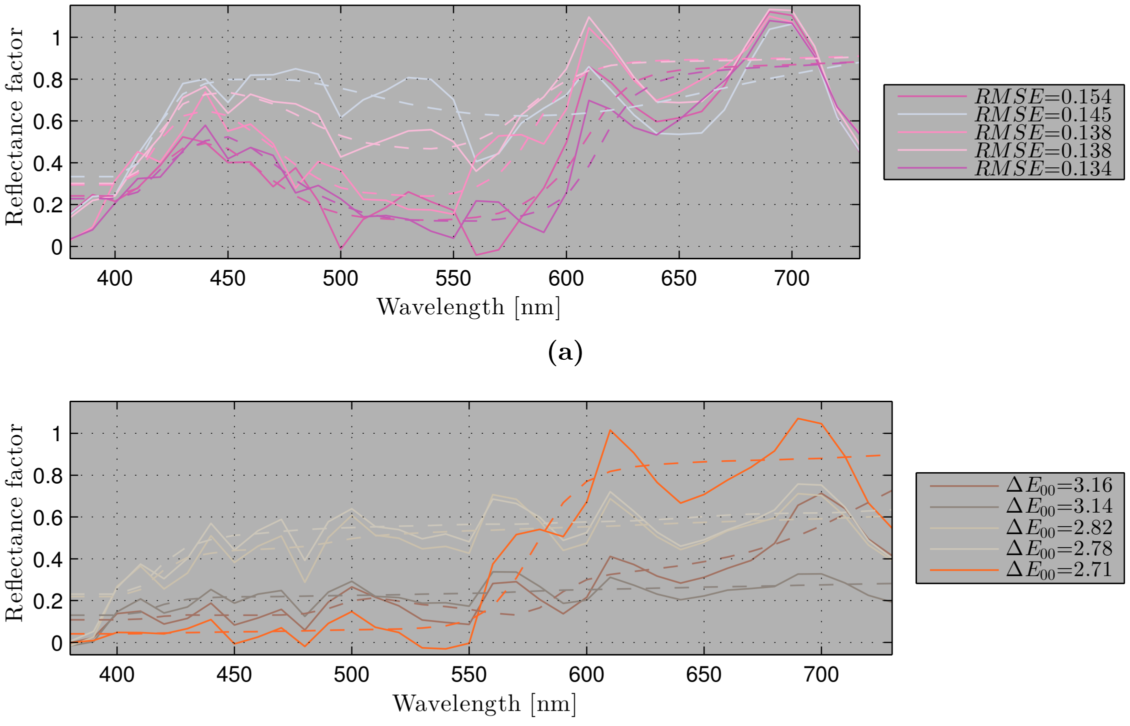
The performance of both types of kernel functions is comparable, with E-KER-Gauss having a slightly higher performance for $\mathbf{SC3_{6C}}$.

---

[4] As mentioned in Section 2.3.3.B, this means that we assume signal independent noise to be correct, and that signal dependent noise is modeled according to noise variances specified by the camera manufacturer.

Table 5.3: Numerical results of the experiment in Section 5.5.2: comparison of estimation performance for system configurations with different number of camera spectral channels.

	Avg. $RMSE$			Avg. ΔE_{00}		
	$\mathbf{SC4_{3C}}$	$\mathbf{SC3_{6C}}$	$\mathbf{SC2_{12C}}$	$\mathbf{SC4_{3C}}$	$\mathbf{SC3_{6C}}$	$\mathbf{SC2_{12C}}$
P-MIN	0.232	0.211	0.066	14.00	3.65	0.88
P-MW	0.040	0.033	0.010	2.21	0.96	0.14
P-WN	0.043	0.027	0.010	2.16	0.43	0.16
E-PINV-SVD	0.039	0.022	0.007	2.10	0.41	0.11
E-KER-Log	0.025	0.014	0.003	1.35	0.26	0.09
E-KER-Gauss	0.025	0.013	0.003	1.35	0.24	0.09



(a)

(b)

Figure 5.9: Five estimated reflectances (solid lines) and the measured counterparts (dashed lines) corresponding to the largest $RMSE$ (a) and ΔE_{00} (b) error for $\mathbf{SC2_{12C}}$ using P-MIN.

5.5.3 Estimation with simulated as compared to estimation with measured camera response data

As mentioned earlier, $\mathbf{SC2_{12C}}$ corresponds to our laboratory LSMSC system, so we have measured camera responses available. In this experiment, we compared the estimation performance for the following three cases:

Case 1: Noiseless simulated camera responses (**Noise Scenario 1**, $\sigma_i = \sigma_d = 0$)

Case 2: Noisy simulated camera responses (**Noise Scenario 2**)

Case 3: Measured camera responses

We report average and maximum $RMSE$ and ΔE_{00} error in Table 5.4. Further, first order statistics corresponding to **Case 3** are given in Table 5.5. The results are summarized as follows:

General performance trend with respect to the three cases:
As expected, the performance when considering noiseless camera response data (**Case 1**) is the highest. For noisy simulated camera response data (**Case 2**), the performance is lower for all methods. The lowest performance is found for measured camera response data (**Case 3**).

Physical model vs. empirical model based estimation:
The performance discrepancy between considering simulated and measured camera response data is much larger for physical model based estimation methods than for empirical models. We discuss various reasons that can explain this finding below.

Performance of physical model based estimation:
The performance with measured camera response data is very low for P-MIN and P-MW. Low performance was already identified in previous experiments with noise contaminated data for P-MIN, but not so much for P-MW. Among the physical model based methods, P-WN is the only one whose performance is reasonable to be considered in practice.

Performance of empirical model based estimation:
E-KER-Gauss has the highest spectral estimation performance with measured camera response data, followed by E-KER-Log and E-PINV-SVD. The color measurement performance is similar for E-PINV-SVD and E-KER-Log, and slightly higher for E-KER-Gauss. When comparing maximum errors from Table 5.5, it can be seen that E-KER-Log performs the worst among the empirical methods. Spectrally, the lowest maximum error is found for E-KER-Gauss, and colorimetrically for E-PINV-SVD. We note that the average colorimetric and spectral error values corresponding to measured camera response data for empirical methods are in ranges that are adequate for various color and spectral measurement applications.

Table 5.4: Numerical results of the experiment in Section 5.5.3: average and maximum colorimetric and spectral estimation errors for noiseless (**Case 1**), noisy simulated (**Case 2**), and measured camera response data (**Case 3**) corresponding to $\mathbf{SC2_{12C}}$.

	Avg. $RMSE$			Avg. ΔE_{00}		
	Case 1	**Case 2**	**Case 3**	**Case 1**	**Case 2**	**Case 3**
P-MIN	0.063	0.066	2.697	0.83	0.89	44.21
P-MW	0.004	0.010	0.110	0.05	0.14	14.90
P-WN	0.003	0.010	0.027	0.04	0.16	1.70
E-PINV-SVD	0.003	0.007	0.009	0.04	0.11	0.41
E-KER-Log	0.001	0.003	0.006	0.01	0.09	0.41
E-KER-Gauss	0.001	0.003	0.005	0.01	0.09	0.36

Table 5.5: Numerical results of the experiment in Section 5.5.3: first order statistics of colorimetric and spectral estimation errors for measured camera response data (**Case 3**) corresponding to $\mathbf{SC2_{12C}}$.

	$RMSE$			ΔE_{00}		
	95th Prctl.	**Std.**	**Max.**	**95th Prctl.**	**Std.**	**Max.**
P-MIN	5.297	1.416	6.395	76.15	19.89	103.83
P-MW	0.209	0.056	0.251	28.87	7.31	38.04
P-WN	0.051	0.014	0.123	4.26	1.38	10.48
E-PINV-SVD	0.019	0.005	0.034	0.89	0.25	1.76
E-KER-Log	0.013	0.004	0.050	0.93	0.29	2.96
E-KER-Gauss	0.011	0.003	0.028	0.87	0.25	1.96

There exists a large discrepancy between estimation with measured and simulated camera response data. Specifically for physical model based estimation, this discrepancy is larger than for empirical estimation. In an attempt to explain this disagreement, we need to recall some of the assumptions that limit the camera response simulations for our laboratory LSMSC system:

1. **Observation angle**
   Camera response simulations are based on Equation 2.19, which is a specific case of Equation 2.18, assuming that the observed scene radiance at any pixel location along the sensor line is constant. As explained in Section 2.3.2.C when discussing general limitations of the LSMSC system regarding color measurement, this assumption does not hold in practice.

2. **Illumination angle**
   The angle of the scene illumination with respect to the scanning surface normal is approximately 45 °. Spatial non-uniformity is accounted for by white-level correction, while spectral non-uniformity on the other hand is not addressed.

3. **Range of spectral data**
   We consider the spectral range from 380 nm to 730 nm in steps of 10 nm. The choice of this range was justified by the requirement for color and reflectance measurements. As mentioned before in Section 3.5 when considering filter selection in system design, the raw RGB sensor of the LSMSC is sensitive beyond 730 nm. Taking into account

the spectral power distribution of the LED line-light, effective responsivity beyond this range is low, but not zero. Hence, part of the measured camera response in respective channels has a larger signal as compared with simulations for the limited spectral range. A simple fix to this problem is to incorporate a cut-off filter in the LSMSC for the spectral range beyond 730 nm.

4. **Responsivity measurement errors**
   The physical model based estimation approaches considered in this thesis are based on accurate knowledge of the spectral responsivity of the LSMSC and spectral properties of the scene illumination. While it is relatively easy to obtain and verify scene illumination measurements, it is rather difficult to measure responsivities of a LSMSC line-scan camera system. The data used in this dissertation was obtained and verified by the camera manufacturer and corresponds to the raw RGB sensor responsivity. The effect of sub-image filtering was computationally modeled on the basis of spectral transmittance measurements of the corresponding color glass filters. Hence, the validity of the effective spectral responsivity was not confirmed other than by comparison of simulated and measured camera response data, which is not adequate to narrow down the source of discrepancy beyond the aspects discussed here.

The observed scene radiance deviation due to the first and second item of the list influences empirical as well as physical model based estimation. For empirical estimation, such effects can be accounted for by considering an independent estimation model for each pixel location along the sensor line. Preliminary experiments with this kind of pixel-wise estimation models have indeed shown that the overall estimation performance can be improved by that, proving that such effects actually do influence the overall performance of the system. However, having an independent estimation model for each pixel location can result in spatial discontinuities of the estimated spectral data. To improve this situation, spatial smoothness could be enforced in the estimation stage via additional constraints.

The third and fourth items do not influence empirical estimation, as corresponding models are not based on assumptions about the acquisition process like physical estimation models. We therefore believe those last two items to explain most of the discrepancy between the results corresponding to simulated and measured camera response data. On one hand, this motivates future work with respect to modeling the acquisition process more accurately and potentially accounting for some of the aforementioned effects in the estimation model. On the other hand, this particular aspect illustrates a general limitation of physical model based estimation, namely the requirement of an accurate model of the image acquisition process and availability of corresponding data. Empirical models do not have this requirement, and as we have shown experimentally, can be used to achieve high performance in color and spectral measurements. However, the requirement of observation data for model training can also be a limiting factor in practical applications and should not be disregarded.

5.6 Summary and conclusions

We have introduced the spectral reflectance estimation problem in the general context of multi-spectral imaging. Two categories of approaches were distinguished, namely physical model based spectral estimation and empirical model based spectral estimation. We have presented the mathematical basis of the models that are most commonly used and linked them to related literature. In [P1], we proposed the logarithmic kernel for the specific case of RKHS ridge regression in spectral reflectance estimation. For the data considered in that study, we found that this kernel function results in comparable estimation performance with the Gaussian kernel, but when few training samples are used, the performance is considerably lower. With respect to model parametrization using the logarithmic kernel, we identified circumstances that allow sequential model parameter selection instead of grid search. Further, fixing the corresponding kernel parameter ($\beta = 2$) was found to be a good choice in practice. Both aspects simplify parameter selection considerably.

Most reflectance estimation algorithms do not constrain the estimation result, even though spectral reflectances are known to be smooth and bounded. In practical applications, such constraints are often desired and sometimes required. Various constrained estimation schemes have been proposed and we have summarized related literature. In [P2], we have applied constrained estimation to the specific application of spectral print density measurement, for which positivity of estimated reflectances is a requirement.

The theoretical aspects of various spectral reflectance estimation schemes were substantiated by analyzing colorimetric and spectral measurement performance experimentally with data corresponding to 1761 Pantone solid chips (ie. dataset **DS1**). Results within the scope of our experimental configurations allow concluding the following:

(1) Overall performance of estimation algorithms:
We considered P-MIN, P-MW, P-WN, E-PINV-SVD and E-KER. For E-KER, the Gaussian and the logarithmic kernel function were evaluated.

For the same number of channels and the same level of noise (except noiseless), the following performance ranking holds: P-MIN < P-MW/P-WN < E-PINV-SVD < E-KER.

(2) Robustness to noise: P-MIN < P-MW/P-WN/E-PINV-SVD/E-KER
We found that all methods compared are influenced by simulated camera response noise, resulting in performance decrease, but not to the same degree. The camera response simulations were performed with various levels of signal independent and signal dependent noise. It was shown that regularized physical model based estimation (P-MW), and the Wiener model (P-WN) that incorporates a noise term, are substantially more robust with respect to noise than a non-regularized approach (P-MIN). However, the performance of these models depends strongly on accurate statistical knowledge about the reflectances to be estimated and the camera acquisition process. For empirical models, we identified a generally higher performance of E-KER than for a simple regression model with linear feature mapping (E-PINV-SVD) and physical model based methods.

(3) Estimation performance regarding the number of camera spectral channels: 3-channel < 6-channel < 12-channel
We evaluated estimation performance for system configurations with different number of spectral channels ($\mathbf{SC4_{3C}}$, $\mathbf{SC3_{6C}}$ and $\mathbf{SC2_{12C}}$), and considered simulated camera responses with a level of camera response noise that is realistic. As expected, the average estimation performance was found to increase with the number of channels.

(4) Estimation performance with simulated as compared to measured camera response data: measured < simulated noisy < simulated noiseless
This finding is exclusive for $\mathbf{SC2_{12C}}$, for which we obtained camera response data by measurement and simulation. This LSMSC system incorporates the optimal

filters selected in Section 3.4.4. It was found that the estimation performance based on measured camera response data is significantly lower than for simulated camera response data. We have identified several potential reasons for this and also discussed possible approaches to overcome these limitations.

In summary, it can be said that when empirical models are considered, the overall colorimetric and spectral estimation performance is satisfying and according to our appraisal suitable for various color or spectral reflectance measurement applications. From our experimental results, we can not claim the same for physical model based estimation. We believe that the main reasons explaining its low performance are an oversimplified view of the acquisition process (ie. ignoring observation and illumination angle deviations), measurement uncertainty for spectral components of the LSMSC (ie. scene illumination SPD and camera responsivity) and missing spectral information in the near infrared where the sensor is still responsive. As pointed out, we are aware that our system is sensitive beyond 730 nm, but did not have corresponding spectral data available to use in physical models. We explicitly note that the low performance of physical based models found experimentally shall not be understood as if these models are generally not suitable for LSMSC systems. In fact, overcoming these limitations motivates future research. Despite the difficulties in using a physical model based estimation correctly, an advantage of corresponding methods is that no observation data is required for training, like in the case of empirical estimation models.

6
Model training in empirical estimation

Contents

6.1 Preface

In empirical spectral estimation methods (see Section 5.3), the model is obtained from training data. Without doubt, model training influences the system performance, and in practice one is interested in obtaining the estimation model under the premise of optimizing performance. So far, we haven't considered the construction of a training set in particular. For instance in experiments where we compared different estimation algorithms in Section 5.5, the training data was randomly selected, as we were only interested in a fair comparison between different spectral estimation methods without particular consideration of the model training process. In this chapter, we will take a closer look at model training in empirical estimation. We concentrate on the following two aspects:

Training sample selection: refers to the process of selecting a subset of available samples for model training. The model is then used in the estimation process of new observations. This model training scheme can be considered as *global* training.

Adaptive estimation: refers to the scheme where an estimation model is trained specifically for every new observation. The corresponding training process can be considered as *local* training.

The remainder of this chapter is as follows: in the next section, we will consider the training sample selection problem. After discussing various motivations for performing a global training, we revise related literature and present a summary of [P3]. In Section 6.3, we introduce a generalized framework for adaptive estimation that conceptually unifies most methods from related literature in the spectral reflectance estimation domain. Section 6.4 contains an experimental evaluation of training sample selection and adaptive estimation. We summarize and conclude this chapter with Section 6.5.

Notation
We denote $\mathcal{S} = \{v_1, v_2, ...v_l\} = \{(\mathbf{x}_1, \mathbf{r}_1), (\mathbf{x}_2, \mathbf{r}_2), ...(\mathbf{x}_l, \mathbf{r}_l)\} \subset \mathbb{R}^{n\times 1} \times \mathbb{R}^{m\times 1}$ as the set of available measurement data, consisting of l camera responses $\{\mathbf{x}_i \in \mathbb{R}^{n\times 1} | i = 1, 2, ...l\}$ and corresponding reflectances $\{\mathbf{r}_i \in \mathbb{R}^{m\times 1} | i = 1, 2, ...l\}$. Further, let $\mathcal{S}_{tr} \subset \mathcal{S}$ be a training set with k elements, ie. with cardinality $|\mathcal{S}_{tr}| = k$. Samples not included in the training set are in $\overline{\mathcal{S}}_{tr} = \mathcal{S} \setminus \mathcal{S}_{tr}$.

6.2 Training sample selection schemes

A global training sample selection results in a subset $\mathcal{S}_{tr}$ with k elements from the set of available data $\mathcal{S}$. One reason for using a subset of the available data instead of the full set is that it reduces the computational burden. For some estimation methods, the computational load increases with $|\mathcal{S}_{tr}|$ only in the model training phase (eg. E-PINV-SVD in Section 5.3.1), while for others methods, it increases with $|\mathcal{S}_{tr}|$ in both model training and prediction (eg. E-KER in Section 5.3.3).

Another reason for considering training sample selection is that the model trained from a subset might better adapt to a particular application. For example, when it is not feasible to obtain training data (or sufficient amount of training data) directly related to an application, we can construct a training set from data of another application or from a general purpose data set.

Whether or not a selection scheme considers application data is one criterion according to which global training sample selection schemes can be categorized. We follow [P3] and distinguish the following:

General purpose vs. application dependent selection refers to a general premise of the selection: In *general purpose selection*, neither application data nor the estimation model are considered in the selection process, whereas *application dependent selection* takes these aspects into account.

Bottom-up vs. top-down refers to the way how the training set is constructed. In *bottom-up*, we start with an empty set $\mathcal{S}_{tr} = \emptyset$ and sequentially add samples from $\mathcal{S}$ until $|\mathcal{S}_{tr}| = k$. In *top-down*, we start with the full set $\mathcal{S}_{tr} = \mathcal{S}$ and sequentially remove samples until $|\mathcal{S}_{tr}| = k$.

6.2.1 Related literature

We have compared various global training sample selection schemes in [P3]. Most items in the following list are taken from this work.

- The simplest training set selection scheme is general purpose *random selection* (further: **RD**) with the objective of reducing the size of the training set. On average, RD is a rather bad choice. This was shown in several experiments related to machine learning in different application domains [112, 113, 114].

- Wu *et al.* [112] proposed a bottom-up approach based on the *Kennard-Stone algorithm*, which aims at sequentially selecting k samples that are uniformly spaced over the sample space. The i-th sample ($i = 1, 2, ...k$) is selected as

$$\arg\max_{v_0} \{D(v_0) | v_0 \in \overline{\mathcal{S}_{tr}}\}, \tag{6.1}$$

  where $D(v_0) = \min\{d(v_1, v_0) | v_1 \in \mathcal{S}_{tr}\}$ is the minimum distance between v_0 and any element in $\mathcal{S}_{tr}$ and $d(v_1, v_0)$ is defined as the Euclidean distance between element v_1 and v_0. Because the application considered in [112] is classification of spectral data, elements in $\mathcal{S}$ are spectral vectors.

- Hardeberg [102] (further: **HD**) proposed a bottom-up scheme of training sample selection for spectral characterization of a multi-spectral imaging system. His iterative method is based on a criterion of minimum condition number and can be regarded as a general purpose selection scheme. The objective of the selection is to create a set of k samples that are most distinct from each other. The first sample is selected as the one with maximum variance in spectral reflectance space among reflectances from $\mathcal{S}$. The i-th sample ($i = 2, 3, ...k$) is selected as

$$\arg\max_{v_0} \{\mathrm{cond}(\mathcal{S}_c(v_0)) | v_0 \in \overline{\mathcal{S}_{tr}}\}, \tag{6.2}$$

  where $\mathrm{cond}(\mathcal{S}_c(v_0))$ denotes the condition number of the $m \times i$ matrix containing i spectral reflectances of dimension m in columns, taken from the set $\mathcal{S}_c(v_0) = \mathcal{S}_{tr} \cup \{v_0\}$.

- Kang *et al.* [114] proposed a bottom-up selection scheme for active learning in text classification. They used k-means to cluster elements from $\mathcal{S}$ into k clusters. Then, from each cluster, the sample with the smallest Euclidean distance to the cluster centroid is added to $\mathcal{S}_{tr}$.

In [P3], we have proposed a modification of this approach for training a spectral estimation model. The modification involves substituting the Euclidean distance measure in k-means by the *spectral similarity value* (SSV) distance metric, which takes into account magnitude as well as scale differences between spectral data. Experimental performance evaluation of both approaches indicated an advantage when using the SSV distance metric.

- Mohammadi *et al.* [115] grouped reflectances from $\mathcal{S}$ into n clusters using agglomerative hierarchical clustering with average linkage and correlation distance to measure dissimilarity of reflectances. For each cluster, the reflectance sample having the minimum average vectorial angle to any other reflectances in the set is selected for training. The approach follows a bottom-up scheme and belongs to the group of general purpose selection schemes.

- Cheung and Westland [113] proposed several objective functions for a general purpose iterative bottom-up approach. It is based on the assumption that representative reflectance samples to be selected for training should be most distinct from each other. Shen *et al.* [116] tested the proposed objective functions and found that MAXSUMS perform better than alternative proposals of Cheung and Westland. The method works as follows: Like in **HD**, the first sample is selected as the one with maximum variance in spectral space among $\mathcal{S}$. The i-th sample ($i = 2, 3, ...k$) is selected as

$$\arg\max_{v_0} \{D(v_0) | v_0 \in \overline{\mathcal{S}_{tr}}\}, \tag{6.3}$$

  where $D(v_0) = \sum_{v_1 \in \mathcal{S}_{tr}} \sqrt{d(v_1, v_0)}$ and $d(v_1, v_0)$ is defined as the Euclidean distance between element v_1 and v_0. The square-root term is introduced in order to penalize small spectral differences.

- Shen *et al.* [116] proposed two iterative bottom-up sample selection schemes for spectral estimation with a multi-spectral imaging system. The first approach is referred to as *eigenvector-based* method. It is based on the premise that the subspace containing the training reflectances should be the one that is most similar to the subspace of all reflectances in $\mathcal{S}$[1].

  The other approach is referred to as a *virtual-imaging-based* method, whose aim is to select samples iteratively that minimize the error between measured reflectances in the training set and corresponding reflectances estimated from simulated camera responses of a virtual imaging system. We propose an extension of this approach: Instead of considering a virtual imaging system, we considered our LSMSC. The i-th sample ($i = 2, 3, ...k$) is selected such as to minimize the average spectral estimation error ($RMSE$) for samples from an application dataset with known spectral reflectances (further: **TE**). Accordingly, this approach works in a bottom-up fashion and is application-dependent.

- For the specific case when $|\mathcal{S}_{tr}|$ is very large, most iterative sample selection schemes can not be used in practice because they are computationally too demanding. In [P3], we proposed a method to overcome this limitation using a clustering based

---

[1] We note that our simplified representation of this algorithm in [P3] contains an error and the reader should therefore refer to the original description in [116] for more details.

application dependent selection scheme. The recursive algorithm processes $\mathcal{S}_{tr}$ in top-down fashion such that elements in $\mathcal{S}_{tr}$ that do not reduce estimation quality are removed from the training set. A summary of this article follows next.

6.2.2 Summary of author publication [P3]

In [P3], we considered in-line print inspection based on reflectance data from a 12-channel LSMSC system, similar to $\mathbf{SC1_{12C}}$. For this particular application, we proposed a flexible and time-saving scheme for constructing a training set without the need of measuring camera response or reflectance. A requirement for this scheme is that a spectral model of the printing process is known[2]. Then, any printable reflectance can be predicted from relative amounts of primary inks available in the printing process. Based on the Yule-Nielson spectral Neugebauer model [117], we have predicted more than 14600 reflectances that sample printable colors within the printer gamut. Corresponding camera responses for the LSMSC system were then simulated based on an acquisition model similar to that described in Section 2.3.3.B, assuming the acquisition process is noise free.

We further proposed a *recursive rejection* (RR) algorithm to construct an application dependent training set from the estimated reflectances. The RR algorithm works in top-down fashion (ie. starts with a full set of available data and successively rejects samples from the training set). Accordingly, the available set of data is recursively processed by division into clusters of varying size using the k-means algorithm. Clusters of samples are rejected from the set if they do not enhance the performance of spectral estimation based on the E-PINV-SVD model (See Section 5.3.1.A).

We evaluated this algorithm in comparison with various other global training sample selection schemes for spectral estimation in terms of colorimetric and spectral performance on several datasets. The most important findings and conclusions of this work can be summarized as follows:

- In most training sample selection schemes, the number of training samples as a parameter has to be defined a priori. For spectral estimation, this parameter depends on the dataset considered. The optimization is not trivial and might be time consuming. The RR algorithm determines an optimal number of training samples automatically based on the premise of enhancing estimation performance.
- Several training sample selection schemes were found not to be applicable to very large datasets. The RR method does not have this limitation.
- The performance of RR is comparable or even better than other training sample selection schemes.

6.3 Adaptive estimation schemes

In spectral estimation using global training, a fixed prediction function is used for estimating reflectances of new observations. On the contrary, in adaptive estimation, each new observation has its own prediction function. Under the premise that an estimation model performs better if it is trained by data in a local region of the new observation instead of the entire training set, the similarity between training data and new observations can be factored into the construction of the prediction functions. We can achieve this by

---

[2] We note that obtaining this model involves spectral reflectance measurements.

employing a weighting scheme in the training process, with larger weights assigned to the training samples that are more similar to the new observation.

A disadvantage of adaptive estimation as compared with conventional estimation with global training is that it is computationally expensive, since the estimation process for each new observation involves weight assignment, model training and prediction. However, this might not be a limitation in practice.

To employ the weighting scheme in the training process, we can reformulate the initial minimization problem in Equation 5.10 for empirical estimation. Instead of minimizing the sum of estimation loss $\mathcal{L}$ over all training samples in $\mathcal{S}_{tr}$, we minimize a weighted sum in the adaptive estimation problem:

$$\arg\min_f \left(\frac{1}{2} \sum_{i=1}^{l} \alpha_i \mathcal{L}(f(\mathbf{x}_i) - \mathbf{r}_i) \right), \tag{6.4}$$

where α_i is a weighting factor for element i ($i = 1, 2, ...l$) in the training set. There are various methods for selecting weights and we distinguish two categories:

Continuous weighting:

In continuous weighting, α_i is expressed as a function $\mathcal{F}$, which describes the similarity between samples from the training set $\{v_i \in \mathcal{S}_{tr} | i = 1, 2, ...l\}$ and the new observation v:

$$\alpha_i = \mathcal{F}(v_i, v). \tag{6.5}$$

Binary weighting:

In binary weighting, $\alpha_i \in \{0, 1\}$. The weight for training sample i ($i = 1, 2, ...l$) is assigned according to the dissimilarity between v_i and v, such that

$$\alpha_i = \begin{cases} 1 & \text{if} \quad \mathcal{D}(v_i, v) \leq t \\ 0 & \text{if} \quad \mathcal{D}(v_i, v) > t \end{cases}, \tag{6.6}$$

where t is a dissimilarity threshold parameter and $\mathcal{D}$ can be considered as a distance function.

An intuitive interpretation of this weighting scheme is that training samples similar to the new observation are considered in the training process (ie. $\alpha = 1$), and others not (ie. $\alpha = 0$). If the number of training samples k is fixed (ie. t is considered as a dynamic parameter), binary weighting selects the k training samples that are most similar to the new observation for training.

Continuous weighting is nonparametric, while binary weighting has one parameter to be optimized (the dissimilarity threshold t or the number of training samples k to be selected). But both methods require the specification of similarity or distance function, which is critical for the performance of adaptive estimation. Note that we do not specify whether the similarity or distance is computed in camera response space or in reflectance space, as both cases are possible.

6.3.1 Example of continuous weighting in least squares estimation

If we incorporate continuous weighting scheme in the least squares estimation with linear feature mapping (ie. model E-PINV-SVD, see Section 5.3.1.A), we can express the model training process for a new observation as the following minimization problem:

$$\arg\min_{\hat{A}} \left(\frac{1}{2} \sum_{i=1}^{l} \|\alpha_i \hat{A}^T \mathbf{x}_i - \alpha_i \mathbf{r}_i\|^2 \right), \tag{6.7}$$

where weighting factors $\{\alpha_i \in \mathbb{R} | i = 1, 2, ...l\}$ are included in the expression for estimation loss for convenience. We denote $\hat{X} = [\alpha_1 \mathbf{x}_1, \alpha_2 \mathbf{x}_2, ...\alpha_l \mathbf{x}_l] \in \mathbb{R}^{n \times l}$ as a matrix of weighted training camera response vectors and $\hat{R} = [\alpha_1 \mathbf{r}_1, \alpha_2 \mathbf{r}_2, ...\alpha_l \mathbf{r}_l] \in \mathbb{R}^{m \times l}$ as a matrix of corresponding weighted reflectances from $\mathcal{S}_{tr}$.

The weighting factors can be computed from camera response data for instance by using the probability density function of a multi-variate Gaussian as similarity measure [98], which is defined as

$$\mathcal{F}(\mathbf{x}_i, \mathbf{x}) = (2\pi)^{-\frac{n}{2}} \det(\Sigma)^{-\frac{1}{2}} \exp\left(-\frac{1}{2} (\mathbf{x}_i - \mathbf{x})^T \Sigma^{-1} (\mathbf{x}_i - \mathbf{x}) \right), \tag{6.8}$$

where n is the dimensionality of the camera response data, $\Sigma = \text{cov}(\mathbf{x}_i) \in \mathbb{R}^{n \times n}$ is a positive definite covariance matrix of the multivariate random variable $\mathbf{x}_i$ and $\det(\Sigma)$ is the determinant of Σ. Note that this function only exists if $\text{rank}(\Sigma) = n$ (ie. Σ is full rank). Covariance matrix Σ can be estimated from observations (ie. samples in $\mathcal{S}_{tr}$)[3].

The solution to the quadratic minimization problem in Equation 6.7 is found by setting the derivative of the cost function with respect to $\hat{A}$ equal to zero and we get

$$\hat{A}_{LIN} = \hat{R}\hat{X}^{+}, \tag{6.9}$$

where $\hat{A}_{LIN} \in \mathbb{R}^{m \times n}$ is the estimation matrix obtained from weighted training data. Reflectance $\widetilde{\mathbf{r}} \in \mathbb{R}^{m \times 1}$ can then be estimated from camera response $\mathbf{x} \in \mathbb{R}^{n \times 1}$ by evaluating $\widetilde{\mathbf{r}} = \hat{A}_{LIN}\ \mathbf{x}$.

6.3.2 Example of binary weighting in least squares estimation

Suppose we have the same setting as before, but this time we want to incorporate binary weighting. The minimization problem is identical to Equation 6.7 and the corresponding solution was given in Equation 6.9. The dissimilarity function in the weighting scheme is selected as Euclidean distance in camera response space, ie.

$$\mathcal{D}(\mathbf{x}_i, \mathbf{x}) = \|\mathbf{x}_i - \mathbf{x}\|. \tag{6.10}$$

If we use a dynamic threshold t instead of a fixed value, the samples corresponding to the k smallest distances are assigned $\alpha = 1$ and all other samples $\alpha = 0$. If k is small, $\hat{X}$ and $\hat{R}$ are sparse. The zero-columns of $\hat{X}$ and $\hat{R}$ do not influence the model and can therefore be removed, often resulting in a great reduction of the sizes of $\hat{X}$ and $\hat{R}$.

6.3.3 Related literature

- Shen and Xin [118] considered adaptive spectral reflectance estimation from camera responses of a 3-channel color scanner using linear least squares regression. They used a binary weighting scheme and selected the k most similar samples from $\mathcal{S}_{tr}$ for training. The dissimilarity function is defined as

$$\mathcal{D}(\mathbf{r}_i, \widetilde{\mathbf{r}}) = \|\mathbf{r}_i - \widetilde{\mathbf{r}}\|, \tag{6.11}$$

---

[3] See Appendix A.4.

where $\mathbf{r}_i$ is the reflectance of the i-th training sample ($i = 1, 2, ...l$) and $\widetilde{\mathbf{r}}$ is the estimated reflectance of the new observation from camera response using a model conventionally trained with all samples.

- In a later work, Shen and Xin [98] expanded their previous approach by using a continuous weighting for the k most similar samples from $\mathcal{S}_{tr}$. The weighting function employed is the multivariate probability density function defined in Equation 6.8, and they found that the performance of the approach is less influenced by the selection of k than in their previous method.
- Agahian *et al.* [119] considered adaptive spectral reflectance estimation from tristimulus values using linear least squares regression in a truncated orthogonalized reflectance space (see Section 5.3.1.A). They used a continuous weighting scheme, and the similarity function is defined as

  $$\mathcal{F}(\mathbf{x_i}, \mathbf{x}) = \frac{1}{\Delta E_i + c}, \tag{6.12}$$

  where ΔE_i refers to the ΔE_{76} color difference between the i-th training sample ($i = 1, 2, ...l$) and the new observation, and the scalar c is a positive constant introduced to avoid division by zero.
- Eslahi *et al.* [120] considered adaptive spectral reflectance estimation from tristimulus values using canonical correlation regression (with non-linear feature mapping using polynomial terms). By means of this feature expansion, additional terms were introduced such that the dimensionality of spectral data and the expanded tristimulus values in feature space were matched. They used a continuous weighting scheme and the similarity function defined in Equation 6.12. They compared this approach with the method used by Agahian *et al.* [119] and found that the weighted canonical correlation regression performed significantly better.
- Babaei *et al.* [121] considered adaptive spectral reflectance estimation from tristimulus values using linear least squares regression. They used a continuous weighting scheme with the similarity function defined in Equation 6.12.
- Zhang *et al.* [108] considered adaptive spectral reflectance estimation from camera responses of a 3-channel camera system. Like Shen and Xin [118], they used a binary weighting scheme and fixed the number of training samples to be selected. They compared non-adaptive and adaptive versions of linear least squares regression, regularized linear least squares regression and reproducing kernel Hilbert space based regression (with the Gaussian kernel function). From experiments with simulated camera response data, they found that, apart from the RKHS based regression, adaptive schemes outperformed non-adaptive schemes. They concluded that the small training set size might have led to over-fitting which resulted in the low performance of RKHS based regression. Another finding was a small performance improvement for regularized linear least squares regression over the non-regularized version.

A tabulated summary of weighting schemes used in these works is given in Table 6.1.

Table 6.1: Adaptive estimation: summary of related literature.

Authors	**Year**	**Weighting scheme**	**Weighting function**
Shen *et al.* [98]	2006	Continuous	$\mathcal{F}(\mathbf{x}_i,\mathbf{x}) = (2\pi)^{-\frac{n}{2}}\det(\Sigma)^{-\frac{1}{2}}$ ... ... $\exp\left(-\frac{1}{2}(\mathbf{x}_i-\mathbf{x})^T\Sigma^{-1}(\mathbf{x}_i-\mathbf{x})\right)$
Agahian *et al.* [119]	2008	Continuous	$\mathcal{F}(\mathbf{x_i},\mathbf{x}) = \frac{1}{\Delta E_i + c}$
Eslahi *et al.* [120]	2009	Continuous	$\mathcal{F}(\mathbf{x_i},\mathbf{x}) = \frac{1}{\Delta E_i + c}$
Babaei *et al.* [121]	2011	Continuous	$\mathcal{F}(\mathbf{x_i},\mathbf{x}) = \frac{1}{\Delta E_i + c}$

Authors	**Year**	**Weighting scheme**	**Dissimilarity function**
Shen *et al.* [118]	2004	Binary	$\mathcal{D}(\mathbf{r}_i,\widetilde{\mathbf{r}}) = \|\mathbf{r}_i-\widetilde{\mathbf{r}}\|$
Zhang *et al.* [108]	2012	Binary	$\mathcal{D}(\mathbf{r}_i,\widetilde{\mathbf{r}}) = \|\mathbf{r}_i-\widetilde{\mathbf{r}}\|$

6.4 Experimental case studies

We have evaluated various global training sample selection and adaptive estimation schemes experimentally with respect to spectral estimation performance. The estimation model used is E-PINV-SVD, ie. least squares regression with linear feature mapping (see Section 5.3.1.A). We considered this model because it is simple and nonparametric[4].

Performance of spectral estimation is measured in terms of average $RMSE$. Whenever applicable, we compare various system configurations (ie. $\mathbf{SC2_{12C}}$, $\mathbf{SC3_{6C}}$, $\mathbf{SC4_{3C}}$). The datasets considered are **DS1**, **DS2** and **DS3**. The experimental setting for each experiment is summarized in Table 6.2.

Table 6.2: Conditions of the experiments in Section 6.4.

System	$\mathbf{SC2_{12C}}$, $\mathbf{SC3_{6C}}$, $\mathbf{SC4_{3C}}$
Dataset	Section 6.4.1: **DS2**: HKS-N; **DS3**: HKS-K Section 6.4.2: **DS1**: Pantone; **DS2**: HKS-N
Performance evaluation	Spectral estimation error (Avg. $RMSE$)
Validation scheme	Separate training and testing data
Parameter selection scheme	-
Camera response data	Simulated (**Noise Scenario 2**) Measured (for $\mathbf{SC2_{12C}}$)

6.4.1 Global training: general purpose vs. application dependent selection

The aim of this experiment is to compare general purpose training selection and application dependent selection in global training schemes, and to analyze the influence of number

---

[4] The problem with parametric approaches is the potential risk of biasing estimation performance by a bad parameter selection.

of training samples on the estimation performance. The methods considered are **RD** (general purpose selection), **HD** (general purpose selection) and **TE** (application dependent selection), which were introduced in Section 6.2.1. Training samples were selected from **DS2**, and samples in **DS3** were used for testing[5].

The results of the experiment are illustrated in Figure 6.1. Our findings are mostly independent of the system configuration, as well as the type of camera response data considered (ie. simulated or measured) and we therefore do not discuss them separately, but at once:

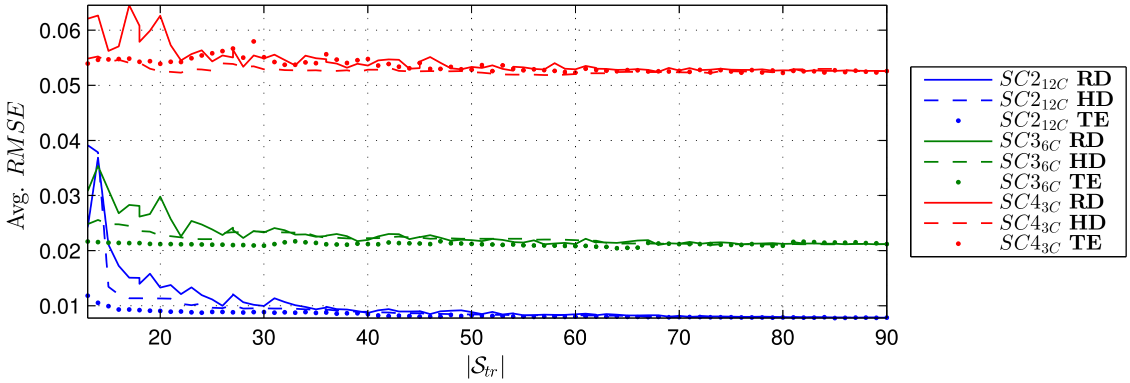
- The performance increases with the number of training samples for all methods compared. The relative increase is larger when $|\mathcal{S}_{tr}|$ is small. For $|\mathcal{S}_{tr}| > 30$, performance is almost constant for all methods.
- **RD**: random selection results in the lowest performance for all system configurations evaluated. The random nature of the selection process results in varying performance for various repetitions. The values shown in the figure correspond to the average over 10 repetitions. The standard deviation over repetitions decreases with increasing $|\mathcal{S}_{tr}|$ (not illustrated), because fewer randomly selected samples in the training set result in a larger variation of the estimation performance.
- **HD**: constructing a training set by iteratively selecting samples that are most distinct from each other is a much better choice than random selection. Not only is the selection stable (ie. not depending on a random number generator), but we also see that the performance is considerably higher.
- **TE**: in most cases, iteratively selecting samples that minimize the estimation error of training reflectances results in the lowest average $RMSE$ when $\mathcal{S}_{tr}$ is small. To some extend this is expected, as those samples are optimal with respect to minimizing average $RMSE$ of estimated samples from the training set. **DS2** and **DS3** might be similar enough so that the model created from **DS2** generalizes well on **DS3**. However, this assumption is by no means general.

We can conclude from this experiment that application dependent training set selection has certain advantages over a general purpose selection scheme. Whenever information about the application data is available, it makes sense to consider it in the training set selection.

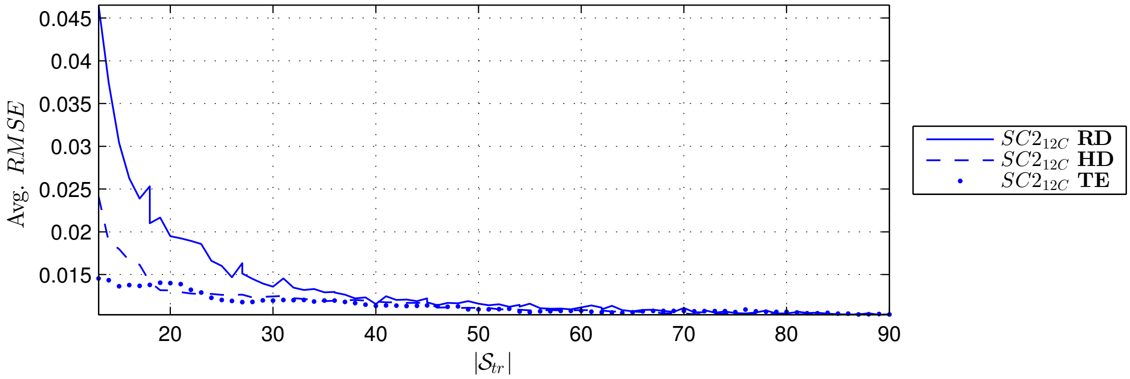
We have also found that it is possible to reduce the number of training samples drastically without or with only little performance reduction. The minimum amount of training data required to maintain high performance depends on the dataset and the estimation model and has to be evaluated for each application case specifically. As a smaller training set demands lower memory usage and often a lighter computational burden, it seems to be worth considering this step in a practical application.

---

[5] Although training set and test set contain intrinsically different data (samples in **DS3** are produced on coated paper substrate, whereas samples in **DS2** are produced on uncoated paper substrate), a fair comparison is not prevented by that.



(a) Simulated camera response data



(b) Measured camera response data

Figure 6.1: Influence of the number of training samples on estimation performance.

6.4.2 Adaptive estimation: performance evaluation

In this experiment, we evaluated different adaptive estimation methods, including continuous and binary weighting schemes, and compared them with conventional training.

A Continuous weighting

In continuous weighting, the estimation matrix is computed from weighted training samples as shown in Equation 6.9. In this experiment, we have considered the following functions:

- $\mathcal{F}_1(\mathbf{x}_i, \mathbf{x}) = \frac{1}{\|\mathbf{x}_i - \mathbf{x}\|}$ is the inverse of Euclidean distance between the i-th training sample ($i = 1, 2, \ldots l$) and the new observation in camera response space.
- $\mathcal{F}_2(\mathbf{r}_i, \widetilde{\mathbf{r}}) = \frac{1}{\|\mathbf{r}_i - \widetilde{\mathbf{r}}\|}$ is similar to the above, but the similarity is computed in reflectance space, with the reflectance of the new observation estimated using a model conventionally trained with all samples.
- $\mathcal{F}_3(\mathbf{x}_i, \mathbf{x}) = (2\pi)^{-\frac{n}{2}} \det(\Sigma)^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(\mathbf{x}_i - \mathbf{x})^T \Sigma^{-1} (\mathbf{x}_i - \mathbf{x})\right)$ is the multivariate Gaussian weighting function which takes into account statistics of camera responses.

Dataset **DS1** contains 1761 samples and **DS2** 90 samples. We used **DS1** for training and **DS2** for testing in the first part of the experiment (**Scenario 1**), then switched the training and test sets for the second part of the experiment (**Scenario 2**). Numerical results are reported in Table 6.3. Findings from results of **Scenario 1** can be summarized as follows:

Conventional vs. adaptive estimation:

Adaptive estimation outperforms training with the full set, regardless system configuration, weighting function and whether simulated or measured camera response data was considered. The decrease of average $RMSE$ ranges from 21 % to 40 % as compared with conventional estimation.

Number of channels:

There is no clear indication that the relative performance increase of adaptive over conventional estimation is depending on the number of channels.

Weighting functions:

Adaptive estimation using the multivariate Gaussian weighting function $\mathcal{F}_3$ performs better than $\mathcal{F}_2$ or $\mathcal{F}_1$ in all cases when simulated camera response data was considered. A slightly lower performance was found for the case of measured camera response data.

From results of **Scenario 2**, we discuss findings that are contrary to those of **Scenario 1**. For adaptive estimation with $\mathcal{F}_1$ or $\mathcal{F}_2$ and simulated camera response data, average performance is again higher as compared with conventional estimation. For measured camera response data, average performance is similar. When $\mathcal{F}_3$ is used, performance is lower than for conventional estimation in all cases apart from when $\mathbf{SC4_{3C}}$ is used.

One major difference between the two scenarios is the number of training samples and their data distribution in camera response space. If we look at the CIE-Lab color coordinates of the two datasets (illustrated in Figure 2.18), we see that **DS2** samples are much more sparsely and unevenly distributed over the color space, as compared with samples from **DS1**. It is reasonable to assume that the situation is similar in camera response space.

To find out whether or not the data distribution can explain the low performance with $\mathcal{F}_3$, we have analyzed weight distributions of adaptive training sets for various test samples. For instance for the test sample that has the lowest similarity to any other training sample (which can be considered *isolated*), we found that the assigned weighs for the closest five samples are much larger for $\mathcal{F}_3$ than for $\mathcal{F}_1$. This means that even though these few training samples are not similar to the test sample, they are assigned a large weight because they are the most similar ones. And for this particular case, adaptive estimation with $\mathcal{F}_1$ and conventional estimation using the full training set outperformed adaptive estimation with $\mathcal{F}_3$. We illustrate this case in Figure 6.2.

The explanation above might be one reason for the low performance of this particular test sample when $\mathcal{F}_3$ is used. Another important aspect to consider is that, as stated before, the efficiency of these schemes are based on the premise that an estimation model performs better if it is trained by data in a local region of the new observation instead of the entire training set. While this being a reasonable assumption, in practice it may happen that the training data far away from the test sample generalizes a better model for predicting this sample (due to the specific property of the data structure), and vice versa.

Table 6.3: Numerical results of average $RMSE$ for the continuous weighting experiment in Section 6.4.2. The average $RMSE$ for conventional estimation with the full training set is added for reference and denoted "Conv.".

	$\mathcal{S}_{tr}$ from **DS1**				$\mathcal{S}_{tr}$ from **DS2**			
	Conv.	$\mathcal{F}_1$	$\mathcal{F}_2$	$\mathcal{F}_3$	Conv.	$\mathcal{F}_1$	$\mathcal{F}_2$	$\mathcal{F}_3$
$\mathbf{SC4_{3C}}$ simu.	0.061	0.042	0.050	**0.039**	0.057	0.047	0.048	**0.044**
$\mathbf{SC3_{6C}}$ simu.	0.027	0.019	0.020	**0.016**	0.029	**0.026**	0.027	0.030
$\mathbf{SC2_{12C}}$ simu.	0.010	0.007	0.006	**0.005**	0.010	**0.009**	**0.009**	0.013
$\mathbf{SC2_{12C}}$ meas.	0.014	**0.011**	**0.011**	0.012	**0.013**	**0.013**	**0.013**	0.038

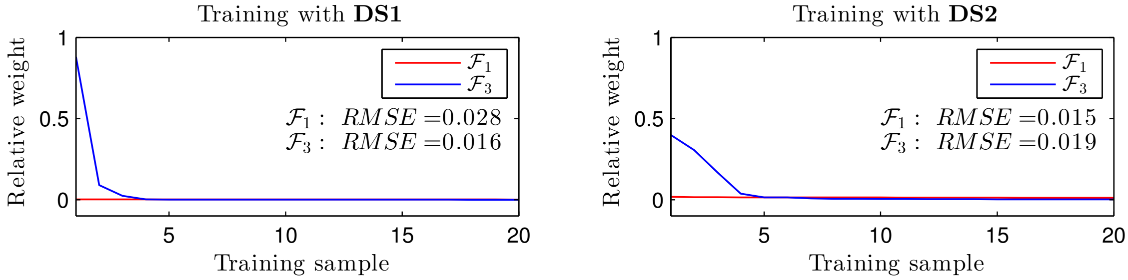


Figure 6.2: Training sample weights corresponding to the test sample with largest distance in camera response space to any sample of the training set. The plots show relative weights $\mathcal{F}_1$ and $\mathcal{F}_3$ (normalized by the sum of all weights) of each training sample in decreasing order.

B Binary weighting

In binary weighting, either the dissimilarity threshold parameter or the number of training samples to be selected is fixed. In this experiment, we evaluated methods of the latter case, ie. for each new observation we selected a fixed number of training samples for model training. One aspect of our analysis is the influence of number of training samples k on the model performance, and we considered values between 1 and $|\mathcal{S}_{tr}|$. The dissimilarity functions considered are the following:

- $\mathcal{D}_1(\mathbf{x}_i, \mathbf{x}) = \|\mathbf{x}_i - \mathbf{x}\|$ quantifies the dissimilarity between the i-th training sample ($i = 1, 2, ...l$) and the new observation in camera response space.
- $\mathcal{D}_2(\mathbf{r}_i, \widetilde{\mathbf{r}}) = \|\mathbf{r}_i - \widetilde{\mathbf{r}}\|$ is similar to the above, but the dissimilarity is computed in reflectance space, with the reflectance of the new observation estimated using a model conventionally trained with all samples.

The results are shown in Figure 6.3 and Figure 6.4, and our findings can be summarized as follows:

Interpretation of the curve shapes: (Figure 6.3a, 6.4a, 6.3b, 6.4b)
The shape of the estimation error curves of different system configurations using different dissimilarity functions follows a similar pattern. For $1 \leq k < n$, the error first decreases rapidly with increasing k, before increasing again drastically to reach a point with large error when $k = n$. For $n < k \leq l$ the error first decreases again rapidly, then stabilizes or starts some turbulences.

As mentioned in Section 5.3.1, the system is underdetermined when the number of training samples is smaller than the number of camera response channels, ie. $k < n$, and E-PINV-SVD gives the minimum norm solution. As the main limiting factor at this point is the small number of training samples, enlarging the training set can effectively improve the performance. When $k = n$, the system is fully determined and a unique solution exists. However the very large condition number of the system matrix (not illustrated) suggests that this is a bad solution, which explains the drastic error increase at this point. When $k > n$, the system is overdetermined, and E-PINV-SVD gives the least squares solution. The error first decreases with increasing training data, but after a certain point where training set size is no longer a limiting factor for the performance, the estimation error stabilizes or starts to be influenced by other factors, which are not straightforward to explain.

Number of camera channels: (Figure 6.3a, 6.3b)

Similar to the other experiments, estimation performance generally increases with the number of camera channels.

Dissimilarity measure $\mathcal{D}_1$ vs. $\mathcal{D}_2$: (Figure 6.3a, 6.4a, 6.3b, 6.4b)

From corresponding figures, we can see that differences in performance exist between using one or the other function. However, in most cases the differences are small and it is not straightforward to explain them. For $\mathbf{SC4_{3C}}$ (Figure 6.3a and Figure 6.3b), we see that the performance of $\mathcal{D}_2$ is lower than that of $\mathcal{D}_1$, but this is not so clear for other system configurations. The estimation performance of $\mathbf{SC4_{3C}}$ is generally rather low, hence measuring dissimilarity using estimated reflectances does not seem to be favorable in that case. Because $\mathcal{D}_1$ does not perform particularly bad under any condition evaluated, we conclude that this measure is the better choice.

Global training vs. adaptive estimation: (Figure 6.3a, 6.4a)

When **DS1** is used for training and **DS2** for testing, the optimal number of training (k_{opt}) is smaller than the total amount of available samples (ie. $k_{opt} < l$). This means that adaptive estimation outperforms conventional estimation. The situation is different when **DS2** is used for training, and **DS1** for testing. We discuss this aspect separately below.

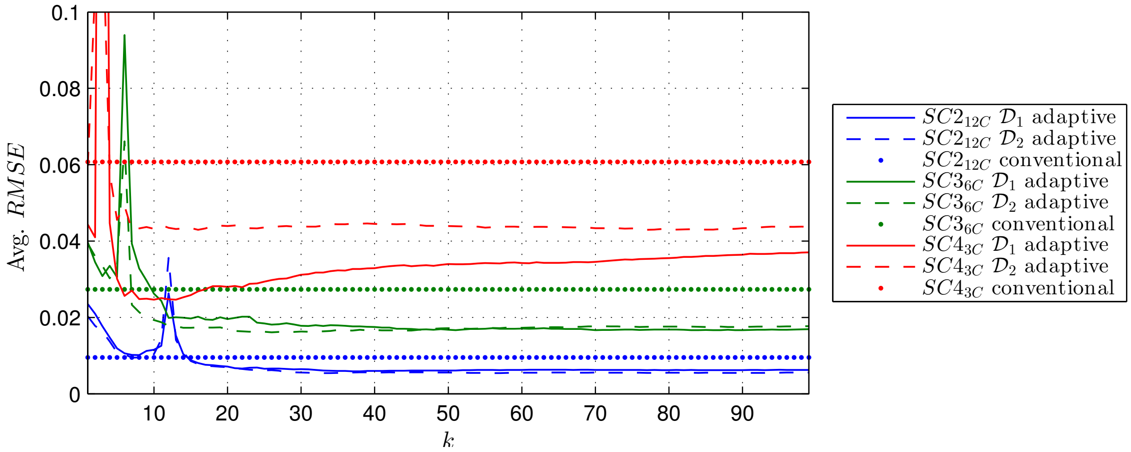
Training with DS2, test with DS1: (Figure 6.3b, 6.4b)

For simulated and measured camera response data, the improvement of adaptive over conventional estimation diminishes or even disappears as compared with using **DS1** for training.

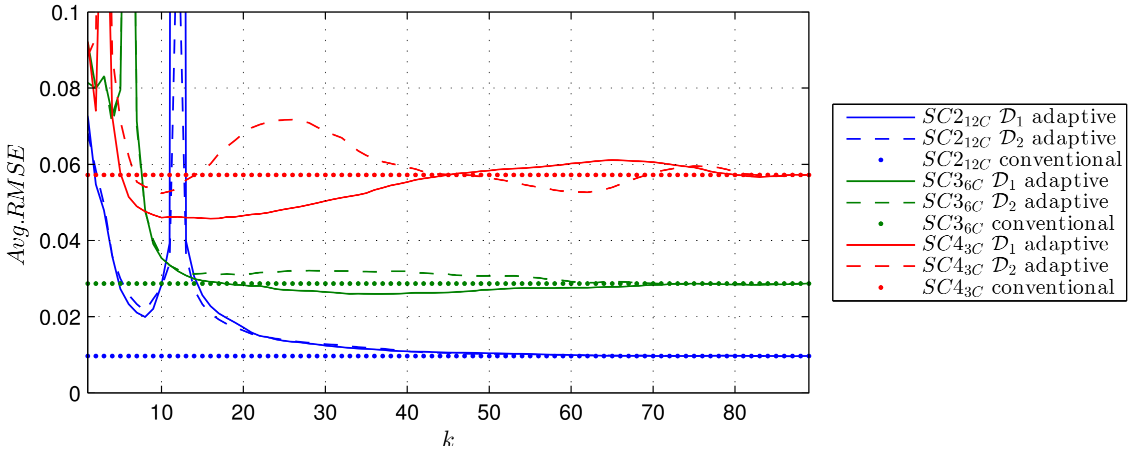
As pointed out earlier, **DS2** contains much less data than **DS1**. Hence, using **DS2** for training means that for a fixed value k, the selected samples may be much less similar to a new observation as compared with those selected from **DS1**. It also means that for a fixed dissimilarity threshold t, there may be much less samples selected from **DS2** as compared with **DS1**. We illustrate an example by considering a test sample with CIE-Lab color coordinate $\mathbf{x}_{Lab} = [60, 50, 60]^T$ using $\mathcal{D}_1$ as dissimilarity measure and a fixed threshold $t = 34$. From the results shown in Figure 6.5, we can see that there are 20 samples selected from **DS2**, and they are more sparsely and unevenly distributed as compared with the 95 samples selected from **DS1**.

Simulated vs. measured camera response data: (Figure 6.4a, 6.4b)

As before, performance is lower when measured camera response data is considered. Apart form this, there is no significant difference from other findings reported earlier.

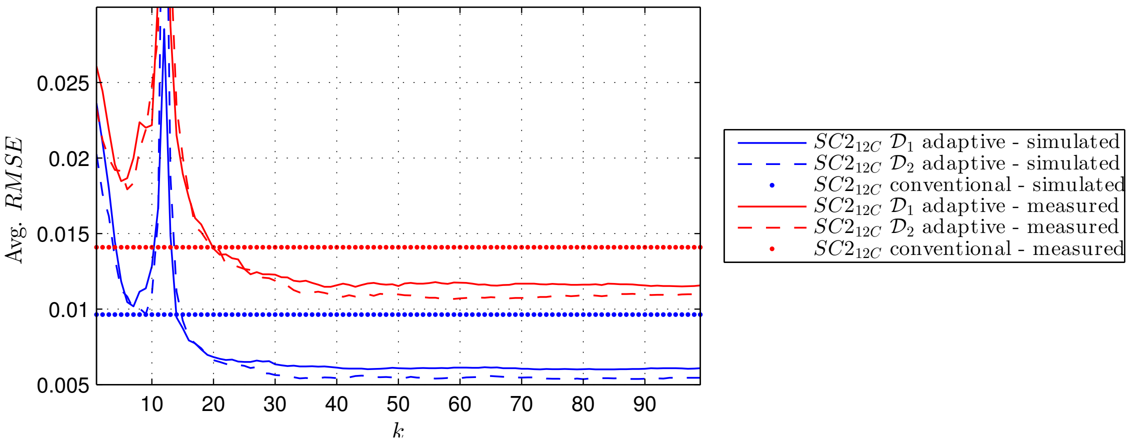


(a) DS1 for training

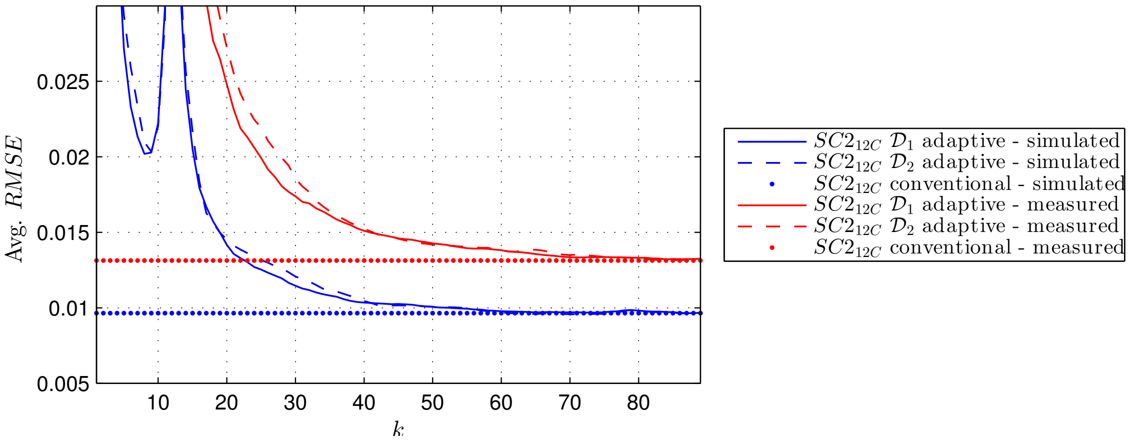


(b) DS2 for training

Figure 6.3: Estimation performance of conventional estimation and adaptive estimation with binary weighting plotted as a function of number of selected training samples, using three system configurations and two dissimilarity measures. Note that for **DS1**, we only show $RMSE$ for $1 < k \leq 100$ for better illustration.

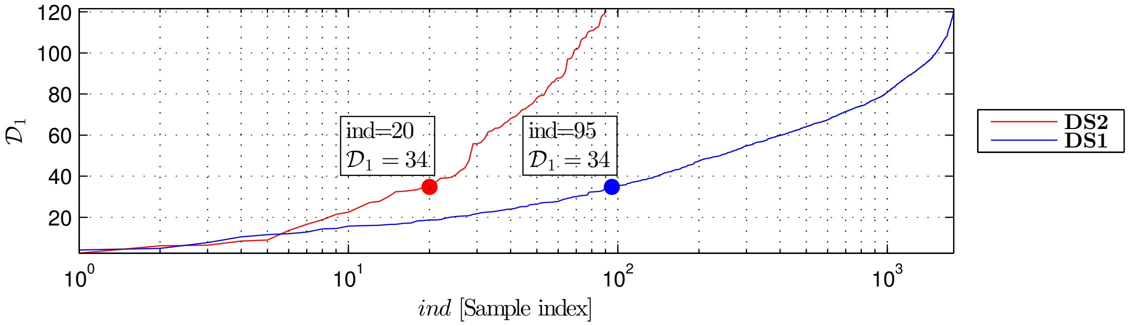


(a) DS1 for training

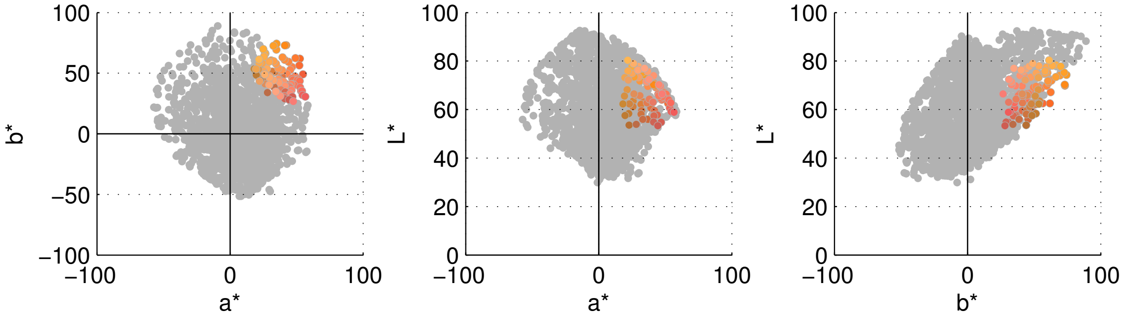


(b) DS2 for training

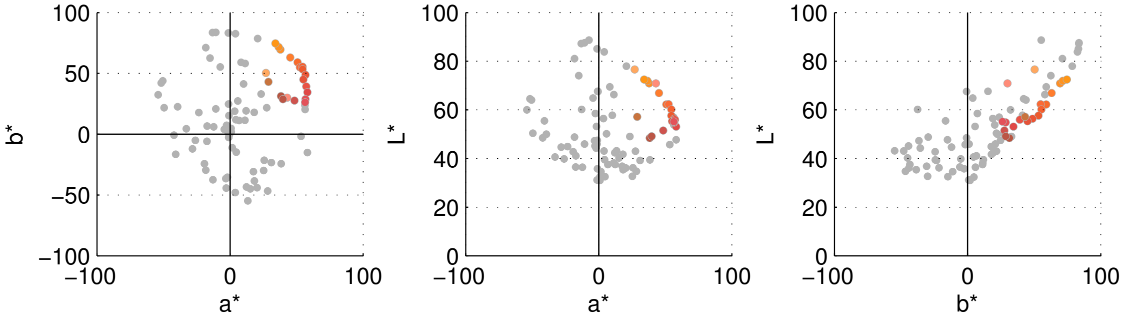
Figure 6.4: Similar with the illustration in Figure 6.3, but considering only $\mathbf{SC2_{12C}}$ with both simulated and measured camera response data.



(a)



(b) CIE-Lab color coordinates of elements in **DS1**.



(c) CIE-Lab color coordinates of elements in **DS2**.

Figure 6.5: Dissimilarity between a test sample and each element in training set **DS1** and **DS2** (a). CIE-Lab color coordinates for the samples in **DS1** (b) and **DS1** (c). The selected samples with dissimilarity $\mathcal{D}_1 \leq 34$ are marked in colors.

6.5 Summary and conclusions

We have considered two particular aspects in model training for empirical estimation: training sample selection and adaptive estimation.

Training sample selection is the process of choosing a subset from available samples for estimation model training. Once the model is constructed, it is used for estimation of new observations.

We have discussed motivations behind training sample selection and proposed a categorization scheme for related literature. In [P3], we considered training sample selection in the context of in-line print inspection by spectral reflectance measurements from multi-spectral image data. In that work, we proposed a recursive rejection algorithm (the so-called RR method) to obtain an application specific training set. A useful property of the RR method is that the algorithm determines an optimal number of training samples automatically. We showed in various experiments that estimation performance with a training set constructed by the RR method is comparable with methods that require an additional optimization process to determine training set size.

We also performed experiments to specifically compare general purpose with application dependent schemes. It can be concluded that application dependent training set selection should be used whenever application data is available, because general purpose selection schemes miss mechanisms that ensure generalization of the training set to application specific data. Another experimental finding for our data was that the number of training samples could be reduced drastically with only little or no estimation performance decrease.

In adaptive estimation, each observation has its own prediction function. This function takes into account the similarity between training samples and the new observation by incorporating a weighting scheme. We have proposed a general framework for adaptive estimation schemes and conceptually unified related works from the spectral estimation domain. Further, we illustrated scenarios of adaptive estimation using continuous and binary weighting and least squares regression estimation with linear feature mapping.

We performed various experiments that allowed drawing the following conclusions: Adaptive estimation schemes were found to outperform conventional estimation with a full training set, if the elements in the training set are distributed densely in the entire application space. This is not the case when the training set is relatively sparse with respect to the test set.

"Now this is not the end.
It is not even the beginning of the end.
But it is, perhaps, the end of the beginning."
Sir Winston Churchill (1874 – 1965)

7

Summary of the dissertation

Contents

7.1 Summary and conclusions

This dissertation is related to the design of line-scan multi-spectral LSMSC imaging systems, though many aspects apply also to other types of multi-spectral imaging systems. The design of such systems and the development of a specific imaging application require the consideration of many aspects. We focus on optimizing spectral properties of the optical components in such systems (Module 1), registration of LSMSC images (Module 2) and mathematical models for the recovery or estimation of surface spectral reflectance data from multi-channel image data (Module 3 and 4).

As this work covers studies from various domains, we structured the dissertation in chapters related to the modules, each containing problem statement, methods, experiments and detailed conclusions. In what follows, we present the modules in a general context and hence only give a summary of the most relevant findings and conclusions.

Module 1 is linked to acquisition system design. We focused on spectral properties of optical system components and considered system design as an optimization problem. Corresponding optimality criteria based on subspace projection and spectral estimation error were studied and summarized. We demonstrated that spectral estimation error based optimality criteria are often a better choice in practice.

For the specific task of LSMSC sub-image filter selection, we proposed a framework that incorporates heuristics in order to make the evaluation of large amounts of filter combinations feasible. Further, by using the Pareto optimality principle, we simplify the selection of an optimal filter combination. We have demonstrated our framework for 3-, 6- and 12-channel LSMSC systems based on a specific imaging principle, using a RGB line sensor with multiple lenses and color filters. This imaging principle was recently invented by company Chromasens GmbH. Our optimal filters for the 12-channel system configuration were considered for practical implementation and assembled in our laboratory setup. Experiments with measured data illustrated the high colorimetric and spectral estimation performance of the system (average $RMSE < 0.01$ and average $\Delta E_{00} < 0.6$), but also a shortcoming of the acquisition model considered for simulations.

Module 2 deals with the registration of multi-channel line-scan image data. We have demonstrated that image channel misalignment can degrade spectral image quality severely. A general framework for LSMSC image registration of planar and non-planar scanning objects was proposed.

For planar scanning objects, the registration process can be considered as geometrical calibration. We have designed a specific calibration target for line-scan cameras and a semi-automatic scheme to extract pixel displacement of key-points from image data. Based on that, we identified polynomial shaped displacement curves for our laboratory LSMSC system. Accordingly, a polynomial curve model and a uniform cubic B-spline based model with grid refinement were developed. Both models were evaluated empirically and it was confirmed that sub-pixel accurate image registration can be achieved (average displacement error below 0.1 pixel).

For non-planar scanning objects, we demonstrated that geometrical calibration alone is not adequate because of object viewpoint variations between geometrically calibrated RGB sub-images. We proposed a solution to this problem that uses a scene-dependent pixel-wise correspondence estimation scheme based on block-matching with sub-pixel estimation. The block-matching is performed on feature images that result from a novel color-invariant feature mapping that was developed specifically for LSMSC systems. Finally, RGB sub-images can be registered using the estimated pixel-wise image correspondence.

We compared physical and empirical model based feature mappings for our laboratory LSMSC system and found that the empirical model performance is considerably higher than the performance of the physical model. Evaluation of correspondence estimation showed that sub-pixel accurate image registration can be achieved also for non-planar scanning objects.

Module 3 is about the spectral reflectance recovery or estimation problem. This can be considered as the task of finding a mapping from camera response space to spectral reflectance space. Unlike camera responses, spectral reflectances are device independent object surface properties. Hence, spectral estimation is the basis for spectral reflectance factor measurements with LSMSC systems.

We first summarized mathematical concepts for physical and empirical model based spectral estimation: for estimation based on a physical model of the acquisition process, we considered a direct inverse approach based on the minimum norm solution (P-MIN), a regularized approach that estimates coordinates in an orthogonalized reflectance space (P-MW) and the Wiener estimation (P-WN). For empirical estimation, we considered a linear model. We discussed the least-squares solution with linear feature mapping (E-PINV-SVD), non-linear feature mapping, and regularized versions of the solution. We also considered reproducing kernel Hilbert space ridge regression (E-KER) and proposed the conditionally positive definite logarithmic kernel function for spectral reflectance estimation [P1].

Then, approaches that impose constraints on the estimation were summarized (ie. positivity or physical realizability of spectral reflectances). We showed in [P2] that constrained estimation is a necessary requirement for spectral density measurements based on estimated reflectance data using a LSMSC systems. We used E-KER with link functions to address the problem.

Performance of spectral estimation approaches was compared experimentally for simulated and measured camera response data. We considered various levels of camera response noise and LSMSC system configurations with different number of spectral channels.

We found that average color and spectral measurement errors decrease with the number of spectral channels and increases with noise level. Further, empirical estimation outperformed physical model based recovery under all conditions evaluated. Performance when using measured data was found to be lower than that of simulated data and the discrepancy was specifically high for physical model based estimation. We discussed various potential reasons for this.

The highest performance using measured camera response data was achieved by E-KER (using Gaussian kernel) with average $RMSE = 0.005$ and average $\Delta E_{00} = 0.36$.

Module 4 is related to the training process in empirical spectral estimation. We concentrate on two main aspects, namely training sample selection and adaptive training. Most training sample selection methods reduce the number of available samples based on the premise of preserving relevant information. However, training with this data does not necessarily result in a good estimation model. Application dependent selection schemes account for this and we proposed the *recursive rejection* algorithm for application dependent training set construction [P3]. An advantage of this method over others is that the training set size is determined automatically.

In adaptive training, the estimation model is trained specifically for each individual test sample. This increases the computation burden, but as compared with conventional training, the estimation model often approximates better in a local region around the test sample. We conceptually unified methods from related literature and proposed a general adaptive estimation framework. We experimentally compared some adaptive estimation schemes and found that a densely sampled training set is required to outperform conventional training.

In summary, we showed that the four modules of line-scan multi-spectral imaging system design considered in this work are critical for spectral and color measurement. Both theoretical and practical design aspects were contemplated and relevance of the proposed methods was substantiated consequently by experiments based on image data acquired with our laboratory LSMSC system. Therefore, this dissertation can be highly beneficial for application specific design and development of multi-spectral camera systems.

7.2 Contributions

The author's contributions can be summarized as follows:

Module 1: Acquisition system design

- Development of a computationally economic framework for LSMSC filter selection based on a heuristic reduction of possible filter combinations that have to be evaluated (Section 3.3).
- Usage of Pareto optimality criteria to ease selection of a single filter combinations from a large set, based on multiple objectives (Section 3.3.3).
- Demonstration of experimental case studies for various application scenarios (Section 3.4).
- Selection of an optimal set of color filters for a 12-channel LSMSC (Section 3.4.4).

Module 2: Image registration

- Development of a general image registration framework for LSMSC systems that applies to planar and non-planar scanning objects (Section 4.1.1).
- Design of a geometrical calibration target with rotated checkerboard pattern that is specifically useful for LSMSC systems and allows semi-automatic extraction of dense key-points (Section 4.2.1).
- Development of a polynomial registration model and a uniform cubic B-spline based registration model with grid refinement for LSMSC systems (Section 4.2).
- Development of scene-adaptive registration by means of the stereo imaging principle (Section 4.3)[1]. This contains the proposal of color invariant feature mapping algorithms for pixel correspondence estimation (Section 4.3.1).
- Proof of principle for combined height profile and spectral measurements of non-planar scanning objects by LSMSC imaging.

---

[1] We note that the idea to use the stereo imaging principle for imaging non-flat scanning objects with a LSMSC system is not the author's contribution, but was motivated by Chromasens GmbH.

Module 3: Spectral reflectance estimation

- Systematic summary and conceptual unification of spectral estimation algorithms from literature (Section 5.2 and Section 5.3).
- Implementation and experimental comparison of spectral estimation algorithms (Section 5.5).
- Proposal of the logarithmic kernel function for reproducing kernel Hilbert space based ridge regression in spectral reflectance estimation ([P1], Section 5.3.4).
- Evaluation of multi-spectral imaging based spectral density measurement for print inspection using a constrained estimation scheme ([P2], Section 5.4.3).[2]

Module 4: Empirical estimation model training

- Systematic summary and conceptual unification of global training sample selection and adaptive estimation schemes (Section 6.2, Section 6.3).
- Proposal of an application specific global training set selection method based on a recursive rejection scheme ([P3], Section 6.2.2).
- Illustration of fundamental requirements on training data in adaptive estimation (Section 6.4.2).

7.3 Future research prospects

In the following, we summarize potential future research topics that were identified within scope of this dissertation:

Illumination SPD optimization:

We have considered the filter selection problem in Chapter 3 and mentioned that optimizing spectral properties of the scene illumination is an equally important task. For the specific case of a LSMSC system with LED line light containing individually controllable types of distinct LEDs similar to the one used in our laboratory setup, an open problem is the optimal adjustment of individual LED type currents. This problem could be solved by modeling the physical light mixing process of the scene illumination SPD as a function of LED currents and combining this model with the model of the image acquisition process. Finding optimal LED currents can then be expressed as a minimization problem with respect to a loss function on the basis of system optimality measures presented in Section 3.2.

Image acquisition model characterization:

Experimental results with respect to filter selection (Section 3.4.4), physical model based spectral estimation algorithms (Section 5.5.3) and training of empirical models (Section 6.4) showed a discrepancy between simulated and measured camera response data, which is an indication for a shortcoming in the image acquisition model. We pointed out that for our model, a first improvement could be achieved by obtaining spectral responsivities in a larger spectral range. Other aspects that could lead to an improvement of the model involve for instance systematic experimental evaluation

---

[2] We acknowledge Chromasens GmbH for introducing the negativity problem of spectral estimation based spectral density measurements to the author.

and modeling of the influence of observation and illumination angle deviations on the sensed camera response data, present in LSMSC type systems.

Scene-adaptive color invariant feature mapping:

We have proposed two schemes for color invariant feature mapping of RGB sub-images from LSMSC systems with more than one lens (Section 4.3.1), and achieved good results with a scheme based on a linear transformation derived from empirical data. If corresponding RGB subspaces do not intersect, the empirical solution only minimizes residuals in feature space, but the mapping is not uniquely defined for all elements of the RGB subspace. We conducted a preliminary study that involves a image feature based extraction of corresponding image regions from RGB sub-images. The color invariant feature mapping can then be obtained from these image regions in a scene-adaptive fashion. Results indicate an improvement as compared with the empirical feature mapping described in Section 4.3.1. Further, preliminary experiments also indicate that non-linear transformations (reproducing kernel Hilbert space based ridge regression with Gaussian kernel function) can improve the results, but a systematic evaluation for different scene types is pending.

3D surface reconstruction:

Solving the correspondence problem between RGB sub-images of a LSMSC with more than one lens (Section 4.3) and non-planar scanning objects make pixel-wise disparity of sub-images available. So far, this information is only used for image registration but we have already shown that it can also be used to reconstruct the height profile of the scanning object scene (Section 4.4.4). Further, an additional calibration step would allow transforming pixel height profiles in image coordinate system in a height profile in the object coordinate system and hence allow image based spatial measurements on 3D objects.

Spectral reflectance measurements of 3D objects:

The aforementioned work in Chapter 4 related to non-planar scanning objects is also the basis for spectral reflectance measurements on non-flat scanning objects. Depending on the complexity of the surface structure, measurement conditions will deviate from what is required for norm conform spectral reflectance factor and device independent color measurements, and the influence on measurement quality is yet to be evaluated.

7.4 Amplio Resumen

Esta tesis versa sobre el diseño de sistemas multiespectrales de escáner en línea LSMSC, aunque muchos de los aspectos tratados pueden aplicarse también a otros sistemas multiespectrales. El diseño de estos sistemas y el desarrollo de aplicaciones de imagen específicas requiere considerar muchos aspectos. Nos centramos en la optimización de las propiedades de los componentes ópticos del sistema (Módulo 1), el registro de imágenes en LSMSC (Módulo 2) y los modelos matemáticos para la recuperación o estimación de la información de la reflectancia espectral de superficies a partir de datos de respuestas de sensores en dichos sistemas (Módulos 3 y 4).

Ya que en este trabajo se presentan estudios pertenecientes a varios campos o dominios diferentes, hemos estructurado la tesis en capítulos ligados al contenido de cada módulo, cada uno de los cuales incluye motivación, método experimental, resultados y conclusiones detalladas. A continuación, presentamos los distintos módulos con una perspectiva general, y luego nos limitamos a ofrecer un resumen de las conclusiones y hallazgos más relevantes.

El **Módulo 1** describe el diseño del sistema de captura. Nos hemos centrado en las propiedades espectrales de los componentes ópticos del sistema, considerando el diseño del mismo como un problema de optimización. Hemos estudiado y resumido criterios de optimización basados en la proyección de subespacios vectoriales y en el error obtenido en la estimación espectral. Hemos demostrado que los criterios basados en el error de estimación espectral son en muchas ocasiones la mejor opción desde un punto de vista práctico.

Para abordar la tarea específica de la selección de filtros de color en sistemas LSMSC, hemos propuesto un sistema de trabajo que incorpora relaciones heurísticas que hacen factible llevar a cabo la selección en caso de tener un número muy elevado de combinaciones posibles de filtros como punto de partida. Además, usando el principio de Pareto-optimización, simplificamos notablemente la selección de la combinación óptima de filtros. Hemos demostrado nuestra propuesta de sistema de trabajo para los casos de 3, 6 y 12 canales en sistemas LSMSC basados en un principio específico de formación de imagen, que utiliza un sensor RGB en línea con varias lentes que tienen filtros de color acoplados. Este sistema fue una invención reciente de la empresa Chromasens GmbH.

La combinación óptima de filtros para el sistema de 12 canales se implementó en la práctica en nuestro prototipo de laboratorio. Los resultados de nuestros experimentos con este prototipo ilustran las magníficas prestaciones para la estimación de datos espectrales y colorimétricos (con un promedio de $RMSE < 0.01$ y una diferencia de color promedio $\Delta E_{00} < 0.6$), pero también pusieron de manifiesto claras limitaciones en el modelo utilizado para la simulación del proceso de adquisición de imágenes.

Contribuciones:

- Desarrollo de un sistema económico computacionalmente para la selección de filtros en sistemas LSMSC, basado en una reducción heurística de las posibles combinaciones de filtros que deben ser evaluadas (ver la sección 3.3)
- Uso de criterios de Pareto-optimización para facilitar la selección de una única combinación de filtros a partir de un conjunto amplio de combinaciones, basados en múltiples objetivos de calidad (ver la sección 3.3.3).
- Demostración del estudio experimental de casos para varios escenarios de aplicación (ver la sección 3.4).

- Selección de un conjunto óptimo de filtros de color para un sistema LSMSC de 12 canales (ver la sección 3.4.4).

El **Módulo 2** trata sobre el registro de imágenes multi-canal para el sistema de escáner en línea. Hemos demostrado que la desalineación de las imágenes degrada la calidad de los datos espectrales estimados de forma severa. Hemos propuesto un sistema de trabajo general para el registro de imágenes en sistemas LSMSC.

Para objetos planos, el proceso de registro se puede considerar como una calibración de tipo geométrico. Hemos diseñado una carta de calibrado específica para cámaras de escáner en línea, y un método semi-automático para extraer los datos de desplazamiento de píxeles en los puntos clave obtenidos de las imágenes. Basándonos en estos datos, hemos identificado una forma funcional de tipo polinómica para las curvas de desplazamiento en nuestro sistema LSMSC de laboratorio. Consecuentemente, se ha desarrollado un modelo de ajuste polinómico y también un modelo basado en B-splines que incluye un sistema de afinado progresivo de la malla de puntos. Hemos evaluado empíricamente ambos modelos, y hemos confirmado que es posible registrar las imágenes con una precisión por debajo de un píxel (desplazamiento residual promedio menor de 0.1 pixel).

Para objetos con volumen (no planos), hemos demostrado que aplicar únicamente la calibración geométrica no resulta adecuado, ya que se introducen cambios de perspectiva que esta calibración no es capaz de descontar entre las imágenes multi-canal registradas. Hemos propuesto una solución para este problema que utiliza un método de estimación de correspondencia puntual específico para la escena capturada. Este método se basa en la igualación de bloques con una precisión por debajo del píxel. La igualación de bloques se lleva a cabo en imágenes en escala de gris que se obtienen tras aplicar una transformación invariante a cambios de color que ha sido desarrollada específicamente para sistemas LSMSC. Tras este proceso, pueden registrarse correctamente las sub-imágenes RGB utilizando la correspondencia entre píxeles estimada.

Hemos comparado transformaciones invariantes a los cambios de color basadas en modelos físicos y empíricos para nuestro sistema LSMSC de laboratorio, encontrando que los modelos empíricos superan con mucho en prestaciones a los modelos físicos. Los resultados de la evaluación de la estimación de puntos correspondientes muestran que es posible registrar de forma muy precisa (con errores residuales por debajo de un píxel) las sub-imágenes de objetos con volumen.

Contribuciones:

- Desarrollo de un sistema de trabajo para el registro de imágenes en sistemas LSMSC, que puede aplicarse tanto a objetos planos como a objetos con volumen (ver la sección 4.1.1).
- Desarrollo de una carta de calibrado geométrico con un patrón de tablero de ajedrez rotado, que es útil específicamente para sistemas LSMSC, y que permite la extracción semi-automática de una nube densa de puntos (ver la sección 4.2.1).
- Desarrollo de un modelo de registro polinómico y de un modelo basado en B-splines con refinado de la malla de puntos, específicos para sistemas LSMSC (ver la sección 4.2).
- Desarrollo de un método de registro adaptativo utilizando el principio de imagen

estéreo (ver la sección 4.3)[3]. Este método contiene la propuesta de algoritmos basados en transformaciones invariantes a los cambios de color, para estimar la correspondencia píxel a píxel (ver la sección 4.3.1).

- Presentación de un ejemplo característico del principio de operación para la obtención combinada de imágenes de perfil de altura y datos espectrales en medidas realizadas en objetos volumétricos con sistemas de imagen LSMSC.

El **Módulo 3** trata el problema de la recuperación o estimación de reflectancias. Este problema se puede describir como la tarea de encontrar un mapeado desde el espacio de respuestas de sensores al espacio de reflectancia espectral. A diferencia de las respuestas de sensores, la reflectancia espectral es una propiedad de la superficie objeto que es independiente del dispositivo utilizado para capturar su imagen. En consecuencia, la estimación espectral es la base sobre la cual se fundamentan las medidas de factor de reflectancia espectral realizadas con sistemas LSMSC.

En este módulo, primero hemos resumido los conceptos matemáticos necesarios para comprender los algoritmos basados en modelos físicos y también los algortimos puramente empíricos de estimación espectral: en el caso de algoritmos basados en modelos físicos, hemos considerado un enfoque de cálculo de la matriz inversa del proceso de formación de imagen, obteniendo así la solución con mínima norma (P-MIN) para el problema de estimación espectral, así como un enfoque regularizado que estima las coordenadas en un espacio de representación ortogonal de las reflectancias (P-MW), y finalmente el modelo de estimación de Wiener (P-WN).

Para el caso de estimación puramente empírica, hemos considerado como base un modelo lineal de relación entre respuestas de sensores y reflectancias. Hemos discutido la solución por mínimos cuadrados con mapeado lineal (E-PINV-SVD) y también no lineal, con regularización incluida. También hemos utilizado un enfoque regularizado de regresión en espacios kernel de Hilbert reproducibles (E-KER), y hemos propuesto la función de kernel logarítmico, que es condicionalmente definida positiva, para el problema de la estimación espectral [P1].

Además, hemos resumido algunos enfoques del prolema de estimación que imponen condiciones sobre las soluciones obtenidas para las reflectancias espectrales (por ejemplo, positividad o que sean realizables físicamente). Como se muestra en [P2], la imposición de estas condiciones se vuelve necesaria si se pretende utilizar los datos de reflectancia obtenidos para medidas de densidad espectral en muestras de tinta impresa cuyo espectro de reflectancia se ha estimado utilizando un sistema LSMSC. En nuestro caso, utilizamos funciones de enlace y el algoritmo de estimación E-KER para abordar este problema.

Hemos comparado mediante varios experimentos la calidad de la estimación espectral para varios enfoques diferentes, utilizando datos de respuestas de sensores tanto simulados como reales. Para los datos simulados, hemos utilizado varios niveles de ruido en las respuestas de los sensores y varias configuraciones del sistema LSMSC con diferente número de canales.

Nuestros resultados indican que los errores promedio (tanto colorimétricos como espectrales) disminuyen conforme aumenta el número de canales del sistema, y aumentan conforme crece el nivel de ruido en las respuestas de sensores. Además, encontramos que los enfoques empíricos ofrecen mejores resultados para la estimación espectral que los enfoques basados en modelos físicos, en todas las condiciones experimentales evaluadas.

---

[3] Debemos señalar aquí que la idea de utilizar este principo de imagen estéreo para los objetos con volumen no es una contribución del autor, sino que se basó en una sugerencia de Chromasens GmbH.

La calidad de la estimación obtenida utilizando datos reales de respuestas de sensores fue inferior a la encontrada con datos simulados, y la discrepancia era mucho más patente para enfoques basados en modelos físicos de adquisición de imágenes. Hemos discutido varias razones que explican potencialmente este efecto. La mejor calidad en la estimación utilizando datos reales de respuestas de sensores fue obtenida con el algoritmo E-KER (utilizando una función de kernel Gaussiana), resultando en unos valores promedio de $RMSE = 0.005$ y $\Delta E_{00} = 0.36$.

Contribuciones:

- Resumen sistemático y unificación conceptual de los algoritmos de estimación espectral a partir de referencias bibliográficas (ver la sección 5.2 y la sección5.3).
- Implementación y comparación experimental de algoritmos de estimación espectral (ver la sección 5.5).
- Propuesta de la función de kernel logarítmica para la regresión regularizada utilizando espacios kernel de Hilbert reproducibles para la estimación espectral ([P1],ver la sección 5.3.4).
- Evaluación de medidas de densidad espectral basadas en imágenes multiespectrales, obtenidas utilizando un método con imposición de condiciones sobre las soluciones ([P2], ver la sección 5.4.3)[4].

El **Módulo 4** se relaciona con el proceso de entrenamiento para la estimación espectral puramente empírica. Nos concentramos en dos aspectos primordiales, a saber, la selección del conjunto de muestras de entrenamiento y el entrenamiento adaptativo. La mayoría de los métodos de selección de muestras de estrenamiento reduce el número de muestras a partir de un conjunto inicial basándose en las premisas de preservar la información relevante en el conjunto de datos. Sin embargo, entrenar con esta selección de datos no resulta necesariamente en la obtención de un modelo adecuado para la estimación espectral.

Los métodos de selección que son dependientes de la aplicación final son capaces de solucionar este inconveniente, y nosotros hemos propuesto el algoritmo de *rechazo recursivo* para la construcción de conjuntos de muestras de entrenamiento dependientes de la aplicación final [P3]. Este algoritmo presenta la ventaja sobre otros enfoques de que el tamaño final del conjunto de muestras seleccionadas se determina automáticamente.

En el entrenamiento adaptativo, el modelo de estimación se entrena específicamente para cada una de las muestras que van a estimarse. Esto supone incrementar el coste computacional, pero si se compara con el sistema de entrenamiento convencional, el modelo de estimación obtenido a menudo es capaz de aproximar mejor el resultado en una región local alrededor de la muestra que se intenta recuperar. Hemos unificado conceptualmente varios métodos extraídos de la literatura, y hemos propuesto un sistema de trabajo general para el entrenamiento adaptativo. También hemos comparado varios métodos de estimación adaptativa y convencional, y hemos encontrado que para que los métodos adaptativos obtengan mejores resultados que los convencionales, es necesario contar con un muestreado lo suficientemente denso del espacio de datos de entrenamiento.

---

[4] El autor agradece a Chromasens el haber introducido el problema causado por la negatividad de algunas soluciones obtenidas en el marco de las medidas de densidad espectral basadas en sistemas de imagen multiespectrales

Contribuciones:

- Resumen sistemático y unificación conceptual de métodos de entrenamiento globales y adaptativos (ver la sección 6.2, y la sección 6.3).
- Propuesta de un algoritmo global de selección de muestras de entrenamiento dependiente de la aplicación final, basado en un esquema recursivo de rechazo ([P3], ver la sección 6.2.2).
- Ilustración de los requisitos fundamentales para el conjunto inicial de muestras de entrenamiento utilizado en algoritmos de selección adaptativos (ver la sección 6.4.2).

A modo de resumen final, hemos mostrado que los cuatro módulos del diseño del sistema de imagen multiespectral de escáner en línea que hemos considerado en el presente trabajo resultan críticos para obtener medidas de calidad tanto espectrales como colorimétricas. Hemos contemplado aspectos del diseño tanto teóricos como prácticos, y la relevancia de los métodos propuestos se ha demostrado suficientemente mediante diferentes experimentos realizados a partir de datos de imágenes adquiridas por nuestro sistema LSMSC de laboratorio. En consecuencia, esta disertación puede resultar altamente beneficiosa para el diseño y el desarrollo de sistemas de captura multiespectrales.

A
Appendix 1:

Contents

A.1 Conventions for notation and graphical illustrations

Boxplot

Boxplots are used to graphically depict statistical descriptors of numerical datasets. Some items of a boxplot are not defined uniquely in the mathematical literature. Boxplots used in this work obey the following conventions (see Figure A.1 for an example of a boxplot):

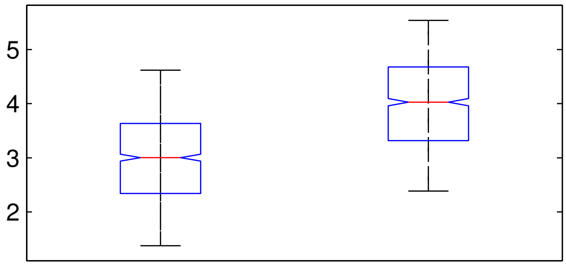
Blue top and bottom line of the box: indicate the first and third quartiles (p_{25} and p_{75} respectively).

Red line inside the box: second quartile (p_{50}), ie. the median of the dataset.

End of lower whisker: 5th percentile (p_5).

End of upper whisker: 95th percentile (p_{95}).

Notch in the box: In plots with multiple boxes, the notches offer a rough indication on the significance of the difference of medians, ie. two medians are statistically different at the 5% significance level, if their notch intervals do not overlap. The extremes of the notches for a dataset with n data points correspond to $p_{50} - 1.57(p_{75} - p_{25})/\sqrt{n}$ and $p_{50} + 1.57(p_{75} - p_{25})/\sqrt{n}$.



Variable 1	**Variable 2**
Mean: 3.0	Mean: 4.0
Min: -1.0	Min: 1.2
Max: 6.1	Max: 7.8

Figure A.1: Example boxplots of a bi-variate dataset of $n = 1000$ normally distributed samples per group with different data mean.

List of symbols and operators

Table A.1: List of symbols and operators.

Symbol or operator	Description
$a \in \mathbb{R}$	Scalar
$\mathbf{a} = [a_1, a_2, ...a_m]^T \in \mathbb{R}^{m\times 1}$	Vector
$\mathbf{1} = [1, ...1]^T \in \mathbb{R}^{m\times 1}$	All-ones vector
$\mathbf{0} = [0, ...0]^T \in \mathbb{R}^{m\times 1}$	All-zeros vector
$A = \begin{pmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,n} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m,1} & a_{m,2} & \cdots & a_{m,n} \end{pmatrix} \in \mathbb{R}^{m\times n}$	Matrix
$\mathrm{diag}(\mathbf{a}) = \begin{pmatrix} a_1 & 0 & 0 & 0 \\ 0 & a_2 & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & a_m \end{pmatrix} \in \mathbb{R}^{m\times m}$	Operator that creates a square diagonal matrix with the element of vector **a** on the main diagonal
$I_m = \mathrm{diag}(\mathbf{1}) \in \mathbb{R}^{m\times m}$	Identity matrix
$\mathcal{S} = \{\mathbf{a}_1, \mathbf{a}_2, ...\mathbf{a}_l\}$	Set of l vectors
$\mathcal{T} = \{A_1, A_2, ...A_l\}$	Set of l matrices
$\lvert\mathcal{S}\rvert$	Cardinality of set $\mathcal{S}$
$\lfloor a \rfloor = max\{m \in \mathbb{Z} \mid m \leq a\}$	Operator that rounds a to the nearest integer less than or equal to a
$\mathrm{Tr}(A) = \sum_{i=1}^{m} a_{i,i}$	Trace of a square matrix A
$\lVert\mathbf{a}\rVert = \sqrt{\sum_{i=1}^{m} a_i^2}$	Euclidean norm of vector **a**
$\lVert A\rVert_F = \sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n} a_{i,j}^2}$	Frobenius norm of matrix A

List of variables

Table A.2: List of variables.

Variable	**Description**
$n \in \mathbb{N}^+$	Number of channels
$m \in \mathbb{N}^+$	Number of discrete wavelength bands
$\mathbf{x} = [x_1, x_2, ...x_n]^T \in \mathbb{R}^{n\times 1}$	Camera response vector
$\mathbf{r} = [r_1, r_2, ...r_m]^T \in \mathbb{R}^{m\times 1}$	Spectral reflectance vector
$\tilde{\mathbf{r}} = [\tilde{r}_1, \tilde{r}_2, ...\tilde{r}_m]^T \in \mathbb{R}^{m\times 1}$	Estimated spectral reflectance vector
$X = [\mathbf{x}_1, \mathbf{x}_2, ...\mathbf{x}_l] \in \mathbb{R}^{n\times l}$	Matrix of l camera response vectors stacked in columns
$R = [\mathbf{r}_1, \mathbf{r}_2, ...\mathbf{r}_l] \in \mathbb{R}^{m\times l}$	Matrix of l reflectance vectors stacked in columns
$Y = [\mathbf{y}_1, \mathbf{y}_2, ...\mathbf{y}_n] \in \mathbb{R}^{m\times n}$	Spectral responsivity of n-channel mutlispectral acquisition device
$\mathbf{l} = [l_1, l_2, ...l_m]^T \in \mathbb{R}^{m\times 1}$	Spectral power distribution of scene illumination
$W = \text{diag}(\mathbf{l})Y \in \mathbb{R}^{m\times n}$	Effective responsivity matrix that combines the effect of the spectral responsivity of the acquisition system Y and scene illumination $\mathbf{l}$
$\mathcal{S} = \{(\mathbf{x}_1, \mathbf{r}_1), (\mathbf{x}_2, \mathbf{r}_2), ...(\mathbf{x}_l, \mathbf{r}_l)\} \subset \mathbb{R}^{n\times 1} \times \mathbb{R}^{m\times 1}$	Set of l camera responses $\mathbf{x} \in \mathbb{R}^{n\times 1}$ and corresponding reflectances $\mathbf{r} \in \mathbb{R}^{m\times 1}$

Units of measurements

Table A.3: Units of measurement.

Symbol	**Unit**	**Description**
λ	nm	Wavelength
Y	AU	Camera responsivity
W	AU	Effective camera responsivity
f	Hz	Frequency of light
Q	J	Radiant energy
Φ	W	Radiant flux
L	W/m^{2}sr	Spectral radiance

A.2 Solution to vectorial least-squares regression problem

Recall the minimization problem for the vectorial case of the linear regression problem, using sum of least squares cost function:

$$\arg\min_{A} \left(\frac{1}{2} \sum_{i=1}^{l} \| A\boldsymbol{\phi}(\mathbf{x}_i) - \mathbf{r}_i \|^2 \right), \tag{A.1}$$

To find the solution, we set the derivate with respect to A zero and solve for A

$$\frac{1}{2} \left(\frac{d}{dA} \sum_{i=1}^{l} (A\boldsymbol{\phi}(\mathbf{x}_i) - \mathbf{r}_i)^T (A\boldsymbol{\phi}(\mathbf{x}_i) - \mathbf{r}_i) \right) = 0, \tag{A.2}$$

which is

$$\frac{1}{2} \left(\frac{d}{dA} \operatorname{Tr} \left\{ (A\Phi(X) - R)^T (A\Phi(X) - R) \right\} \right) = 0, \tag{A.3}$$

where $\Phi(X) \in \mathbb{M}(p, l)$ and $R \in \mathbb{M}(m, l)$. To simplify, let $\Phi = \Phi(X)$. It follows that

$$\begin{aligned} \frac{1}{2} \left(\frac{d}{dA} \operatorname{Tr} \left\{ \Phi^T A^T A \Phi - \Phi^T A^T R - R^T A \Phi + R^T R \right\} \right) &= 0 \\ \frac{1}{2} \left(\frac{d}{dA} \operatorname{Tr}\{\Phi^T A^T A \Phi\} - \frac{d}{dA} \operatorname{Tr}\{\Phi^T A^T R\} - \frac{d}{dA} \operatorname{Tr}\{R^T A \Phi\} + \frac{d}{dA} \operatorname{Tr}\{R^T R\} \right) &= 0. \end{aligned} \tag{A.4}$$

We recall from [122] the 2nd order derivate $\frac{d}{dA} \operatorname{Tr}(B^T A^T A B) = 2ABB^T$. Further, $\frac{d}{dA} \operatorname{Tr}\{B^T A^T C\} = \frac{d}{dA} \operatorname{Tr}\{C^T A B\} = CB^T$, and $\frac{d}{dA} \operatorname{Tr}\{R^T R\} = 0$. Substituting Φ for B and R for C it follows that

$$\begin{aligned} \frac{1}{2} \left(2A\Phi\Phi^T - 2R\Phi^T \right) &= 0 \\ A\Phi\Phi^T - R\Phi^T &= 0 \\ A &= R\Phi^T (\Phi\Phi^T)^{-1}. \quad \square \end{aligned} \tag{A.5}$$

A.3 Computation of the CIEDE2000 color difference

Let $[L_1^*, a_1^*, b_1^*]$ and $[L_2^*, a_2^*, b_2^*]$ be a pair of CIE-L*a*b* color coordinates, for which CIEDE2000 color difference ΔE_{00} is to be computed[1]. By definition, let subscript 1 denote the reference sample for which the color difference is computed with respect to the sample with subscript 2. Conceptually following Sharma et al. [123], computation of the CIEDE2000 color difference [40] can be divided into three parts:

1. Calculate C_i' and h_i':

$$C_{i,ab}^* = \sqrt{(a_i^*)^2 + (b_i^*)^2},\ i = 1,2 \tag{A.6}$$

$$\overline{C}_{ab}^* = \frac{C_{1,ab}^* + C_{2,ab}^*}{2} \tag{A.7}$$

$$G = \frac{1}{2}\left(1 - \sqrt{\frac{(\overline{C}_{ab}^*)^7}{(\overline{C}_{ab}^*)^7 + (25)^7}}\right) \tag{A.8}$$

$$a_i' = (1 + G)\, a_i^*,\ i = 1,2 \tag{A.9}$$

$$C_i' = \sqrt{(a_i')^2 + (b_i^*)^2},\ i = 1,2 \tag{A.10}$$

$$h_i' = \begin{cases} 0 & b_i^* = a_i' = 0 \\ \tan^{-1}(b_i^* a_i') & \text{otherwise} \end{cases} \tag{A.11}$$

2. Calculate $\Delta L'$, $\Delta C'$ and $\Delta H'$

$$\Delta L' = L_2^* - L_1^* \tag{A.12}$$

$$\Delta C' = C_2^* - C_1^* \tag{A.13}$$

$$\Delta h' = \begin{cases} 0 & C_1' C_2' = 0 \\ h_2' - h_1' & C_1' C_2' \neq 0;\ |h_2' - h_1'| \leq 180^\circ \\ (h_2' - h_1') - 360 & C_1' C_2' \neq 0;\ (h_2' - h_1') > 180^\circ \\ (h_2' - h_1') + 360 & C_1' C_2' \neq 0;\ (h_2' - h_1') < -180^\circ \end{cases} \tag{A.14}$$

$$\Delta H' = 2\sqrt{C_1' C_2'}\, \sin\left(\frac{\Delta h'}{2}\right) \tag{A.15}$$

---

[1] Note that in this section, some scalar quantities such as L_1^* or others are denoted in uppercase. This notation is used to be consistent with existing literature.

3. Calculate ΔE_{00} color difference

$$\overline{L}' = \frac{1}{2}(L_1^* + L_2^*) \tag{A.16}$$

$$\overline{C}' = \frac{1}{2}(C_1' + C_2') \tag{A.17}$$

$$h' = \begin{cases} \frac{h_1'+h_2'}{2} & C_1'C_2' \neq 0;\ |h_1' - h_2| \leq 180^\circ \\ \frac{h_1'+h_2'+360^\circ}{2} & C_1'C_2' \neq 0;\ |h_1' - h_2| > 180^\circ;\ (h_1' + h_2') < 360^\circ \\ \frac{h_1'+h_2'-360^\circ}{2} & C_1'C_2' \neq 0;\ |h_1' - h_2| > 180^\circ;\ (h_1' + h_2') \geq 360^\circ \\ (h_2' + h_1') & C_1'C_2' = 0 \end{cases} \tag{A.18}$$

$$T = 1 - 0.17\ \cos(\overline{h}' - 30^\circ) + 0.24\ \cos(2\overline{h}') + 0.32\ \cos(3\overline{h}' + 6^\circ) - 0.20\ \cos(4\overline{h}' - 63^\circ) \tag{A.19}$$

$$\Delta\theta = 30\ \exp\left\{ -\left[\frac{\overline{h}' - 275^\circ}{25}\right]^2 \right\} \tag{A.20}$$

$$R_C = 2\sqrt{\frac{\overline{C}'^7}{\overline{C}'^7 + 25^7}} \tag{A.21}$$

$$S_L = 1 + \frac{0.015(\overline{L}' - 50)^2}{\sqrt{20 + (\overline{L}' - 50)^2}} \tag{A.22}$$

$$S_C = 1 + 0.045\,\overline{C}' \tag{A.23}$$

$$S_H = 1 + 0.015\,\overline{C}'\,T \tag{A.24}$$

$$R_T = -sin(2\Delta\theta)\,R_C \tag{A.25}$$

$$\Delta E_{00} = \sqrt{\left(\frac{\Delta L'}{k_L S_L}\right)^2 + \left(\frac{\Delta C'}{k_C S_C}\right)^2 + \left(\frac{\Delta H'}{k_H S_H}\right)^2 + R_T \left(\frac{\Delta C'}{k_C S_C}\right)\left(\frac{\Delta H'}{k_H S_H}\right)}, \tag{A.26}$$

where k_L, k_C and k_H are application dependent parametric weighting factors.

The definition of a reference color sample (indicated by subscript 1) is important when differences in lightness ($\Delta L'$), hue ($\Delta H'$) and chroma ($\Delta C'$) are to be computed from a sample pair. For the calculation of the CIEDE2000 color difference, the definition of a reference color sample is insignificant because the color difference equation is a symmetric function.

In this work, the weighting factors k_L, k_C and k_H are set to unity, so $k_L = k_C = k_H = 1$.

A.4 Unbiased estimator of the covariance matrix of a random variable

Consider a set of l independent observations $\mathbf{x}_1, \mathbf{x}_2, ...\mathbf{x}_l$ of a n-dimensional vector describing a random variable. The unbiased estimator of the covariance matrix $\mathrm{cov}(\mathbf{x}) = E[(\mathbf{x} - E[\mathbf{x}])(\mathbf{x} - E[\mathbf{x}])^T] \in \mathbb{R}^{n \times n}$ of the random variable $\mathbf{x}$ is the matrix

$$K_{\mathbf{x}} = \frac{1}{l-1} \sum_{k=1}^{l} (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^T, \tag{A.27}$$

where $\mathbf{x}_i$ is the i-th observation of the random variable and

$$\bar{\mathbf{x}}_i = \frac{1}{l} \sum_{k=1}^{l} \mathbf{x}_i \tag{A.28}$$

is the n-dimensional mean vector of the observations. We also use $\mathrm{cov}(X)$ to denote the unbiased estimator of random variable $\mathbf{x} \in \mathbb{R}^{n \times 1}$ when the observations are stacked in columns such that $X = [\mathbf{x}_1, \mathbf{x}_2, ...\mathbf{x}_l] \in \mathbb{R}^{n \times l}$.

B

Appendix 2:

Contents

List of Figures

List of Tables

Glossary

BSRDF Bidirectional spectral reflectance distribution function 9, 10, 211

CIE Commission Internationale de l'Eclairage - the International Commission on Illumination 12, 15, 16, 24, 32, 34, 53, 150, 152, 155, 207, 209, 211

CMFs CIE color matching functions 15, 16, 43, 45, 47, 57, 211

CPD Conditionally positive definite. 119, 211

DS1 Dataset 1 (Pantone solid chips). See Table 2.1 in Section 2.5. 38–40, 48, 54, 95–97, 99, 125, 126, 128, 136, 147, 149–155, 207–209, 211

DS2 Dataset 2 (HKS-N chips). See Table 2.1 in Section 2.5. 38–40, 147–155, 209, 211

DS3 Dataset 3 (HKS-K chips). See Table 2.1 in Section 2.5. 38–40, 147, 148, 211

DS4 Dataset 4 (RAL chips). See Table 2.1 in Section 2.5. 38–40, 95, 96, 98, 100, 207, 208, 211

E-KER Empirical estimation using RKHS ridge regression. See Section 5.3.3. 119, 121, 123–126, 129, 131–134, 136, 140, 159, 165, 166, 209, 211

E-PINV-SVD Empirical estimation using linear least-squares method and linear feature mapping, computed using SVD. See Section 5.3.1.A. 116, 124–126, 129–134, 136, 140, 143, 144, 147, 152, 159, 165, 209, 211

LSMSC Line-scan multi-spectral camera 3, 4, 12, 18, 20, 22–24, 27–32, 37, 42, 49, 51, 53, 61–63, 66, 67, 70, 71, 74–76, 83–85, 87, 89, 90, 97–101, 105–107, 120, 124–126, 133–137, 142, 143, 158–165, 167, 207, 208, 211

Noise Scenario 1 Camera response noise simulation scenario. See Subsection 2.3.3.B. 48, 125, 126, 133, 211

Noise Scenario 2 Camera response noise simulation scenario. See Subsection 2.3.3.B. 53, 125, 131, 133, 147, 211

P-MIN Physical model based estimation using minimum norm solution. See Section 5.2.1. 112, 125–128, 131–134, 136, 159, 165, 209, 211

P-MW Physical model based estimation using a regularized minimum norm solution. See Section 5.2.1. 112, 125–134, 136, 159, 165, 209, 211

P-WN Physical model based recovery using Wiener estimation. See Section 5.2.2. 113, 125, 126, 128, 129, 131–134, 136, 159, 165, 209, 211

PCA Principal component analysis 14, 116, 211

PD Positive definite. 119, 120, 211

PRNU Pixel response non-uniformity 26, 211

RKHS Reproducing kernel Hilbert space. See Subsection 5.3.3. 119, 122, 136, 146, 211

RR Recursive rejection algorithm. See Subsection 6.2.2. 143, 156, 211

SC1 MSLSC system configuration 1. See Section 2.5. 37, 61–65, 67, 95–98, 120, 124, 143, 208, 211

SC2 MSLSC system configuration 2. See Section 2.5. 37, 49, 50, 55, 59–67, 70, 71, 76, 84, 90, 95, 96, 99–101, 106, 125–127, 130–134, 136, 147, 151, 154, 208, 209, 211

SC3 MSLSC system configuration 3. See Section 2.5. 37, 49, 50, 58–61, 66, 125, 130–132, 136, 147, 151, 208, 211

SC4 MSLSC system configuration 4. See Section 2.5. 37, 49, 55–58, 66, 125, 130–132, 136, 147, 150–152, 208, 211

SNR Signal-to-noise ratio 27, 62, 125, 126, 128–130, 211

SPD Spectral power distribution 12, 15, 24, 28, 30, 67, 137, 161, 211

SVD singular value decomposition 14, 115–118, 211

Author Index

Author's Bibliography

[P1] T. Eckhard, E. M. Valero, J. Hernández-Andrés, and V. Heikkinen, "Evaluating logarithmic kernel for spectral reflectance estimation—effects on model parametrization, training set size, and number of sensor spectral channels," *J. Opt. Soc. Am. A*, vol. 31, no. 3, pp. 541–549, 2014.

[P2] T. Eckhard, M. Klammer, E. M. Valero, and J. Hernández-Andrés, "Improved spectral density measurement from estimated reflectance data with kernel ridge regression," in *Image and Signal Processing*, pp. 79–86, Springer, 2014.

[P3] T. Eckhard, E. M. Valero, J. Hernández-Andrés, and M. Schnitzlein, "Adaptive global training set selection for spectral estimation of printed inks using reflectance modeling," *Applied Optics*, vol. 53, no. 4, pp. 709–719, 2014.

Bibliography

[1] A. Davenport, *The history of photography: an overview.* UNM Press, 1991.

[2] J. M. Palmer and B. G. Grant, *The art of radiometry.* SPIE Press Bellingham, 2010.

[3] N. Ohta and A. Robertson, *Colorimetry: fundamentals and applications.* John Wiley & Sons, 2006.

[4] J. Dorsey, H. Rushmeier, and F. Sillion, *Digital modeling of material appearance.* Morgan Kaufmann, 2010.

[5] H.-C. Lee, *Intorduction to Color Imaging Science.* Cambridge University Press Cambridge, 2005.

[6] "Graphic technology – Spectral measurement and colorimetric computation for graphic arts images," 2009.

[7] M. G. K. Vogl Howard E., "Color measurement on substrates with optical brightening agents," in *Eighteenth IAPRI World Packaging Conference* (J. Singh, ed.), 2012.

[8] L. T. Maloney, "Evaluation of linear models of surface spectral reflectance with small numbers of parameters," *J. Opt. Soc. Am. A*, vol. 3, no. 10, pp. 1673–1683, 1986.

[9] P. Morovic and G. D. Finlayson, "Metamer-set-based approach to estimating surface reflectance from camera rgb," *J. Opt. Soc. Am. A*, vol. 23, no. 8, pp. 1814–1822, 2006.

[10] J. Cohen, "Dependency of the spectral reflectance curves of the munsell color chips," *Psychonomic Science*, vol. 1, no. 1-12, pp. 369–370, 1964.

[11] J. Parkkinen, J. Hallikainen, and T. Jaaskelainen, "Characteristic spectra of munsell colors," *J. Opt. Soc. Am. A*, vol. 6, no. 2, pp. 318–322, 1989.

[12] J. L. Dannemiller, "Spectral reflectance of natural objects: how many basis functions are necessary?," *J. Opt. Soc. Am. A*, vol. 9, no. 4, pp. 507–515, 1992.

[13] J. Shlens, "A tutorial on principal component analysis," *arXiv preprint arXiv:1404.1100*, 2014.

[14] C. L. Lawson and R. J. Hanson, *Solving least squares problems*, vol. 161. SIAM, 1974.

[15] C.-I. Chang, *Hyperspectral data processing: algorithm design and analysis.* John Wiley & Sons, 2013.

[16] F. Agahian and B. Funt, "Outlier modeling for spectral data reduction," *J. Opt. Soc. Am. A*, vol. 31, no. 7, pp. 1445–1452, 2014.

[17] L. T. Maloney and B. A. Wandell, "Color constancy: a method for recovering surface spectral reflectance," *J. Opt. Soc. Am. A*, vol. 3, no. 1, pp. 29–33, 1986.

[18] P. D. Burns, *Analysis of image noise in multispectral color acquisition.* PhD thesis, Rochester Institute of Technology, 1997.

[19] G. Wyszecki and W. S. Stiles, *Color science, 2nd edition.* Wiley New York, 1982.

[20] W. S. Stiles and J. M. Burch, “Npl colour-matching investigation: final report (1958),” *Journal of Modern Optics*, vol. 6, no. 1, pp. 1–26, 1959.

[21] N. Speranskaya, “Determination of spectrum color coordinates for 27 normal observers,” *Optics and Spectroscopy*, vol. 7, p. 424, 1959.

[22] E. Herrala and J. Okkonen, “Imaging spectrograph and camera solutions for industrial applications,” *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 10, no. 01, pp. 43–54, 1996.

[23] M. Vilaseca, J. Pujol, M. Arjona, and M. de Lasarte, “Multispectral system for reflectance reconstruction in the near-infrared region,” *Applied Optics*, vol. 45, no. 18, pp. 4241–4253, 2006.

[24] J. Brauers, N. Schulte, and T. Aach, “Multispectral filter-wheel cameras: Geometric distortion model and compensation algorithms,” *Image Processing, IEEE Transactions on*, vol. 17, no. 12, pp. 2368–2380, 2008.

[25] J. Hardeberg, F. J. Schmitt, and H. Brettel, “Multispectral image capture using a tunable filter,” in *Electronic Imaging*, pp. 77–88, International Society for Optics and Photonics, 1999.

[26] N. Gat, “Imaging spectroscopy using tunable filters: a review,” in *AeroSense 2000*, pp. 50–64, International Society for Optics and Photonics, 2000.

[27] S. Blais-Ouellette, O. Daiglea, and K. Taylorc, “The imaging bragg tunable filter,” in *Proc. of SPIE Vol*, vol. 6269, pp. 62695H–1, 2006.

[28] J. Eckhard, T. Eckhard, E. M. Valero, J. L. Nieves, and E. G. Contreras, “Outdoor scene reflectance measurements using a bragg-grating-based hyperspectral imager,” *Applied Optics*, vol. 54, no. 13, pp. D15–D24, 2015.

[29] A. Gebejes, J. Orava, N. Penttinen, V. Heikkinen, J. Hiltunen, and M. Hauta-Kasari, “Color and image characterization of a three ccd seven band spectral camera,” in *Image and Signal Processing*, pp. 96–105, Springer, 2014.

[30] R. F. Lyon and P. M. Hubel, “Eyeing the camera: Into the next century,” in *Color and Imaging Conference*, vol. 1, pp. 349–355, Society for Imaging Science and Technology, 2002.

[31] G. Langfelder, A. Longoni, and F. Zaraga, “Implementation of a multi-spectral color imaging device without color filter array,” in *IS&T/SPIE Electronic Imaging*, pp. 787608–787608, International Society for Optics and Photonics, 2011.

[32] R. Hain, C. J. Kähler, and C. Tropea, “Comparison of ccd, cmos and intensified cameras,” *Experiments in Fluids*, vol. 42, no. 3, pp. 403–411, 2007.

[33] D. Litwiller, “Ccd vs. cmos,” *Photonics Spectra*, vol. 35, no. 1, pp. 154–158, 2001.

[34] J. R. Janesick, *Scientific charge-coupled devices*, vol. 83. SPIE press, 2001.

[35] W. Van Drongelen, *Signal processing for neuroscientists: an introduction to the analysis of physiological signals.* Academic press, 2006.

[36] M. A. López-Álvarez, J. Hernández-Andrés, E. M. Valero, and J. L. Nieves, "Colorimetric and spectral combined metric for the optimization of multispectral systems," in *Proceedings of the 10th Congress of the International Colour Association (AIC'05)*, pp. 1685–1688, Association Internationale de la Couleur, 2005.

[37] F. H. Imai, M. R. Rosen, and R. S. Berns, "Comparative study of metrics for spectral match quality," in *Proceedings of the First European Conference on Colour in Graphics, Imaging and Vision*, pp. 492–496, 2002.

[38] J. Romero, A. García-Beltrán, and J. Hernández-Andrés, "Linear bases for representation of natural and artificial illuminants," *J. Opt. Soc. Am. A*, vol. 14, no. 5, pp. 1007–1014, 1997.

[39] CIE, "Industrial colour-difference evaluation," tech. rep., CIE Publication No. 116-1995, 1995.

[40] CIE, "Improvement to industrial colour-difference evaluation," tech. rep., CIE Publication No. 142-2001, 2001.

[41] B. Schöllkopf, "The kernel trick for distances," *Advances in neural information processing systems*, pp. 301–307, 2001.

[42] P. L. Vora and H. Trussell, "Measure of goodness of a set of color-scanning filters," *J. Opt. Soc. Am. A*, vol. 10, no. 7, pp. 1499–1508, 1993.

[43] P. L. Vora and H. Trussell, "Mathematical methods for the design of color scanning filters," *Image Processing, IEEE Transactions on*, vol. 6, no. 2, pp. 312–320, 1997.

[44] G. Sharma and H. Trussell, "Figures of merit for color scanners," *IEEE Transactions on Image Processing*, vol. 6, no. 7, pp. 990–1001, 1997.

[45] D.-Y. Ng and J. P. Allebach, "A subspace matching color filter design methodology for a multispectral imaging system," *Image Processing, IEEE Transactions on*, vol. 15, no. 9, pp. 2631–2643, 2006.

[46] C. D. Meyer, *Matrix analysis and applied linear algebra.* Siam, 2000.

[47] D. Attewell and R. J. Baddeley, "The distribution of reflectances within the visual environment," *Vision research*, vol. 47, no. 4, pp. 548–554, 2007.

[48] H.-L. Shen, J.-F. Yao, C. Li, X. Du, S.-J. Shao, and J. H. Xin, "Channel selection for multispectral color imaging using binary differential evolution," *Applied Optics*, vol. 53, no. 4, pp. 634–642, 2014.

[49] H. Haneishi, T. Hasegawa, A. Hosoi, Y. Yokoyama, N. Tsumura, and Y. Miyake, "System design for accurately estimating the spectral reflectance of art paintings," *Applied Optics*, vol. 39, no. 35, pp. 6621–6632, 2000.

[50] R. Piché, "Nonnegative color spectrum analysis filters from principal component analysis characteristic spectra," *J. Opt. Soc. Am. A*, vol. 19, no. 10, pp. 1946–1950, 2002.

[51] I. B. Styles, “Selection of optimal filters for multispectral imaging,” *Applied Optics*, vol. 47, no. 30, pp. 5585–5591, 2008.

[52] M. Hauta-Kasari, W. Wang, S. Toyooka, J. Parkkinen, and R. Lenz, “Unsupervised filtering of munsell spectra,” in *Computer Vision—ACCV’98*, pp. 248–255, Springer, 1997.

[53] N.-C. Hu and C.-C. Wu, “Optimal selection of commercial sensors for linear model representation of daylight spectra,” *Applied Optics*, vol. 47, no. 17, pp. 3114–3123, 2008.

[54] G. Novati, P. Pellegri, and R. Schettini, “Selection of filters for multispectral acquisition using the filter vectors analysis method,” in *Electronic Imaging 2004*, pp. 20–26, International Society for Optics and Photonics, 2003.

[55] F. H. Imai, S. Quan, M. R. Rosen, and R. S. Berns, “Digital camera filter design for colorimetric and spectral accuracy,” in *Proceedings of the 3rd International Conference on Multispectral Color Science (Joensuu, Finland)*, pp. 23–26, 2001.

[56] I. S. Chatzis, V. A. Kappatos, and E. S. Dermatas, “Filter selection for multi-spectral image acquisition using the feature vector analysis methods,” in *Intelligent Production Machines and Systems*, p. 283, Elsevier, 2011.

[57] C. M. Bishop, *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Springer, 1st ed. 2006. corr. 2nd printing ed., Oct. 2007.

[58] M. Ehrgott, *Multicriteria optimization*, vol. 2. Springer, 2005.

[59] J. Brauers and T. Aach, “Geometric calibration of lens and filter distortions for multispectral filter-wheel cameras,” *Image Processing, IEEE Transactions on*, vol. 20, no. 2, pp. 496–505, 2011.

[60] T. Eckhard, J. Eckhard, E. M. Valero, and J. L. Nieves, “Nonrigid registration with free-form deformation model of multilevel uniform cubic b-splines: application to image registration and distortion correction of spectral image cubes,” *Applied Optics*, vol. 53, no. 17, pp. 3764–3772, 2014.

[61] T. Tuytelaars and K. Mikolajczyk, “Local invariant feature detectors: a survey,” *Foundations and Trends in Computer Graphics and Vision*, vol. 3, no. 3, pp. 177–280, 2008.

[62] J.-Y. Bouguet, “Camera calibration toolbox for matlab,” 2004.

[63] E. Hecht, *Optics*. Addison Wesley, 4 ed., 2001.

[64] X. Shen, F. Mohd-Zaid, and R. Francis, “Runge phenomenon: A virtual artifact in image processing,” in *The 16th International Conference on Image Processing, Computer Vision, & Pattern Recognition (IPCV’12)*, pp. 208–214, 2012.

[65] C. De Boor, *A practical guide to splines*, vol. 27. Springer-Verlag New York, 1978.

[66] S. Lee, G. Wolberg, K.-Y. Chwa, and S. Y. Shin, “Image metamorphosis with scattered feature constraints,” *Visualization and Computer Graphics, IEEE Transactions on*, vol. 2, no. 4, pp. 337–354, 1996.

[67] S. Lee, G. Wolberg, and S. Y. Shin, "Scattered data interpolation with multilevel b-splines," *Visualization and Computer Graphics, IEEE Transactions on*, vol. 3, no. 3, pp. 228–244, 1997.

[68] E. Catmull and J. Clark, "Recursively generated b-spline surfaces on arbitrary topological meshes," *Computer-aided design*, vol. 10, no. 6, pp. 350–355, 1978.

[69] D. Rueckert, L. I. Sonoda, C. Hayes, D. L. Hill, M. O. Leach, and D. J. Hawkes, "Nonrigid registration using free-form deformations: application to breast mr images," *Medical Imaging, IEEE Transactions on*, vol. 18, no. 8, pp. 712–721, 1999.

[70] K. I. Joy, "Cubic uniform b-spline curve refinement," *On-Line Geometric Modeling Notes*, 2000.

[71] D. R. Forsey and R. H. Bartels, "Hierarchical b-spline refinement," in *ACM SIGGRAPH Computer Graphics*, vol. 22-4, pp. 205–212, ACM, 1988.

[72] G. Wolberg, *Digital image warping*, vol. 10662. IEEE computer society press Los Alamitos, CA, 1990.

[73] M. Z. Brown, D. Burschka, and G. D. Hager, "Advances in computational stereo," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 25, no. 8, pp. 993–1008, 2003.

[74] F. Essannouni, R. O. H. Thami, D. Aboutajdine, and A. Salam, "Simple noncircular correlation method for exhaustive sum square difference matching," *Optical Engineering*, vol. 46, no. 10, pp. 107004–107004, 2007.

[75] G. Blanchet, A. Buades, B. Coll, J.-M. Morel, and B. Rougé, "Fattening free block matching," *Journal of mathematical imaging and vision*, vol. 41, no. 1-2, pp. 109–121, 2011.

[76] Q. Tian and M. N. Huhns, "Algorithms for subpixel registration," *Computer Vision, Graphics, and Image Processing*, vol. 35, no. 2, pp. 220–233, 1986.

[77] E. Z. Psarakis and G. D. Evangelidis, "An enhanced correlation-based method for stereo correspondence with subpixel accuracy," in *Computer Vision, 2005. ICCV 2005. Tenth IEEE International Conference on*, vol. 1, pp. 907–912, IEEE, 2005.

[78] R. Szeliski and D. Scharstein, "Symmetric sub-pixel stereo matching," in *Computer Vision—ECCV 2002*, pp. 525–540, Springer, 2002.

[79] M. Shimizu and M. Okutomi, "Sub-pixel estimation error cancellation on area-based matching," *International Journal of Computer Vision*, vol. 63, no. 3, pp. 207–224, 2005.

[80] N. Sabater, J.-M. Morel, and A. Almansa, "Block matching reaches theoreti-cal accuracy bounds," in *ICIP*, vol. 3-4, pp. 168–172, 2010.

[81] B. Zitova and J. Flusser, "Image registration methods: a survey," *Image and vision computing*, vol. 21, no. 11, pp. 977–1000, 2003.

[82] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, "Speeded-up robust features (surf)," *Computer vision and image understanding*, vol. 110, no. 3, pp. 346–359, 2008.

[83] J. Shawe-Taylor and N. Cristianini, *Kernel methods for pattern analysis.* Cambridge university press, 2004.

[84] C. M. Bishop *et al.*, *Pattern recognition and machine learning*, vol. 1. springer New York, 2006.

[85] N. Shimano, K. Terai, and M. Hironaga, “Recovery of spectral reflectances of objects being imaged by multispectral cameras,” *J. Opt. Soc. Am. A*, vol. 24, no. 10, pp. 3211–3219, 2007.

[86] W. K. Pratt and C. E. Mancill, “Spectral estimation techniques for the spectral calibration of a color image scanner,” *Applied Optics*, vol. 15, no. 1, pp. 73–75, 1976.

[87] P. Urban, M. R. Rosen, and R. S. Berns, “Spectral image reconstruction using an edge preserving spatio-spectral wiener estimation,” *J. Opt. Soc. Am. A*, vol. 26, no. 8, pp. 1865–1875, 2009.

[88] N. Shimano, “Recovery of spectral reflectances of objects being imaged without prior knowledge,” *IEEE Transactions on Image Processing*, vol. 15, no. 7, pp. 1848–1856, 2006.

[89] A. Neumaier, “Solving ill-conditioned and singular linear systems: A tutorial on regularization,” *Siam Review*, vol. 40, no. 3, pp. 636–666, 1998.

[90] Y. Zhao and R. S. Berns, “Image-based spectral reflectance reconstruction using the matrix r method,” *Color Research & Application*, vol. 32, no. 5, pp. 343–351, 2007.

[91] 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,” *J. Opt. Soc. Am. A*, vol. 25, no. 10, pp. 2444–2458, 2008.

[92] J. L. Nieves, E. M. Valero, S. Nascimento, J. Hernández-Andrés, and J. Romero, “Multispectral synthesis of daylight using a commercial digital CCD camera,” *Applied Optics*, vol. 44, no. 27, pp. 5696–5703, 2005.

[93] E. M. Valero, Y. Hu, J. Hernández-Andrés, T. Eckhard, J. L. Nieves, J. Romero, M. Schnitzlein, and D. Nowack, “Comparative performance analysis of spectral estimation algorithms and computational optimization of a multispectral imaging system for print inspection,” *Color Research & Application*, vol. 39, no. 1, pp. 16–27, 2014.

[94] J. Herrera-Ramírez, M. Vilaseca, and J. Pujol, “Portable multispectral imaging system based on light-emitting diodes for spectral recovery from 370 to 1630 nm,” *Applied Optics*, vol. 53, no. 14, pp. 3131–3141, 2014.

[95] M. A. López-Álvarez, J. Hernández-Andrés, and J. Romero, “Developing an optimum computer-designed multispectral system comprising a monochrome ccd camera and a liquid-crystal tunable filter,” *Applied Optics*, vol. 47, no. 24, pp. 4381–4390, 2008.

[96] N. Tsumura, n. H. Sato, T. Hasegawa, H. Haneishi, and Y. Miyake, “Limitation of color samples for spectral estimation from sensor responses in fine art painting,” *Optical review*, vol. 6, no. 1, pp. 57–61, 1999.

[97] P. Stigell, K. Miyata, and M. Hauta-Kasari, “Wiener estimation method in estimating of spectral reflectance from rgb images,” *Pattern Recognition and Image Analysis*, vol. 17, no. 2, pp. 233–242, 2007.

[98] H.-L. Shen and J. H. Xin, "Spectral characterization of a color scanner based on optimized adaptive estimation," *J. Opt. Soc. Am. A*, vol. 23, no. 7, pp. 1566–1569, 2006.

[99] F. H. Imai and R. S. Berns, "Spectral estimation using trichromatic digital cameras," in *Proceedings of the International Symposium on Multispectral Imaging and Color Reproduction for Digital Archives*, vol. 42, 1999.

[100] G. Hong, M. R. Luo, and P. A. Rhodes, "A study of digital camera colorimetric characterization based on polynomial modeling," *Color Research & Application*, vol. 26, no. 1, pp. 76–84, 2001.

[101] H.-L. Shen, H.-J. Wan, and Z.-C. Zhang, "Estimating reflectance from multispectral camera responses based on partial least-squares regression," *Journal of Electronic Imaging*, vol. 19, no. 2, pp. 020501–020501, 2010.

[102] J. Hardeberg, *Acquisition and reproduction of color images: colorimetric and multispectral approaches*. Ph.D. thesis, 2001.

[103] V. Heikkinen, T. Jetsu, J. Parkkinen, M. Hauta-Kasari, T. Jääskeläinen, and S. D. Lee, "Regularized learning framework in the estimation of reflectance spectra from camera responses," *J. Opt. Soc. Am. A*, vol. 24, no. 9, pp. 2673–2683, 2007.

[104] V. Heikkinen, *Kernel methods for estimation and classification of data from spectral imaging*. PhD thesis, University of Eastern Finland, Faculty of Science and Forestry, 2011.

[105] C. Berg, J. Christensen, and P. Ressel, *Harmonic analysis on semigroups theory of positive definite and related functions*. Springer New York, 1984.

[106] B. Schöllkopf and A. Smola, *Learning with kernels*. MIT Press, Cambridge, MA, 2002.

[107] V. Heikkinen, A. Mirhashemi, and J. Alho, "Link functions and matern kernel in the estimation of reflectance spectra from rgb responses," *J. Opt. Soc. Am. A*, vol. 30, pp. 2444–2454, Nov 2013.

[108] W.-F. Zhang, P. Yang, D.-Q. Dai, and A. Nehorai, "Reflectance estimation using local regression methods," in *Advances in Neural Networks–ISNN 2012*, pp. 116–122, Springer, 2012.

[109] G. D. Finlayson and P. Morovic, "Metamer sets," *J. Opt. Soc. Am. A*, vol. 22, no. 5, pp. 810–819, 2005.

[110] C. Li and M. R. Luo, "The estimation of spectral reflectances using the smoothness constraint condition," in *Color and Imaging Conference*, vol. 2001-1, pp. 62–67, Society for Imaging Science and Technology, 2001.

[111] J.-I. Park, M.-H. Lee, M. D. Grossberg, and S. K. Nayar, "Multispectral imaging using multiplexed illumination," in *Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on*, pp. 1–8, IEEE, 2007.

[112] W. Wu, B. Walczak, D. Massart, S. Heuerding, F. Erni, I. Last, and K. Prebble, "Artificial neural networks in classification of nir spectral data: design of the training set," *Chemometrics and intelligent laboratory systems*, vol. 33, no. 1, pp. 35–46, 1996.

[113] V. Cheung and S. Westland, “Methods for optimal color selection,” *Journal of Imaging Science and Technology*, vol. 50, p. 481, 2006.

[114] J. Kang, K. Ryu, and H. Kwon, “Using cluster-based sampling to select initial training set for active learning in text classification,” *Advances in Knowledge Discovery and Data Mining, Lecture Notes in Computer Science*, vol. 3056, pp. 384–388, 2004.

[115] M. Mohammadi, M. Nezamabadi, R. S. Berns, and L. Taplin, “A prototype calibration target for spectral imaging,” in *Tenth Congress of the International Colour Association*, pp. 387–390, 2005.

[116] H.-L. Shen, H. Zhang, J. H. Xin, and S.-J. Shao, “Optimal selection of representative colors for spectral reflectance reconstruction in a multispectral imaging system,” *Applied Optics*, vol. 47, no. 13, pp. 2494–2502, 2008.

[117] R. Hersch and F. Crété, “Improving the yule-nielsen modified spectral neugebauer model by dot surface coverages depending on the ink superposition conditions,” in *Proceedings of SPIE*, vol. 5667, pp. 434–445, 2005.

[118] H.-L. Shen and J. H. Xin, “Spectral characterization of a color scanner by adaptive estimation,” *J. Opt. Soc. Am. A*, vol. 21, no. 7, pp. 1125–1130, 2004.

[119] F. Agahian, S. A. Amirshahi, and S. H. Amirshahi, “Reconstruction of reflectance spectra using weighted principal component analysis,” *Color Research & Application*, vol. 33, no. 5, pp. 360–371, 2008.

[120] N. Eslahi, S. H. Amirshahi, and F. Agahian, “Recovery of spectral data using weighted canonical correlation regression,” *Optical Review*, vol. 16, no. 3, pp. 296–303, 2009.

[121] V. Babaei, S. H. Amirshahi, and F. Agahian, “Using weighted pseudo-inverse method for reconstruction of reflectance spectra and analyzing the dataset in terms of normality,” *Color Research & Application*, vol. 36, no. 4, pp. 295–305, 2011.

[122] M. Brookes, “The matrix reference manual,” *Imperial College London*, 2005.

[123] G. Sharma, W. Wu, and E. N. Dalal, “The ciede2000 color-difference formula: Implementation notes, supplementary test data, and mathematical observations,” *Color Research & Application*, vol. 30, no. 1, pp. 21–30, 2005.