FORESTRY AND
NATURAL SCIENCES

VILLE HEIKKINEN

Kernel methods for estimation and classification of data from spectral imaging

Publications of the University of Eastern Finland
Dissertations in Forestry and Natural Sciences No 31

UNIVERSITY OF
EASTERN FINLAND

VILLE HEIKKINEN

Kernel methods for estimation and classification of data from spectral imaging

Publications of the University of Eastern Finland
Dissertations in Forestry and Natural Sciences
No 31

Academic Dissertation
To be presented by permission of the Faculty of Science and Forestry
for public examination in the Louhela Auditorium in Science Park, Joensuu,
on May 20, 2011, at 12 o'clock noon.

School of Computing

Kopijyvä Oy
Joensuu, 2011
Editors: Prof. Pertti Pasanen, PhD. Sinikka Parkkinen, Prof. Kai Peiponen

Distribution:
University of Eastern Finland Library / Sales of publications
P.O. Box 107, FI-80101 Joensuu, Finland
tel. +358-50-3058396
http://www.uef.fi/kirjasto

ISBN: 978-952-61-0423-2 (printed)
ISSNL: 1798-5668
ISSN: 1798-5668
ISBN: 978-952-61-0424-9 (pdf)
ISSNL: 1798-5668
ISSN: 1798-5676

Author's address:
University of Eastern Finland
School of computing
P.O.Box 111
80101 Joensuu
Finland
email: ville.heikkinen@uef.fi

Supervisors:
Professor Juha Alho, Ph.D.
University of Eastern Finland
School of computing
P.O.Box 111
80101 Joensuu
Finland
email: juha.alho@uef.fi

Professor Markku Hauta-Kasari, Ph.D.
University of Eastern Finland
School of computing
P.O.Box 111
80101 Joensuu
Finland
email: markku.hauta-kasari@uef.fi

Professor Jukka Tuomela, Ph.D.
University of Eastern Finland
Department of Physics and Mathematics
P.O.Box 111
80101 Joensuu
Finland
email: jukka.tuomela@uef.fi

Reviewers:
Associate Professor Javier Hernández-Andrés, Ph.D.
Universidad de Granada
Departamento de Óptica, Facultad de Ciencias
Campus Fuentenueva
18071 Granada
Spain
email: javierha@ugr.es

Markus Koskela, Ph.D.
Aalto University School of Science
Department of Information and Computer Science
P.O. Box 15400
00076 Aalto
Finland
email: markus.koskela@tkk.fi

Opponent:
Professor Erkki Oja, Ph.D.
Aalto University School of Science
Department of Information and Computer Science
P.O. Box 15400
00076 Aalto
Finland
email: erkki.oja@hut.fi

ABSTRACT

In this dissertation I have studied color calibration of a camera, estimation of reflectance spectra, and supervised classification of spectral information, based on reproducing kernel Hilbert space methods (RKHS). A unifying characteristic of our spectral data was that imaging was performed with small number of broad-band spectral response functions. We considered reflectance estimation as a generalized color calibration problem and mainly focused on an empirical regression approach that assumes relatively large ensembles of training data. The connections of several reflectance estimation and color calibration models to more general RKHS models are discussed. Several RKHS models and transformations based on physical a priori knowledge are introduced and evaluated for the reflectance estimation from responses of an ordinary RGB camera. The results suggest that new models lead to better accuracy in reflectance estimation and color calibration than some classical, more widely used models. The data classification is discussed in remote sensing context, where data are simulated to correspond measurements from a multispectral airborne camera. In the classification of birch (Betula pubescens Ehrh., Betula pendula Roth), pine (Pinus sylvestris L.) and spruce (Picea abies (L.) H. Karst.) trees, a Support Vector Machine classifier (SVM) and RKHS feature space mappings were used to validate the performance of several simulated sensor systems. The results indicate a need for careful data pre-processing, a higher number of sensor bands, decrease in the bandwidths or new positioning of the bands in order to improve pixel-based classification accuracy for these tree species.

Universal Decimal Classification: 004.93, 519.6, 535.3, 535.6
PACS Classification: 02.60.-x, 02.60.Ed, 02.70.-c, 07.05.Mh
Keywords: color imaging; machine vision; machine learning; pattern recognition; remote sensing; spectral imaging; supervised learning

Preface

The research for this study was carried out in the Department of Computer Science at the University of Joensuu during 2005-2009 and in the School of Computing at the University of Eastern Finland during 2010-2011. This study was partly supported by the Academy of Finland (”High Resolution Remote Sensing Potential for Measure Single Trees and Site Quality”, project no. 123193.). This support is gratefully acknowledged.

I want to thank prof. Jussi Parkkinen and prof. Timo Jääskeläinen for providing me a position in the color research group. During these years, I have had the possibility to work in several projects dealing with spectral color science and remote sensing. I want to express my gratitude to the department staff and all members (past and present) of the color research group in Joensuu with whom I have worked during these years.

Especially, I wish to thank Dr. Tuija Jetsu, Dr. Ilkka Korpela and prof. Timo Tokola for collaborations in several projects. I also want thank Dr. Reiner Lenz for collaboration and being a host during my visit at the Linköping University, Sweden. I am indebted for my supevisors, prof. Juha Alho, prof. Markku Hauta-Kasari and prof. Jukka Tuomela and for the reviewers Dr. Javier Hernández-Andrés and Dr. Markus Koskela for their valuable comments. Especially, I am deeply grateful for Juha for several discussions and his invaluable guidance in scientific writing.

I thank my parents Irma and Esko and my sister Hanna for their support and encouragement.

Finally, I thank Hanna for her endless support and love.

Joensuu, April 2011,
Ville Heikkinen

LIST OF PUBLICATIONS

The core of this dissertation is based on the following refereed publications:

P1 Jetsu, T., Heikkinen, V., Parkkinen, J., Hauta-Kasari, M., Martinkauppi, B., Lee, S.D., Ok, H.W., Kim, C.Y., “Color Calibration of Digital Camera Using Polynomial Transformation”. In *Proceedings of CGIV2006 - 3rd European Conference on Colour in Graphics, Imaging, and Vision*, Leeds, UK, June 19-22, 163–166, 2006.

P2 Heikkinen, V., Jetsu, T., Parkkinen, J., Hauta-Kasari, M., Jääskeläinen, T., Lee, S.D., “Regularized learning framework in estimation of reflectance spectra from camera responses,” *J. Opt. Soc. Am. A* **24**(9), 2673-2683 (2007).

P3 Heikkinen, V., Lenz, R., Jetsu, T., Parkkinen, J., Hauta-Kasari, M., Jääskeläinen, T. “Evaluation and unification of some methods for estimating reflectance spectra from RGB images,” *J. Opt. Soc. Am. A* **25**(10), 2444-2458 (2008).

P4 Heikkinen, V., Tokola, T., Parkkinen, J., Korpela, I., Jääskeläinen, T. “Simulated Multispectral Imagery for Tree Species Classification Using Support Vector Machines,” *IEEE Trans. Geosci. Remote Sens.* **48**(3), 1355–1364 (2010).

[P1] is peer reviewed conference article, [P2]–[P4] are peer reviewed journal articles.

Throughout the core text, these papers will be referred to by [P1], [P2], [P3] and [P4].

AUTHOR'S CONTRIBUTION

This dissertation consists of a lengthy core text and four publications. The publications selected for this dissertation are original research papers on reflectance estimation, color calibration and tree species classification using spectral data.

The ideas for papers [P1], [P2] and [P4] were mainly proposed by the author in collaboration with the co-authors. The ideas in [P3] originated from discussions between the author and co-author Dr. R. Lenz. The author has carried out numerical computations in [P1]. In [P2], [P3] and [P4] the author has carried out all numerical computations, the selection of data and selection of optimization methods. The hyperspectral and characteristic camera data which were used in [P1]-[P4] were measured by the co-authors.

The author has written parts of [P1] as a co-author and has written the papers [P2], [P3] and [P4]. In [P3], the contribution of co-author Dr. R. Lenz has been especially significant.

The author has written the core text, which is based on extensive discussions with the supervisors prof. J. Alho, prof. M. Hauta-Kasari, and prof. J. Tuomela. The core text in this dissertation is in the nature of long introduction and is partly written to be a foundation for future work for several topics in color science research.

CONTENT OF THE DISSERTATION

The core text in this dissertation summarizes the four publications (see Fig. 1), and explains the physical and mathematical background of the models more extensively than could have been done in the four publications. The text is multidisciplinary and its considerable length has been motivated by the fact that in color science, there appears not to exist a unified treatment of the types of methods that have been used in [P1]-[P3].

Chapter 1 gives an introduction to the dissertation. A general machine vision reflection model is presented in chapter 2 which links the used reflection model to the standardized physical characteristics of surfaces. Some essential theory of reproducing kernel Hilbert spaces is presented in chapter 3. It also introduces feature mappings used for classification and estimation in chapters 4 and 5. Chapters 4 and 5 also discuss and summarize the experimental results. Chapter 6 gives conclusions and discusses future work.

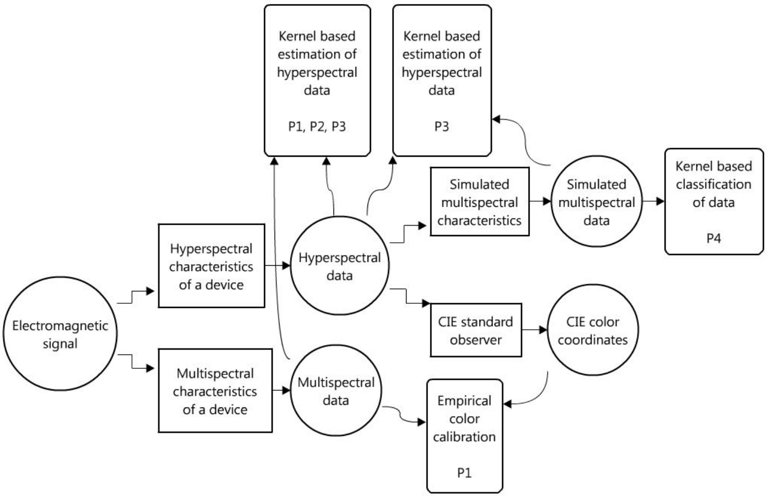

Figure 1: Content of the publications [P1], [P2], [P3] and [P4]. The arrows explains the type of data (real measurements and simulated camera responses), which were used in the publications.

NOTATION

Table 1: Symbols for radiometric quantities and for concepts in color science.

Symbol	Meaning
λ	wavelength
Λ	wavelength domain
ω	solid angle
θ	zenith angle
ϕ	azimuth angle
Φ	spectral radiant flux (Table 2: Feature map)
E	spectral irradiance
L	spectral radiance
L_I, L_R	incident and reflected radiance, respectively
f_r	Bidirectional Spectral Reflectance Distribution Function
r	spectral factor of f_r (reflectance spectrum)
g	geometrical factor of f_r
l	spectral factor of spectral radiant power (flux)
$L_{I,2}$	geometrical factor of spectral radiant power (flux)
r_f	reflectance factor
k	number of spectral response functions
s_i	ith spectral response function of sensor
t_g	geometrical factor
t_e	exposure time
ν	transmittance function
Γ	non-linearity of sensor (also Gamma-function)
$\mathbf{x}$	multi- or hyperspectral measurement, $\mathbf{x} \in \mathbb{R}^k$
$\mathbf{q}$	hyperspectral measurement, $\mathbf{q} \in \mathbb{R}^n$
$D65$	CIE standard illuminant
A	CIE standard illuminant
R, G, B	Red-Green-Blue color values
L, M, S	cone sensitivity functions
$\overline{x}, \overline{y}, \overline{z}$	CIE 1931 color matching functions
X, Y, Z	CIEXYZ color coordinates
L^*,a^*,b^*	CIEL*a*b* color coordinates
C^*_{ab}	CIEL*a*b* chroma
h_{ab}	CIEL*a*b* hue-angle
ΔE^*	CIE 1976 color difference
ΔE^*_{00}	CIEDE2000 color difference

Table 2: Mathematical symbols.

Symbol	Meaning
$\mathbb{R}$	the real numbers
$\mathbb{R}^+$	the positive real numbers
$\mathbb{R}^n$	n-dimensional real vector space
H	real vector space
$\mathcal{H}$	Hilbert space
$\mathcal{X}$	input space
l^2	space of square summable series
$L^2(\mathcal{X})$	space of square integrable functions on $\mathcal{X}$
$\mathcal{C}(\mathcal{X})$	space of continuous functions on $\mathcal{X}$
$\langle \cdot, \cdot \rangle$	inner product of $L^2(\mathcal{X})$
$\langle \cdot, \cdot \rangle_{\mathcal{H}}$	inner product of $\mathcal{H}$
Id	identity map $\mathrm{Id}(\mathbf{x}) \to \mathbf{x}$
k	dimension of the input space
m	number of the training samples
$\mathbf{x}_i$	the i^{th} training input in $\mathbb{R}^k$
x_i	the i^{th} coordinate of vector $\mathbf{x}$
$\begin{pmatrix} n \\ k \end{pmatrix}$	binomial coefficient $= n!/(k!(n-k)!)$
$\Phi(\mathbf{x})$	feature mapping $\Phi : \mathbf{x} \to \mathcal{F}$
N	dimension of the finite dimensional feature space
$\kappa(\cdot, \cdot)$	kernel function $\Phi : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$
X	$m \times k$ matrix of the training inputs $\{\mathbf{x}_i\}_{i=1}^m$
$\Phi(X)$, $\Psi(X)$	matrix X, where rows are mapped to feature space
$\mathbf{x}^T$ or X^T	transpose of vector $\mathbf{x}$ or matrix X
$\|\| \cdot \|\|$	L^2 or l^2 norm
$\|\| \cdot \|\|_{\mathcal{H}}$	$\mathcal{H}$ (semi) norm
$\|\| \cdot \|\|_F$	Frobenius norm of a matrix
I (I_n)	identity matrix (of size n)
$\mathbf{1}$	vector of ones
Σ	covariance matrix
$\|\Sigma\|$	determinant of matrix Σ
$\mathrm{Tr}(X)$	trace of matrix, i.e. $\sum_i x_{ii}$
$\mathcal{L}$	loss function
σ^2	regularization parameter in regression
C	penalization parameter in SVM classification
K	kernel matrix
$\mathbf{k}(\mathbf{x})$, $\mathbf{k}_{\mathbf{x}}$	vector of kernel evaluations between $\mathbf{x}$ and elements in the training set

Contents

1 Introduction

Spectral imaging gathers information from across the emitted electromagnetic spectrum of an object in some wavelength range. This technique has become important part of information gathering in several real-life applications in medical imaging [40], [75], [154] remote sensing [19], [141], quality control [36], [60], color engineering [44], [48], [52], [171] etc. Also many applications in cosmetics, paint, textile and plastics industries use the spectral imaging for information gathering. A common goal for all these applications is accurate color measurement and representation or object characterization in radiometric sense. An object characterization can be critical in medical imaging, whereas in industrial applications, a color quality control can be important due to economical reasons. Similarly, accurately measured spectral images can be used for the purposes of electronic commerce and electronic museums.

Spectral imaging over visible (VIS) wavelength range of 380–780 nm is sometimes called spectral color imaging and can be seen as an extension of standard RGB color imaging. Imaging can be also extended to measure invisible radiation in infrared (IR) and ultraviolet (UV) ranges. In this dissertation we utilize spectral data from the visible and near-infrared range of 390–850 nm.

A property which characterizes an imaging sensor spectrally, is the number and location of the individual wavelength bands sensed in the electromagnetic spectrum. Every spectral band of a sensor has a corresponding spectral response function with some shape and bandwidth. The number of the independent bands define the dimensionality of the measurement vectors. Current sensor technology allows the capture of spectral data using hundreds of contiguously positioned narrow spectral bands simultaneously. This so called hyperspectral data provide a possibility to use well-known linear methods to extract representative spectral space features from the data or to estimate surface reflectance information.

For example, it has been shown that several identification tasks based on hyperspectral reflectance data can be carried out accurately and efficiently using linear mappings to a low-dimensional subspaces for each class [19], [76], [120], [122].

Although the most informative spectral data are obtained with hyperspectral systems with several spectral bands, the use of such imaging devices is impractical or expensive in many applications. Usually a high-dimensional hyperspectral data also involve a high level of redundancy, implying an inefficient data management and storage. These issues usually restrict the imaging device to be a multispectral systems with only small amount of broadband characteristics (e.g. [51], [52], [171]). Usually in these applications, surface reflectance is estimated from the measurements by using physical or empirical computational models. It has been suggested that the most widely used multispectral devices, RGB-devices, could be used for the reflectance estimation for the purposes of color engineering [14], [15], [69] - [72], [108], [112], [113], [148], [152], [153].

RGB devices have several shortcomings due to three spectral bands with broad bandwidth characteristics. Without any a priori information, the RGB imaging is considered as an inadequate technique for measuring reflectance information and rendering between different illumination or observer conditions. However, due to the rapid development of color cameras, the RGB technique has also several benefits which motivate its use for reflectance estimation purposes. The acquisition system is capable for imaging moving objects, it is practical, inexpensive and it can produce images with large spatial resolution. It is also evident, that currently almost all digital video cameras have three-band spectral characteristics.

Another class of imaging systems which try to maximize the use of small amount spectral bands are multispectral airborne cameras which were originally designed for photogrammetric applications [39]. Photogrammetric multispectral sensors can provide additional benefits over hyperspectral sensors with small spatial resolution. In this context it is assumed that it is possible to construct all-purpose sensor with small number of spectral response functions, which would allow adequate information extraction from ground objects, both radiometrically and geometrically. In this context, an optimization of sensor spectral bands would increase the applicability of these imaging systems.

This dissertation concentrates on estimation of reflectance spectra, device color calibration and classification of multispectral data:

- In the reflectance estimation, we are considering the reflectance measurements obtained via hyperspectral camera as the objects of interest. Using a priori knowledge, the reflectance is estimated mainly by using empirical models and measurements via multispectral RGB cameras. Here, the reflectance estimation task is seen as a generalized color calibration. Estimation process is also discussed using different levels of available knowledge from the measurement conditions and devices.

- Supervised classification of multispectral data is formulated for the purposes of remote sensing, where the objects of interest are the spectral signatures of the three main tree species in Finland. Measurements are simulated to correspond to a photogrammetric airborne digital sensor with four spectral bands. The classification performance is studied for these simulated measurements, where an optimization of the spectral bands is also considered.

In these tasks, we concentrate on using and estimating pointwise (pixelwise) spectral information from the digital image data.

When the imaging device has a low number of spectral bands, the available data already reside in a fixed low-dimensional subspace defined by spectral response functions of system. Consequently, efficient linear feature extraction might be disrupted due to the low information content of the measured data. In this study, the modeling problems due to low information content of the multispectral measurements is compensated for by means of various data pre-processing methods and non-linear feature space mappings through (conditionally, strictly) positive definite kernel functions. Kernel techniques do not compensate for the lack of information content, but they introduce the tools to model complex data structures nonlinearly. The features derived from the kernel function give representations for the data in high-dimensional feature spaces, where the classification and estimation problems are easier to solve [50], [143], [151], [155], [163]. Especially, the kernels used in this dissertation induce a function space, called reproducing kernel Hilbert space (RKHS). The theory of RKHSs was developed by Aronszajn [5] in 1950s, and it has gained popularity in the field of machine learning due to the algorithmic developments. The generalization properties of the RKHS models can be controlled via regularization functionals, which correspond to the norm or seminorm of the induced function space.

In the field of spectral imaging and color engineering, many reflectance estimation, color calibration and classification models have been previously introduced, without referring to common context of RKHSs and regularized learning. However, in the field of statistics and machine learning, it is known that the RKHS theory unifies the regression estimation and classification when using parametric polynomial expansions and some non-parametric models such as regularization networks, radial basis function networks, ridge regression and smoothing splines [33], [41], [127], [128], [167].

This dissertation gives an overview for the physical foundations of spectral imaging and its implications to reflectance estimation and data classification. We also present the foundations of RKHSs, so that reflectance estimation and color calibration models can be discussed in a unified framework. We evaluate some widely used techniques for the reflectance estimation by using different sources of a priori information and present new approaches using the kernel based learning. In data classification, we concentrate on simulated data, based on a reflection model and evaluate the classification performance by using widely used combination of Support Vector Machine (SVM) and RKHS kernels [151], [163].

2 Physical model for sensation of electromagnetic signal

In this section we discuss on the physical sensation of electromagnetic signal reflecting from an object. Especially, we explicitly define object reflectance as a measurable physical quantity. We mainly focus on the sensitive wavelength range of human vision system, which is approximately 380–780 nm and is called as *visible range* (VIS). However, in the case of artificial device, like digital camera, the sensation is formed using photosensitive sensor chip and the corresponding wavelength range can be extended to *ultraviolet* (UV) and *infrared* (IR) wavelength ranges.

We formulate sensation process for artificial observer, but it can be assumed that under certain conditions, also biological vision systems follow same approach (e.g. [73], [88], [168]). We assume that the sensation process is contributed by four different factors:

1. Illumination and observation geometry.
2. Characteristic reflectance properties of object.
3. Spectral power distribution of illumination.
4. Characteristic sensor properties of the observer.

In addition, a sensation may have significant effects from the medium where the object and the observer are situated, e.g. the atmosphere in remote sensing or water in underwater imaging.

2.1 RADIOMETRIC QUANTITIES AND LIGHT REFLECTION MODEL

In the following we define basic radiometric quantities associated with a light beam by following [88], [89]. We use λ to denote wavelength in nanometers ([nm]) and A to denote area ([m^2]) of a surface. The radiant energy flow per unit time in wavelength λ, through a point $\mathbf{o}$ in surface in a direction (θ, ϕ) is *spectral radiant flux* Φ, i.e.

$$\Phi(\lambda, \theta, \phi, \mathbf{o}) \quad [\mathrm{W\,nm^{-1}}] \tag{2.1}$$

The *spectral irradiance* E is the incident radiant flux (at a point on the surface) per unit area and per unit wavelength, i.e.

$$E(\lambda, \theta, \phi, \mathbf{o}) = \frac{d^2\Phi(\lambda, \theta, \phi, \mathbf{o})}{dAd\lambda} \quad [\mathrm{W\,m^{-2}\,nm^{-1}}]. \tag{2.2}$$

The *solid angle* with unit *steradian* ([sr]), is defined as the area of the radial projection A_{proj} of a surface element to the surface of the sphere with radius ϱ. Assuming surface element and a sphere with radius ϱ, the differential solid angle is $d\omega = dA_{\text{proj}}/\varrho^2$. The solid angle of a full sphere is 4π sr, when viewed from a point inside the sphere. The *spectral radiance* L is the radiant flux per solid angle per projected surface area per unit wavelength, i.e.

$$L(\lambda, \theta, \phi, \mathbf{o}) = \frac{d^3\Phi(\lambda, \theta, \phi, \mathbf{o})}{dA\cos\theta d\omega d\lambda} \quad [\mathrm{W\,m^{-2}\,sr^{-1}\,nm^{-1}}], \tag{2.3}$$

where θ is angle between the surface normal and cone of light beams. The spectral radiance is the quotient of radiant flux in a given direction, leaving or arriving at an element of surface (with area dA) at a point, and propagated through a cone of solid angle $d\omega$ in given direction. Incident spectral radiance on a surface interacts with the material, so that it is *absorbed*, *transmitted* and *reflected* according to properties of material. Fig. 2.1 presents propagation of incident and reflected radiance through cones of light beams.

In order to model geometrical reflectance properties of an object the *Bidirectional Spectral Reflectance Distribution Function* - f_r (BSRDF) is defined as a ratio of differentials (omitting point coordinates)

$$f_r(\lambda, \theta_I, \phi_I, \theta_R, \phi_R) = \frac{dL_R(\lambda, \theta_I, \phi_I, \theta_R, \phi_R)}{dE_I(\lambda, \theta_I, \phi_I)} \quad [\mathrm{sr^{-1}}], \tag{2.4}$$

where $dL_R(\lambda, \theta_I, \phi_I, \theta_R, \phi_R)$ is the reflected spectral radiance in the viewing direction and $dE_I(\lambda, \theta_I, \phi_I)$ is the spectral irradiance inci-

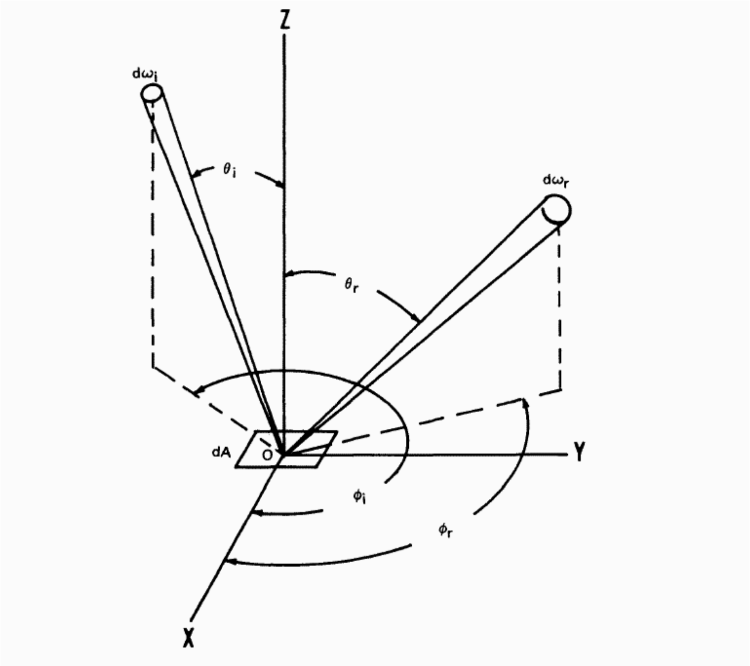

Figure 2.1: Geometry of incident and reflected cones of light beams (Adapted from [117]). Subscripts i and r are used to denote incidence and reflection angles, respectively.

dent on the illumination direction [88], [117]. Here, the *Zenith angles* θ_I and θ_R are defined with respect to the surface normal and the *azimuth angles* ϕ_I and ϕ_R with respect to chosen direction in the surface plane. Quantities dL_R and dE_I depend on the differential incident flux from direction (θ_I, ϕ_I) within differential solid angle and over the differential area element dA in the surface (Fig. 2.1). The differential solid angle can be written as $d\omega = \sin\theta d\theta d\phi$.

Incident spectral irradiance on the surface can be written as

$$dE_I(\lambda, \theta_I, \phi_I) = L_I(\lambda, \theta_I, \phi_I)\cos\theta_I d\omega_I, \tag{2.5}$$

where $L_I(\lambda, \theta_I, \phi_I)$ is the incident spectral radiance and $d\omega_I$ is the differential solid angle in the direction (θ_I, ϕ_I) (Fig. 2.1). For the surface reflected spectral radiance L_R, it can be written

$$L_R = \int dL_R = \int f_r dE_I = \int_{\omega_I} f_r L_I \cos\theta_I d\omega_I. \qquad (2.6)$$

The *at-sensor spectral irradiance* (image spectral irradiance) E_P is obtained by integrating (2.6) over the solid angle occupied by the sensor's entrance aperture

$$\begin{aligned} E_P &= \frac{dA}{dA_{\text{image}}} \int_{\omega_R} L_R \cos\theta_R d\omega_R \\ &= \frac{dA}{dA_{\text{image}}} \int_{\omega_R} \int_{\omega_I} f_r L_I \cos\theta_I d\omega_I \cos\theta_R d\omega_R, \qquad (2.7) \end{aligned}$$

where dA is the infinitesimal area in the object surface and dA_{image} is the area of image patch, which contains the reflected rays from the area dA [64].

2.2 SPECTRAL REFLECTANCE FACTORS

In terms of available radiometric information, usually the most interesting property of the object is a spectral reflectance image, where every spatial location in the image has information about surface reflectance. The information from surface reflectance allows us to investigate the inherent object properties, because it is characteristic property of the object surface and is independent (in simplified, ideal case) of the illumination conditions. Because of this, reflectance information is also highly useful in many spectral based pattern recognition applications. For example, in remote sensing via airborne and satellite imaging, the access to reflectance information is expected to reduce the need of expensive *in situ* gathering of at-target field data between flight campaigns and thus improve the cost efficiency [19], [55], [83]– [86], [141]. In general, the comparison between remotely sensed radiance data from different flight campaigns is complicated due to changes in atmospheric and illumination conditions.

The BSRDF f_r formulated above is only a conceptual quantity and for the measurement purposes, a concept of spectral reflectance factor is needed. A *spectral reflectance factor* r_f is defined as the ratio of the radiant flux at a given wavelength actually reflected by a sample surface to that which would be reflected into the same reflected beam geometry by an *ideal perfectly diffuse (ideal Lambertian) standard surface* irradiated in exactly the same way as the sample [117]. It can be shown that $f_r = 1/\pi$ for the ideal perfect diffuse surface [64], [88]. The reflectance factor (biconical) r_f is defined as

$$\begin{aligned} r_f(\omega_I, \omega_R, \lambda) &= \frac{d\Phi}{d\Phi_{\text{ideal diffuser}}} \\ = & \frac{\int_{\omega_I} \int_{\omega_R} f_r(\theta_I, \phi_I, \theta_R, \phi_R, \lambda) L_I(\lambda, \theta_I, \phi_I) \cos\theta_I d\omega_I \cos\theta_R d\omega_R}{\int_{\omega_I} \int_{\omega_R} \frac{1}{\pi} L_I(\lambda, \theta_I, \phi_I) \cos\theta_I d\omega_I \cos\theta_R d\omega_R}, \end{aligned} \tag{2.8}$$

where $d\omega_I = \sin\theta_I d\theta_I d\phi_I$ and $d\omega_R = \sin\theta_R d\theta_R d\phi_R$ correspond to illumination and viewing apertures, respectively.

2.3 SIMPLIFIED REFLECTION MODELS

For some surfaces and illumination conditions, simplifications of the reflectance factor (2.8) and the at-sensor radiance (2.7) can be derived. In the following we exclude several causes of light, such as diffraction, fluorescence, interference, polarization and refraction [89], [88], pp. 147–150. As an example, usually in a simplified reflection model, a single light source is assumed and the incident spectral radiance on a surface has form

$$L_I(\lambda, \theta_I, \phi_I) = L_{I,1}(\lambda) L_{I,2}(\theta_I, \phi_I), \tag{2.9}$$

which separates geometrical and spectral factors [89]. Similarly, the BSRDF is assumed to have separation to geometrical and spectral factor ([89], [117], p. 31)

$$f_r(\lambda, \theta_I, \phi_I, \theta_R, \phi_R) = r(\lambda) g(\theta_I, \phi_I, \theta_R, \phi_R). \tag{2.10}$$

In this case, the at-sensor radiance signal (2.7) is written as

$$E_P \quad = \quad t_g L_{I,1}(\lambda) r(\lambda) \tag{2.11}$$

where

$$t_g := \frac{dA}{dA_{\text{image}}} \int_{\omega_I} \int_{\omega_R} g(\theta_I, \phi_I, \theta_R, \phi_R) L_{I,2}(\theta_I, \phi_I) \cos\theta_I d\omega_I \cos\theta_R d\omega_R. \tag{2.12}$$

The reflectance factor (2.8) is written as

$$r_f(\omega_I, \omega_R, \lambda) = a_g(\omega_I, \omega_R) r(\lambda), \tag{2.13}$$

where

$$a_g(\omega_I, \omega_R) = \frac{\int_{\omega_I} \int_{\omega_R} g(\theta_I, \phi_I, \theta_R, \phi_R) L_{I,2}(\theta_I, \phi_I) \cos\theta_I d\omega_I \cos\theta_R d\omega_R}{\frac{1}{\pi} \int_{\omega_I} \int_{\omega_R} L_{I,2}(\theta_I, \phi_I) \cos\theta_I d\omega_I \cos\theta_R d\omega_R} \tag{2.14}$$

Simplifications of the terms a_g and t_g are discussed in [64], [89] and [88]. The model (2.10) can be seen as a simplification of Shafer's *dichromatic model* [142] or *neutral-interface-reflection model* [89], which include a term for *interface reflection*.

In the most simplest model, in *Lambertian surface model*, the function $g(\theta_I, \phi_I, \theta_R, \phi_R)$ in (2.10) equals to $1/\pi$, i.e. the BSRDF is written as

$$f_r(\lambda, \theta_I, \phi_I, \theta_R, \phi_R) = \frac{1}{\pi} r(\lambda). \tag{2.15}$$

From this it follows that reflected spectral radiance (2.6) for all view angles (θ_R, ϕ_R) is the same and the reflectance factor is written as

$$r_f(\omega_I, \omega_R, \lambda) = r(\lambda). \tag{2.16}$$

In this case the values of reflectance factor are in region $[0, 1]$, because the total amount of reflected radiance (integrated over the hemisphere above the surface) cannot exceed the amount of the incident irradiance in the surface.

In the following we denote $l(\lambda) := L_{I,1}(\lambda)$ in (2.11), and call function $l : \Lambda \to \mathbb{R}^+$ as a *(relative) spectral power distribution of light source* and function $r : \Lambda \to \mathbb{R}^+$ as a *reflectance spectrum (or factor)* of an object. In this dissertation we have usually assumed that surfaces are Lambertian (e.g. [P3], [P4]). For a Lambertian surface, the value $r(\lambda) \in [0, 1]$ can be seen as a probability for the reflection of incoming photon of wavelength λ [96]. In case the Lambertian assumption is not reasonable, computational models can be constructed for some specific, fixed viewing geometry. Alternatively, a more general reflection model as presented above, may provide reasonable approximation.

2.4 SENSOR

Access to radiometric measurement information is done via some sensor system, where the signal is converted to electronic in photosensitive sensor chip (e.g. CCD, CMOS [88]) and quantized to digital signal.

2.4.1 Properties of a sensor

Let $\Lambda = [\lambda_1, \lambda_2]$ be a fixed interval of the positive real axis. By using a fixed geometry, the interaction of a reflected electromagnetic signal (2.11) with a k-band sensor system, can be modeled as

$$x_i = \Gamma_i \left(t_g t_e \int_\Lambda l(\lambda) r(\lambda) s_i(\lambda) d\lambda \right), \quad i = 1, \ldots, k \tag{2.17}$$

where $s_i : \Lambda \to [0, 1]$ is the i^{th} *spectral response function* (responsivity) of the sensor. Value $s(\lambda)$ defines a probability, that a photon of wavelength λ will generate an output signal in the sensor [96]. The domain of integration can be written as $\Lambda = \bigcup_{i=1}^{k} \Lambda_i$, where Λ_i correspond to support of the responsivity s_i. The scalar t_e is related to exposure time and the function Γ_i collects non-linearity of the system. The function Γ_i is usually modelled as

$$\Gamma_i(x) = (x/a_i)^{\gamma_i} + b_i, \tag{2.18}$$

where $x \in [0, 1]$, $\gamma_i > 1$ and b_i is bias due to electrical current in the device known as "black current" [88].

In the model (2.17) we assume that the system $\{s_i\}_{i=1}^{k}$ includes the combined effects from quantum efficiency of the sensor s (spectral sensitivity), from the transmittance function ν_o of the optics and from the transmittance functions of the filters ν_i, i.e $s_i = s\nu_o\nu_i$. In the model above, the sensation is free of noise and it is assumed that the Γ is the only source for non-linearity. Sensor *quantizes* analog signals $\{x_i\}_{i=1}^{k}$ to digital usually using 8–16 bits. However, the final quantization level for data might depend on its representation. As an example for images in JPEG (Joint Photographic Experts Group) format, the quantization level may be 8 bits.

Ordinary RGB devices have three spectral response functions and the triplets $\{x_i\}_{i=1}^{3}$ are called as *RGB-values*. For some devices (typically for RGB devices), imaging is done via a spatial wavelength filter mosaic (color filter array), and therefore a response (2.17) for a pixel position is obtained only for some channel s_i and response values for other $k-1$ channels in this pixel position are obtained from spatial neighborhood via some interpolation technique [88]. In spite of this, the model (2.17) (with $\Gamma = \mathrm{Id}$) has been successfully used for computational models using data from RGB- and monochrome devices and interference-, absorption- and Liquid Crystal Tunable filters [14], [48], [69]– [72], [112], [113].

Sensors sample the electromagnetic spectrum over a range of wavelengths supported by the spectral response functions. The Full Width at Half Maximum (FWHM) is used to define the *spectral bandwidth* of a spectral response function. The FWHM is defined as the distance between points on the spectral curve at which the function reaches half of its maximum value. A narrower spectral bandwidth does improve the resolution of closely spaced spectral peaks, but it also decreases the signal-to-noise ratio of a sensor.

In the field of remote sensing, the labeling of the sensor system is based on the properties of the sensor system. A *monochromatic* spectral imaging system has only one spectral band in the region Λ. System is called *multispectral* system, if it has several, narrow and discretely located wavelength bands in the region Λ. System is called *hyperspectral* system, if it has several, narrow wavelength bands, located over a contiguous spectral range in the region Λ. It is usually the case for the hyperspectral devices, that the bandwidths are narrower when compared to the multispectral devices in the

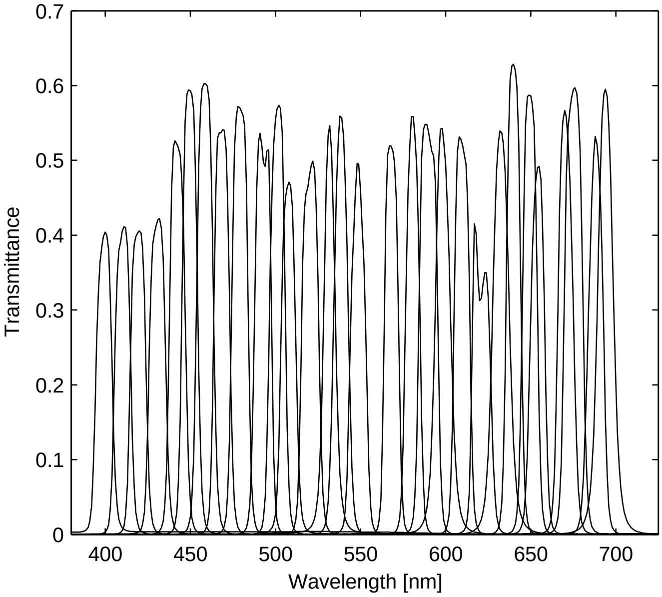

Figure 2.2: Spectral transmittance functions $\{\nu_i\}_{i=1}^{30}$ *of a 30-band interference filter-wheel system in visible wavelength range. The range correspond to 0-70% transmittance.*

same wavelength range. Usually the RGB sensor systems tailored for color photography are considered as a separate system class, because they have broadband characteristics, which cover the whole VIS range. In this dissertation, the class of multispectral devices is extended to consist of systems with narrow- or broad wavelength bands in Λ. In the case of publications [P1],[P2] and [P3], the multispectral systems in interest have three- to six-bands and cover the whole VIS range. In [P4], a four band multispectral system with discretely located bands is considered.

An example of real 30-band hyperspectral system with narrowband characteristics, is depicted in Fig. 2.2. The filters in this system have a bandwidth of 10 nm with overlapping supports and they are almost "regularly" positioned over the VIS range. As a comparison, Fig. 2.3 represents the estimated spectral response functions of an

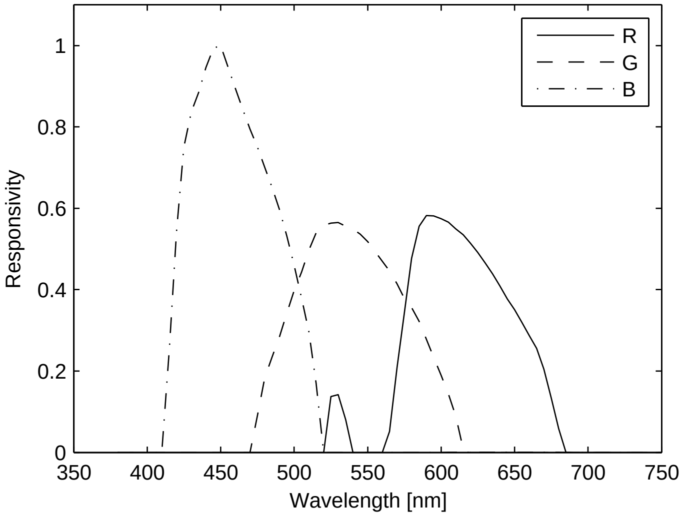

Figure 2.3: Estimated spectral responsivities $\{s_i\}_{i=1}^3$ *of a Nikon D90 RGB-camera. Responsivities* $s_i = s v_o v_i$ *combine all the elements in the optical path.*

RGB camera, which has more broader spectral characteristics than the hyperspectral system.

Spatially the sensor is discrete grid of elements, pixels, which have fixed size in the image space and varying size in the object space. The spatial area covered by one sensor element in the object space depends on the focal length and distance to the object [141], p. 21. Large amount of pixels in image sensor are useful and provide a possibility for large spatial resolution. The spatial properties of the measurement device are usually characterized by the *Point Spread Function*, which defines the response of an imaging system to a point object [88], p. 234. This function covers all the spatial convolution effects due to optics, image motion, detector and electronics [141]. In this dissertation we don't consider the effects from the Point Spread Function, but assume that it is spatially uniform and ideal, infinitely narrow pulse.

The spectral data that are mainly used in this dissertation ([P1]–[P3]) were measured in laboratory conditions using spectrally homogenous and flat color targets. It can be assumed that the pixel size in object surface was fixed and that the effects of spectral response functions and spectral power distributions of light sources were spatially uniform. Exception to these are the data, that were used for the classification of tree species in [P4]. Details of these data are discussed in chapter 5 and [P4].

2.4.2 Reflectance via narrow band sensors

As discussed above, in many practical applications we are interested on the extraction of surface reflectance information from the measurements (2.17). The goal of estimating reflectance information is often not to obtain an explicit representation of the function r but to estimate some of its values. Many computational models assume that the measurements x_i can be written as a product of reflectance r and illumination irradiance l. An example of a class of algorithms using the product form are the *color constancy algorithms* [32], [37], [68]. According to model (2.17), the product form is valid for infinitely narrow-band spectral response functions or light source (i.e. *Dirac delta function* [81], p. 124). In reality, by assuming models (2.11) and (2.17) and narrow bandwidth characteristics $\{s_i\}_{i=1}^{k}$ respectively "centered" at $\{\lambda_i\}_{i=1}^{k}$, we can estimate the values of reflectance factor

$$a_g r(\lambda_i) \approx a_g \frac{\int_\Lambda l(\lambda) r(\lambda) s_i(\lambda) d\lambda}{\frac{1}{\pi}\int_\Lambda l(\lambda) s_i(\lambda) d\lambda}, \quad i = 1, \ldots, k. \tag{2.19}$$

The divisor correspond to the perfect diffuser and a_g is the geometrical factor (2.14). In practice, the measured radiance values $\{x_i\}_{i=1}^{k}$ from object are divided by the measurements from a calibrated diffuser.

In a measurement, it is required that an adequate amount of radiant power exists in the wavelength area of interest. Figure 2.4 depicts two standard illuminants, which are approximated in the laboratory conditions by using some physical light sources. CIE (Commission Internationale de l'Eclairage, International Commission on Illumination), illuminant D65 correspond to a Planckian radiator with 6500 K correlated color temperature and simulates the average sunny mid-day daylight in Europe [67], [88]. CIE illuminant A has the same relative spectral power distribution as a Planckian radiator with 2856 K absolute temperature. It is intended to represent typical, domestic, tungsten-filament lighting [67], [88]. The third curve represents the spectral power distribution of a real fluorescent light source.

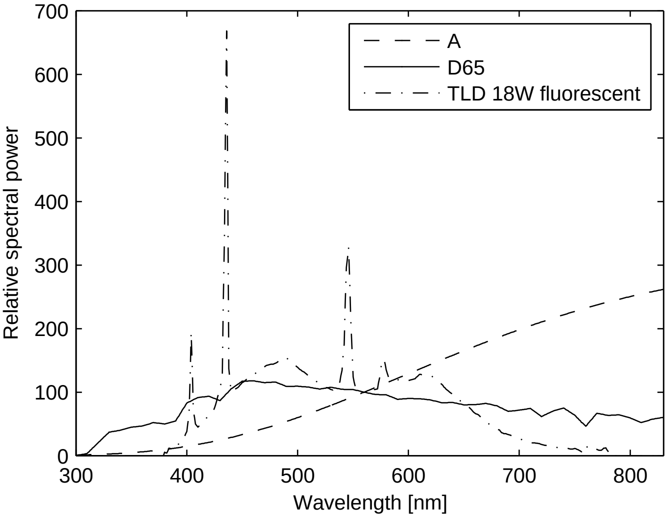

Figure 2.4: The spectral power distributions of CIE standard illuminants A and D65 and fluorescent source TLD 18W.

2.4.3 Practical issues in spectral imaging

Spectral color imaging in the field of artwork and cultural heritage imaging give example of an application where the spectral imaging has efficient use. This approach has been used e.g. in the National Gallery, London, National Gallery of Art, Washington DC, in the Museum of Modern Art, New York, in the Uffizi Gallery, Florence and in the National Museum of Japanese History, Sakura, Chiba, Japan [14], [15], [35], [47], [69]– [72], [109], [157]. In optimal case, a hyperspectral device with high number of bands can be used to derive the reflectance properties of the object. The analysis of measured data gives information about the materials and methods which were used during the period of artist. Characteristic reflectance properties allow also the accurate control for the color rendering in different illumination conditions, detection of counterfeits, cleaning of the artworks and the restoration of the degraded colors. The reflectance data is also highly useful in the construction of databases for digital museums.

Depending on the spectral imaging device, an output of one measurement is either a point measurement from some spatial location, a line scan, or a two-dimensional image. In the line scanning technique (push-broom imaging), a spectral image can be constructed from a sequence of scans. In a band-sequential technique, a set of filters and a monochromatic frame camera are used to produce multiple measurements from an object. In this approach, a spectral image is constructed from a sequence of images corresponding to different filters. A similar, but more novel approach is to consider a light source as a filter [28], [52], [54], [69], [121].

Device manufacturers need to compromise between number of pixels and number of spectral bands when constructing line scanning and imaging devices. Therefore, some application specific constraints guide the choice of a suitable imaging device. In many cases, the line scanning approach provides the best compromise between spectral and spatial accuracy [39], [60].

The practical problems associated to the use of line scans or multiple filters depends on object properties and on the exposure time for single measurement. In practice, the line scans and images corresponding to filters are measured separately. The efficient use of these measurements requires a spatial registration in a post-processing phase. It is possible that object is moving and short exposure time is required. Examples of applications where the object is not stationary are the medical imaging of the human retina (e.g. [40], [75], [154]) and quality control applications (e.g. [36], [60]). Usually fast hyperspectral imaging is possible only with relatively high light power levels [87], pp. 38–40. The line scanning approach is usually more efficient than band-sequential technique in this sense [60]. However, it is possible that the exposure time (or light power level) limits the measurement system to be a ”snapshot” device, such as RGB camera. Albeit, in the case of ordinary RGB camera, the measurements are limited to the VIS range.

In order to address the practical issues (imaging speed, light source restrictions and poor mobility) related to hyperspectral imaging systems for artworks, it has been suggested to construct a system consisting of RGB device and a calibration chart with a known reflectance information (e.g [14], [15], [69]– [72]). This system would simplify the spectral imaging process and decrease investments. A detailed discussion of this approach is presented in chapter 4.

A summary of measurement type, number of pixels and bands, bandwidths (FWHM) and measurement speed (output/second) for different spectral imaging systems are presented in Table 2.1.

Table 2.1: Suggestive properties of measurement and output data for different spectral imaging systems (corresponding to single measurement in VIS area).

Device	Output	No. of pixels	Bands	FWHM [nm]	Outputs/s
Mono	Image	$> 5 \cdot 10^3 \times 5 \cdot 10^3$	1	-	$\leq 10^5$
RGB	Image	$5 \cdot 10^3 \times 5 \cdot 10^3$	3	100	$\leq 10^3$
6-band ([171])	Image	$2 \cdot 10^3 \times 2 \cdot 10^3$	6	50	$\leq 10^2$
Hyperspec.	Image	$10^3 \times 10^3$	10-60	$5 - 10$	≤ 1
4-band ([39])	Line	1×10^4	4	50	$\leq 10^3$
Hyperspec.	Line	1×10^3	100-1000	$3 - 0.3$	$\leq 10^2$

2.5 COLOR AND RESPONSE SPACES

Spectral reflectance information (as a function of wavelength) in visual wavelength range is highly useful in color engineering and colorimetry. This information can be used to represent image data in any color- and device response space or simulate different illumination conditions for a scene. In color processing chains, several devices are used to measure, represent or reproduce information [171]. A communication between devices is normally dealt with the use of conversions between light source and device dependent color spaces. Some color space conversions and light source simulations are difficult to perform with standard trichromatic color data. In many cases, the access to light source independent reflectance data would simplify the device communication significantly.

Reflectance information can be used to produce color responses in device independent color spaces. Currently, CIE spaces are standard device independent color spaces. These spaces are three-dimensional and based on characteristic signal processing in the human visual system. For the color sensation, human retina uses only three types of cone cells. These cells are named L (Long),M (Middle) and S (Short) cells according to the location of sensitivity maxima of 565 nm, 545 nm and 440 nm, correspondingly [168]. Wavelength support of the cells are: L-cells 380-700 nm, M-cells 380-650 nm and S-cells 380-550 nm. It is assumed that the responses of

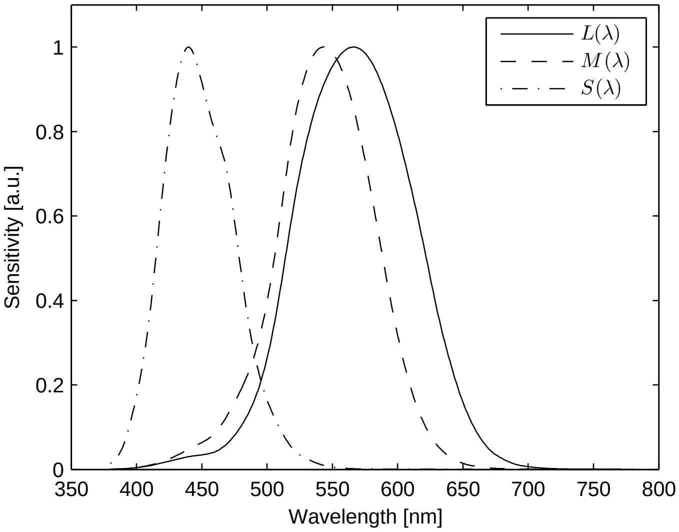

Figure 2.5: Estimate of spectral sensitivity of cone cells (Smith-Pokorny).

the cone cells can be modeled according to (2.17), where the spectral response functions s_i are replaced with the cone sensitivity functions. Smith-Pokorny estimates of the sensitivities of the L,M and S cone cells are depicted in Fig. 2.5 [150].

2.5.1 CIEXYZ-space

The CIE color spaces (and the estimated cone cell sensitivities in Fig. 2.5) are based on the so called *standard observer*, which is an average human observer derived from empirical psychophysical data induced by a group of people in the 1920s and 1930s [88]. At that time, a direct measurement of cone cell sensitivities was impossible.

The CIE tristimulus values X, Y, Z are defined for the spectral reflectance factor r and for the spectral power distribution of light source (or illuminant) l as follows:

$$(X,Y,Z) = \left(t \int_{\Lambda} r_l(\lambda)\overline{x}(\lambda)d\lambda, t \int_{\Lambda} r_l(\lambda)\overline{y}(\lambda)d\lambda, t \int_{\Lambda} r_l(\lambda)\overline{z}(\lambda)d\lambda \right), \tag{2.20}$$

where $r_l = rl$, $\Lambda = [380, 780]$ nm, $t = 100/\int_{\Lambda} l(\lambda)\overline{y}(\lambda)$ and $\overline{x}(\lambda)$, $\overline{y}(\lambda)$ and $\overline{z}(\lambda)$ are the CIE *1931 color matching functions* of the standard observer [13], [88], [170]. Normalization term $100/\int l\overline{y}$ leads to value $Y = 100$, when $r(\lambda) \equiv 1$ for all $\lambda \in \Lambda$. The 1931 color matching functions are represented in Fig. 2.6.

The measurements depend on the view- and light source angles, and therefore the CIE has defined four standard view- and light source conditions for laboratory spectral color measurements [61], [88], [170]. Two of these conditions define the light source and viewing angles as follows: $45°/0°$, $0°/45°$. For example, in $45°/0°$ condition, a surface is illuminated with one or more beams, whose axes are at an angle of $45 \pm 5°$ from the surface normal. The angle between the view direction and the surface normal should not exceed $5°$.

There are some limitations in the CIEXYZ space, which need to be considered, when the tristimulus values are interpreted. First, the tristimulus values are based on standard observer, although it

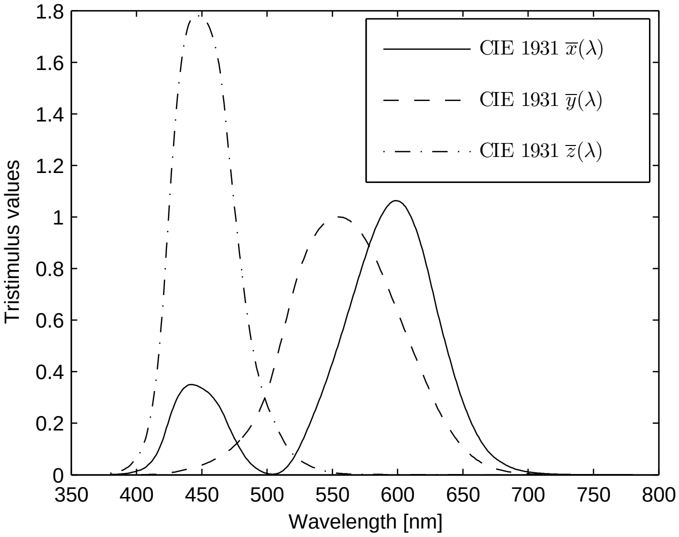

Figure 2.6: CIE 1931 color matching functions.

is known that there are differences to real human observers sensitivity functions. Furthermore, the CIEXYZ space does not take into account the spatial context of reflecting signal, luminance level or chromatic adaptation of the eyes, all of which are known to affect to the color appearance of the object [88]. However, the objective of many CIE color spaces, such as the CIEXYZ, is only to specify physical aspects of color stimuli, not the *color appearance*. The main goal of these colorimetric spaces is to represent the spectral power distribution function of color stimulus using small amount of parameters [88].

In order to allow communication of different devices in a color reproduction chain, standard color spaces are used. International Electrotechnical Commission (IEC) and International Standards Organization (ISO) have accepted standard RGB space (sRGB) as the default color space for multimedia applications [88]. The sRGB coordinates can be transformed from the CIEXYZ coordinates via linear transformation and gamma correction. The gamma correction is included in order to compensate non-linear characteristic of CRT (Cathode Ray Tube) monitors. However, only a limited number of CIE colors can be reproduced additively in the sRGB space. Various RGB spaces have been defined in order to overcome the limitations of the sRGB space, such as Adobe RGB, Apple RGB, RIMM/ROMM RGB [44], [88].

2.5.2 CIEL*a*b* space and color differences

For psychophysical color spaces, a JND (Just Noticeable Difference) - value has been defined as the smallest possible distance between two colors, so that observer is able to discriminate between the colors. The CIEXYZ color space is not perceptually uniform, which means that color differences depend on the location of the space [170].

Another color space, the CIEL*a*b* 1976, have been formulated in order to calculate color differences which correlate to color differences sensed by human observer. The CIEL*a*b* is pseudouniform color space where the JND is almost independent of the location in the space. The CIEL*a*b* coordinates are obtained using the following transformation from the CIEXYZ coordinates [88]:

$$L^* = 116 f(Y/Y_N - 16), \tag{2.21}$$

$$a^* = 500 \left[f(X/X_N) - f(Y/Y_N) \right], \tag{2.22}$$

$$b^* = 200 \left[f(Y/Y_N) - f(Z/Z_N) \right], \tag{2.23}$$

where

$$f(\alpha) = \begin{cases} \sqrt[3]{\alpha} & , \alpha > 0.008856 \\ 7.787\alpha + 16/116 & , \alpha \leq 0.008856 \end{cases} \tag{2.24}$$

In the CIEL*a*b* space coordinate a^* represent the redness - greenness, b^* yellowness - blueness and L^* is the psychometric 1976 lightness of the stimulus. The CIEL*a*b* coordinates include adjustment for the illuminant via normalization with respect to the tristimulus values X_N, Y_N, Z_N of a perfect reflecting diffuser under the same lightning condition. For this perfect reflecting diffuser, called *reference white*, we have $(L^*, a^*, b^*) = (100, 0, 0)$. Color difference in the CIEL*a*b* space is calculated using the Euclidean distance

$$\Delta E^*_{ab} = \sqrt{(\Delta L^*)^2 + (\Delta a^*)^2 + (\Delta b^*)^2} \tag{2.25}$$

According to the literature the JND - value in this space varies from 0.5 to 1 [87], [88], [119].

In the CIEL*a*b* space, the *chroma* C^*_{ab} and the *hue-angle* h_{ab} are defined as

$$C^*_{ab} = \sqrt{a^{*2} + b^{*2}}, \quad h_{ab} = \arctan(b^*/a^*). \tag{2.26}$$

Updated version of the ΔE^*_{ab} is the CIE ΔE^*_{00}, which is still based on the CIEL*a*b* coordinates, but includes the contribution from chroma and hue-angle differences [88], pp. 113–115. It has been reported that the CIE ΔE^*_{00} especially improves the prediction of small color-differences and differences for "blue" colors [98].

Accuracy in terms of colorimetric error doesn't guarantee spectral similarity of two spectral signals. According to Wyszecki's hypothesis (dating back to 1953 [170]), in *fundamental color space* we can represent the signal as a combination of visual components [22], [23]. The components invisible to the observer are called as *visually inactive components*. In the context of fundamental color space, the equivalence classes defined by a sensor system are known as the *metameric sets*. In this set, spectrally different functions produce equal cone responses. The difference in the functions is only due to difference in observer null-space components. Fig. 2.7 gives an example of metameric reflectance with equal color stimulus values under the CIE D65 illumination, but significantly different representation in spectral space. Therefore, when the illumination is changed, the colorimetric error may increase significantly.

2.5.3 Simulated device dependent responses

By using the reflectance factor measurements, we are also able to simulate arbitrary color and device responses using the sensor

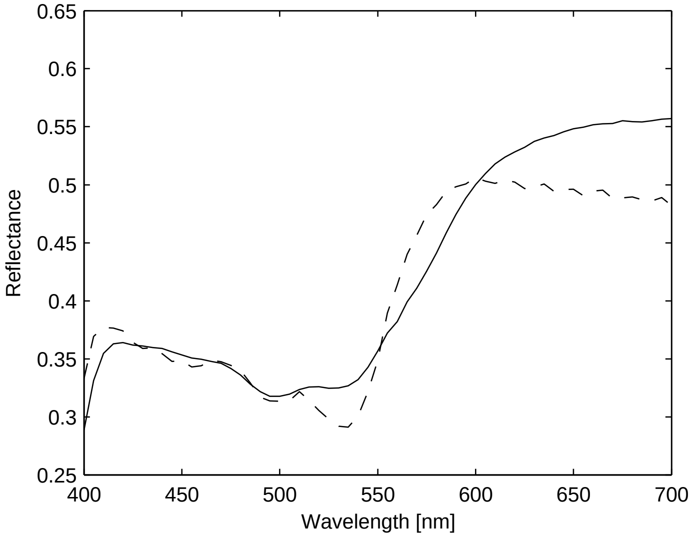

Figure 2.7: Simulated metameric pair of reflectance spectra for the L,M,S system when using the CIE D65 illumination. Dashed line corresponds to hyperspectral measurement of sample from Munsell Matte collection.

model (2.17). Simulated sensor models are highly beneficial in engineering and scientific calculation for constructing or optimizing sensor systems (e.g. [47], [48], [49], [51], [52], [58], [71], [110], [147], [164]). An example of this is presented in publication [P4], where we numerically simulated device response values of a multispectral sensor, which is tailored for photogrammetric airborne applications. These simulations employed hyperspectral reflectance ensembles for pine, spruce, and birch trees measured in the ground-level in conjunction with approximated spectral response functions and hemispheric daylight measurement.

The purpose of the simulation was to study classification performance of the multispectral measurements from these tree species. It is expected, that the results give indication about the suitability of this sensor for remote sensing purposes. The classification was formulated by using the kernel based learning and a Support Vector

Machine algorithm and is summarized in chapter 5.

2.5.4 Spectral color calibration of devices

In the case of digital cameras and scanners, *device dependent* values describe the response to incoming spectrum of light. A *color calibration* (or *characterization* [6], [44], [48], [63]) of a multispectral device, is defined as a construction of transformation from the device response values to the corresponding values in device independent CIE color spaces [34], [88], [165]. Due to the light source dependence of CIE values, it has been discussed that a *spectral color calibration* would provide significant benefits [34]. In color engineering applications, the idea is to calibrate the device responses values to reflectance spectra and then calculate some CIE color values by using this data.

A spectral color calibration can be seen to belong to a class of spectral estimation problems, which includes estimation of spectral transmittance [156] and illumination characteristics [37], [56], [58], [68], [93], [96] and estimation of camera characteristics [2], [7], [130]. Especially, a spectral color calibration can be identified as the surface reflectance estimation problem (see section 2.4.2) in the VIS range via multispectral responses (with broadband characteristics). Usually in this context the reflectance information correspond to reflectance factor measurements from a hyperspectral device.

A reflectance estimation process can be based on empirical modeling or on theoretical physical modeling. In physical modeling the spectral characteristics $\{s_i\}_{i=1}^{k}$, the non-linearities $\{\Gamma_i\}_{i=1}^{k}$ and the light source l are utilized in the inversion of (2.17) (e.g. [29], [47], [52], [59], [112]– [115], [121], [145]– [149], [162]). In some models, the light source can be unknown [99]. This dissertations mainly considers the empirical model, where all the above characteristics are unknown. In this case, a priori calibration data is used to construct the calibration transform via interpolation and regression models. The empirical approach can be motivated by the fact that the physical characterization of the measurement device or light source can be inaccurate, laborious or it may require expensive devices, such as monochromators. The same approach can be used in reversed sense, i.e. in the estimation of device dependent multispectral response for a given hyperspectral input [6].

Empirical models for calibration transform can be roughly divided to following categories: a) Local interpolation methods (by using triangulation- and tetraedrization) [3], [6], [66] b) Neural networks [4], [6], [20], [78], [101], [135], [136], [159], [166], c) Spline interpolation [6] and d) regression models. In regression, models are optimized with respect to some loss function for data fit. By far, the most widely used regression models in color science have been least squares polynomial fittings [6], [14], [15], [24], [48], [69]–[71], [88], [152], [153].

In the following we concentrate on methods from categories (b)–(d), and mainly concentrate on *global* approach in a sense that all training samples are included in model construction [50]. Due to computational costs, these methods are mainly suited for off-line estimation and construction of look-up-tables for faster calibration methods e.g. from category (a) [6], [44], [88].

In chapter 4 and [P1]–[P3], we consider the estimation of hyperspectral reflectance factors from the responses of ordinary RGB devices. We concentrate on reproducing kernel Hilbert space (RKHS) methods, which unify several methods from the categories (b)–(d). Currently, it seems that the categories (b)–(d) are considered as separate in color calibration and spectral estimation contexts. In [P2], we evaluate some RKHS models, which have not been used previously for the reflectance estimation. We also generalize and evaluate some previous approaches and also emphasize connections between previously proposed models in [P3]. Furthermore, in [P3] we make a connection to physical models by using an empirical kernel model with the sensor and light source characteristics. In [P1]–[P3], we evaluate several kernel functions, light sources, color ensembles and sensor systems in order to give information about color- and spectral accuracy of the kernel methods. The results are discussed and summarized in chapter 4.

3 Function spaces defined by a kernel

In the following we concentrate on kernel defined spaces, which give the functional form for the regression function in reflectance estimation and color calibration in chapter 4 and the form of decision function in spectral data classification in chapter 5.

Sometimes models in regression, classification and inverse problems lead to formulations which require a solution of a non-linear optimization problem. In some cases the non-linearity can be dealt by a "feature map" Φ, which defines a map from the original input space to a larger "feature space" $\mathcal{F}$ after which a linear representation is appropriate. Consider regression

$$y = f(\mathbf{x}) + \varepsilon, \text{ where } E[\varepsilon] = 0. \tag{3.1}$$

In many cases a linear representation $f(\mathbf{x}) = \mathbf{c}^T\mathbf{x}$ is not adequate. However, if we define a feature map

$$\Phi(\mathbf{x}) = (x_1, \dots, x_k, x_1^2, \dots, x_k^2, x_1x_2, \dots, x_1x_k, \dots) \tag{3.2}$$

we may be justified using $f(\mathbf{x}) = \mathbf{c}^T\Phi(\mathbf{x})$. Using the feature map we can formulate a class of functions

$$\{f | f(\mathbf{x}) = \mathbf{c}^T\Phi(\mathbf{x}), \mathbf{c} \in \mathcal{F}\}. \tag{3.3}$$

This is the space of functions from which the solutions for optimization problems are searched. The solution for some specific problem is found by selecting the coefficient vector $\mathbf{c}$ that has some optimality properties.

As an example of (3.3) we define a kernel induced function space and demonstrate its properties. In particular we consider the properties of Reproducing Kernel Hilbert Spaces (RKHS) by mainly following [5], [26], [33], [41], [107], [138], [143], [151], [163], [167]. It turns out that the kernel defines an inner product between the elements of feature space, a property which is sometimes called a "kernel trick" in the machine learning literature. This property enables the use of infinite dimensional RKHSs with continuous elements in algorithmic implementations.

In this dissertation we focus on regression, interpolation and classification models to estimate (approximate) multivariate functions. A common feature of the methods we use is that there is a training set $S = \{(\mathbf{x}_1, y_1), ..., (\mathbf{x}_m, y_m)\} \subset \mathbb{R}^k \times \mathbb{R}$ and we minimize

$$V[f] = \sum_{j=1}^{m} \mathcal{L}(y_j, f(\mathbf{x}_j)) + \sigma^2 \|f\|_{\mathcal{H}}^2, \tag{3.4}$$

over $f \in \mathcal{H}$, where $\mathcal{H}$ is a RKHS induced by a kernel. The first term on the right hand side measures the fit of the function to the training data under some *loss function* $\mathcal{L}$. The squared error loss $\mathcal{L}(y, f(\mathbf{x})) = (y - f(\mathbf{x}))^2$, is used in all regression and interpolation models (reflectance estimation and color calibration). Classification (with binary valued function $f : \mathbf{x} \to \{-1, 1\}$) is considered via $\mathcal{L}(y, f(\mathbf{x})) = |1 - yf(\mathbf{x}))|_+$, with $|a|_+ = a$ if $a > 0$ and $|a|_+ = 0$ otherwise. The term $\|f\|_{\mathcal{H}}^2$ in (3.4) is the square of a norm or a semi-norm of f and σ^2 is a scalar parameter, which allows us to control the balance between lack of fit and the RKHS norm of the solution. The term $\sigma^2 \|f\|_{\mathcal{H}}^2$ is called as a *regularization functional* and its purpose is to reduce wiggliness of f and improve generalization properties of models.

The purpose of this chapter is to give some examples of the RKHS space $\mathcal{H}$ in (3.4), which are not well known in color science. The RKHS framework allows us to construct kernels and feature maps, in a way that provides a link between several models that have been used previously in reflectance estimation and color calibration. We will see that the models are independent of the dimension of data and their spacing in the input domain. These are significant benefits in spectral imaging, when the dimension of the input space may vary or the spacing of the data points cannot be controlled.

Let $\mathcal{X}$ be a closed and bounded subset of $\mathbb{R}^k$. In most of our applications it will be assumed that $\mathcal{X} = [0,1]^k$. We will denote $\mathcal{H}$ a Hilbert space of real functions defined on $\mathcal{X}$ and endowed with an inner product $\langle \cdot, \cdot \rangle_{\mathcal{H}}$ and a corresponding norm $\| \cdot \|_{\mathcal{H}}$. The notation $\| \cdot \|$ is used for the usual Euclidean norm.

3.1 HILBERT SPACE

Let H be a real vector space. An *inner product* (or scalar product [81]) on H is function $\langle \cdot, \cdot \rangle$: $H \times H \to \mathbb{R}$, with properties

1. $\langle f, g \rangle = \langle g, f \rangle$
2. $\langle af + bg, v \rangle = a \langle f, v \rangle + b \langle g, v \rangle$
3. $\langle f, f \rangle \geq 0 \quad \text{and} \quad \langle f, f \rangle = 0 \quad \text{iff} \quad f = 0,$

for $f, g, v \in H$ [82]. A pre-Hilbert space (inner product space) is a vector space that has an inner-product $\langle \cdot, \cdot \rangle$ and norm $\langle f, f \rangle = \|f\|^2$ induced by the inner product. A pre-Hilbert space H is called a *Hilbert-space,* if all Cauchy sequences $\{f_n\}_{n \geq 1}$ ($\|f_n - f_m\| \to 0$, as $n, m \to \infty$), converge to some element in H (i.e. $\|f_n - f\| \to 0$ for some $f \in H$, as $n \to \infty$). For all $f, g \in H$ we have *Cauchy-Schwarz inequality*

$$|\langle f, g \rangle| \leq \|f\| \|g\|. \tag{3.5}$$

In the following, the integration is assumed in the sense of Lebesgue ([82]), although questions concerning measurability are neglected. Two well known infinite dimensional Hilbert spaces are presented below.

- Space $L^2(\mathcal{X})$ is a space of square integrable real valued functions

  $$\int_{\mathcal{X}} f^2(x) dx < \infty, \text{ for all } f \in L^2(\mathcal{X}). \tag{3.6}$$

  The $L^2(\mathcal{X})$ is a Hilbert space with inner product

  $$\langle f, g \rangle = \int_{\mathcal{X}} f g dx, \tag{3.7}$$

and a norm

$$\|f\| = \langle f, f \rangle^{\frac{1}{2}}. \tag{3.8}$$

- Space l^2 is a space of square summable series

$$\sum_{i=1}^{\infty} a_i^2 < \infty, \text{ for all } \{a_i\}_{i=1}^{\infty} \in l^2. \tag{3.9}$$

The l^2 is the Hilbert space of sequences $\mathbf{a} = \{a_i\}_{i=1}^{\infty}$, with an inner product

$$\mathbf{a}^T\mathbf{b} = \langle \{a_i\}_{i=1}^{\infty}, \{b_i\}_{i=1}^{\infty} \rangle = \sum_{i=1}^{\infty} a_i b_i, \tag{3.10}$$

and a norm

$$\|\mathbf{a}\| = (\mathbf{a}^T\mathbf{a})^{\frac{1}{2}}. \tag{3.11}$$

In the following $\mathcal{C}(\mathcal{X})$ is the complete normed space (i.e. Banach space) of continuous functions, with norm

$$\|f\|_\infty = \sup_{\mathbf{x} \in \mathcal{X}} |f(\mathbf{x})|. \tag{3.12}$$

3.2 REPRODUCING KERNEL HILBERT SPACE

A function $\kappa : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ of two variables is *symmetric* if

$$\kappa(\mathbf{x}, \mathbf{y}) = \kappa(\mathbf{y}, \mathbf{x}), \tag{3.13}$$

for all $\mathbf{x}, \mathbf{y} \in \mathcal{X}$. It is *positive definite* if for all finite collections $\{\mathbf{x}_i\}_{i=1}^{l} \subseteq \mathcal{X}$ and real numbers $\{a_i\}_{i=1}^{l}$

$$\sum_{i,j=1}^{l} a_i a_j \kappa(\mathbf{x}_i, \mathbf{x}_j) \geq 0, \tag{3.14}$$

Defining a vector $\mathbf{a} = (a_1, \ldots, a_l) \in \mathbb{R}^l$ and a matrix $K \in \mathbb{R}^{l \times l}$ with elements $K_{ij} = \kappa(\mathbf{x}_i, \mathbf{x}_j)$, we can write (3.14) as

$$\mathbf{a}^T K \mathbf{a} \geq 0, \tag{3.15}$$

and identify K as positive semi-definite. Function κ is *strictly positive definite*, if matrix K is positive definite, i.e. $\mathbf{a}^T K \mathbf{a} > 0$ for $\mathbf{a} \neq 0$. Function $\kappa(\cdot,\cdot) \in \mathcal{H}$ is said to be *continuous* if

$$|\kappa(\mathbf{x},\mathbf{z}) - \kappa(\mathbf{v},\mathbf{y})| \to 0, \text{ when } \|(\mathbf{x},\mathbf{z}) - (\mathbf{v},\mathbf{y})\| \to 0 \,. \tag{3.16}$$

A Hilbert space $\mathcal{H}$ is called a *reproducing kernel Hilbert space* (RKHS) if there exists a symmetric function $\kappa : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ and an inner-product $\langle \cdot,\cdot \rangle_{\mathcal{H}}$ such that for all $f \in \mathcal{H}$ and for all $\mathbf{x} \in \mathcal{X}$

$$\kappa(\mathbf{x},\cdot) \in \mathcal{H} \tag{3.17}$$

and

$$\langle f(\cdot), \kappa(\mathbf{x},\cdot) \rangle_{\mathcal{H}} = f(\mathbf{x}). \tag{3.18}$$

It follows that

$$\langle \kappa(\mathbf{x},\cdot), \kappa(\mathbf{z},\cdot) \rangle_{\mathcal{H}} = \kappa(\mathbf{x},\mathbf{z}). \tag{3.19}$$

The properties (3.18) and (3.19) are called *reproducing properties* and the function κ is called the *kernel of the space $\mathcal{H}$*.

The kernel of RKHS is positive definite, because by (3.19) and the bilinearity of the inner-product

$$\begin{aligned} \sum_{i,j=1}^{l} a_i a_j \kappa(\mathbf{x}_i,\mathbf{x}_j) &= \sum_{i,j=1}^{l} a_i a_j \langle \kappa(\mathbf{x}_i,\cdot), \kappa(\mathbf{x}_j,\cdot) \rangle_{\mathcal{H}} \\ &= \langle \sum_{i=1}^{l} a_i \kappa(\mathbf{x}_i,\cdot), \sum_{j=1}^{l} a_j \kappa(\mathbf{x}_j,\cdot) \rangle_{\mathcal{H}} = \| \sum_{i=1}^{l} a_i \kappa(\mathbf{x}_i,\cdot) \|_{\mathcal{H}}^2 \geq 0, \end{aligned} \tag{3.20}$$

where $\|\cdot\|_{\mathcal{H}}$ is the norm induced by $\langle \cdot,\cdot \rangle_{\mathcal{H}}$.

It can be shown, that for every RKHS there is a *unique* positive definite kernel $\kappa : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ such that $\kappa(\mathbf{x},\cdot) \in \mathcal{H}$ for all $\mathbf{x}$ and the reproducing properties hold. Conversely, for every positive definite kernel there exists a unique RKHS of real valued functions on $\mathcal{X}$ [5].

To construct an example of a RKHS, let $\{\varphi_i\}_{i=1}^{\infty}$ be a set of linearly independent functions such that every finite subset of $\{\varphi_i\}_{i=1}^{\infty}$ is linearly independent. A finite set $\{\varphi_i\}_{i=1}^{N}$ is linearly independent if

$$\sum_{i=1}^{N} c_i \varphi_i(\mathbf{x}) = 0, \text{ only if } c_1 = \cdots = c_N = 0. \tag{3.21}$$

Define a function

$$\kappa(\mathbf{x},\mathbf{y}) = \sum_{i=1}^{\infty} \gamma_i \varphi_i(\mathbf{x})\varphi_i(\mathbf{y}), \tag{3.22}$$

where $\gamma_i > 0$ for all i, such that

$$\sup_{\mathbf{x},\mathbf{y}\in\mathcal{X}} \kappa(\mathbf{x},\mathbf{y}) < \infty. \tag{3.23}$$

Suppose the series converges uniformly on $\mathcal{X} \times \mathcal{X}$ [33], i.e. for every $\epsilon > 0$ there exist N' such that for all $\mathbf{x},\mathbf{y} \in \mathcal{X}$ and $N \geq N'$

$$|\kappa(\mathbf{x},\mathbf{y}) - \sum_{k=1}^{N} \gamma_k \varphi_k(\mathbf{x})\varphi_k(\mathbf{y})| < \epsilon. \tag{3.24}$$

Now, consider a class of functions

$$\mathcal{H} = \left\{ f \,|\, f(\mathbf{x}) = \sum_{i=1}^{\infty} c_i \sqrt{\gamma_i} \varphi_i(\mathbf{x}),\ \sum_{i=1}^{\infty} c_i^2 < \infty \right\} \tag{3.25}$$

The elements are well-defined, since by the Cauchy-Schwarz inequality

$$|\sum_{i=1}^{\infty} c_i \sqrt{\gamma_i} \varphi_i(\mathbf{x})| \leq \left(\sum_{i=1}^{\infty} c_i^2 \right)^{1/2} \kappa(\mathbf{x},\mathbf{x})^{1/2}, \tag{3.26}$$

for all $f \in \mathcal{H}$ and $\mathbf{x} \in \mathcal{X}$. This is finite by (3.23) and (3.25). It follows from (3.23) and (3.26) that the series $\sum_{i=1}^{N} c_i \sqrt{\gamma_i} \varphi_i$ converges uniformly to $f \in \mathcal{H}$, when $N \to \infty$ [26], p.26.

An inner product between functions $f = \sum c_i \sqrt{\gamma_i} \varphi_i$ and $g = \sum d_i \sqrt{\gamma_i} \varphi_i$ is defined

$$\langle \sum_{i=1}^{\infty} c_i \sqrt{\gamma_i} \varphi_i(\mathbf{x}), \sum_{i=1}^{\infty} d_i \sqrt{\gamma_i} \varphi_i(\mathbf{y}) \rangle_{\mathcal{H}} = \sum_{i=1}^{\infty} c_i d_i. \tag{3.27}$$

The function $\langle \cdot, \cdot \rangle_{\mathcal{H}}$ defined in (3.27) is an inner-product, because it is real-valued, symmetric, i.e. $\langle f, g \rangle_{\mathcal{H}} = \langle g, f \rangle_{\mathcal{H}}$ and bilinear, i.e $\langle a_1 f_1 + a_2 f_2, g \rangle_{\mathcal{H}} = a_1 \langle f_1, g \rangle_{\mathcal{H}} + a_2 \langle f_2, g \rangle_{\mathcal{H}}$. Furthermore

$$\langle f, f \rangle_{\mathcal{H}} = \sum_{i=1}^{\infty} c_i^2 \geq 0 \tag{3.28}$$

with equality only if $c_i = 0$ for all i.

It can be seen that $\mathcal{H}$ is a RKHS. Firstly, to show that $\kappa(\cdot, \mathbf{x}) \in \mathcal{H}$, note that by (3.22) $\kappa(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^{\infty} \sqrt{\gamma_i} d_i \varphi_i(\mathbf{y})$, where $d_i = \sqrt{\gamma_i} \varphi_i(\mathbf{x})$. It follows that

$$\sum_{i=1}^{\infty} (\sqrt{\gamma_i} \varphi_i(\mathbf{x}))^2 = \kappa(\mathbf{x}, \mathbf{x}) < \infty, \tag{3.29}$$

and therefore $\kappa(\mathbf{x}, \cdot) \in \mathcal{H}$. Secondly, to show (3.18) pick an $f \in \mathcal{H}$ as defined in (3.25). Using the above representation of $\kappa(\mathbf{x}, \mathbf{y})$, we get

$$\langle f(\cdot), \kappa(\mathbf{x}, \cdot) \rangle_{\mathcal{H}} = \sum_{i=1}^{\infty} c_i d_i = \sum_{i=1}^{\infty} c_i \sqrt{\gamma_i} \varphi_i(\mathbf{x}), \tag{3.30}$$

where $d_i = \sqrt{\gamma_i} \varphi_i(\mathbf{x})$. It follows that function in (3.22) is the kernel of $\mathcal{H}$ and positive definite by (3.20).

Some examples of kernels follow. Recall that we have $\mathcal{X} \subset \mathbb{R}^k$. Probably the most widely used kernel can be defined as

$$\kappa(\mathbf{x}, \mathbf{y}) = \mathbf{x}^T A \mathbf{y}, \tag{3.31}$$

where A is some symmetric positive definite matrix. Using the eigenvalue decomposition $A = V \Lambda V^T$, where $V = [\mathbf{v}_1, \ldots, \mathbf{v}_k]$ is an orthogonal matrix and Λ is a diagonal matrix with elements $\lambda_i > 0$, $i = 1, \ldots, k$. The kernel can be written

$$\mathbf{x}^T A \mathbf{y} = \sum_{i=1}^{k} \lambda_i \mathbf{x}^T \mathbf{v}_i \mathbf{v}_i^T \mathbf{y}, \tag{3.32}$$

which is a positive definite and symmetric quadratic form.

As a second example, consider

$$\kappa(\mathbf{x}, \mathbf{y}) = g(\mathbf{x}) g(\mathbf{y}), \tag{3.33}$$

where $g : \mathcal{X} \to \mathbb{R}$ is some function. In this case it can be assumed that $\varphi_i(\mathbf{x}) = g(\mathbf{x})$ and $\gamma_i = 1$ for $i = 1$ and $\varphi_i(\mathbf{x}) \equiv 0$ for $i > 1$. The elements in the space $\mathcal{H}$ are now scalings of function φ_1.

One can gain additional insight into kernels by considering interpretations from probability theory. Let us consider a real random vector $\mathbf{c} = (c_1, \ldots, c_k)^T$, covariance matrix $\mathrm{Cov}(\mathbf{c}) = A$ and real random variables $\mathbf{x}^T\mathbf{c}$ and $\mathbf{y}^T\mathbf{c}$. The kernel (3.31) can be interpreted as a covariance of random variables, i.e. $\mathrm{cov}(\mathbf{x}^T\mathbf{c}, \mathbf{y}^T\mathbf{c}) = \mathbf{x}^T A\mathbf{y}$. Furthermore, it follows that a random vector $\mathbf{z} = (\mathbf{x}_1^T\mathbf{c}, \ldots, \mathbf{x}_l^T\mathbf{c})$ has covariance matrix $\mathrm{Cov}(\mathbf{z}) = K$, $K_{ij} = \kappa(\mathbf{x}_i, \mathbf{x}_j)$. This is a useful point of view for deriving rule 2 below.

Let positive definite kernels κ_1 and κ_2 be defined on $\mathcal{X} \times \mathcal{X}$. The following algebraic rules lead to new positive definite kernels:

1. $\kappa(\mathbf{x}, \mathbf{y}) = a\kappa_1(\mathbf{x}, \mathbf{y}) + b\kappa_2(\mathbf{x}, \mathbf{y}),\ a, b \in \mathbb{R}^+$.
2. $\kappa(\mathbf{x}, \mathbf{y}) = \kappa_1(\mathbf{x}, \mathbf{y})\kappa_2(\mathbf{x}, \mathbf{y})$.
3. $\kappa(\mathbf{x}, \mathbf{y}) = P^+(\kappa_1(\mathbf{x}, \mathbf{y}))$, where P^+ is polynomial with positive coefficients, i.e. $P^+(z) = \{\sum_{i=1}^d \alpha_i z^i, \alpha_1, \ldots, \alpha_n \in \mathbb{R}^+\}$.
4. $\kappa(\mathbf{x}, \mathbf{y}) = \exp(\kappa_1(\mathbf{x}, \mathbf{y}))$.

The fact that rule 1 defines a kernel is easily proven using the properties of κ_1 and κ_2. To verify the rule 2 (proved originally in [124] and [151]), assume zero-mean random variables $\mathbf{s} = (s_1, \ldots, s_l)^T$ and $\mathbf{v} = (v_1, \ldots, v_l)^T$ with associated covariance matrices $K_1(i, j) = E[s_i s_j]$ and $K_2(i, j) = E[v_i v_j]$ in $\mathbb{R}^l$. Element-wise product of the random variables is written as $\mathbf{s} \cdot \mathbf{v} = (s_1 v_1, \ldots, s_l v_l)^T$ with covariance matrix $K(i, j) = E[s_i v_i s_j v_j]$. Assuming that the random variables are independent, we have $K(i, j) = E[s_i s_j]E[v_i v_j] = K_1(i, j)K_2(i, j)$. The positive definite matrix K is now associated to the product kernel in rule 2. The rule 3 directly follows from the rules 1 and 2. Rule 4 is also valid because the exponential function can be written using the Taylor expansion

$$\exp(z) = \sum_{i=0}^{\infty} \frac{1}{i!} z^i, \tag{3.34}$$

and therefore the kernel can be defined as the limit of kernels in rule 3, when $d \to \infty$ [143], p. 77.

As a third example of a well-known kernel, consider first kernels $\kappa_1(\mathbf{x}, \mathbf{y}) = \exp(-\mathbf{x}^T\mathbf{x}/2\sigma^2)\exp(-\mathbf{y}^T\mathbf{y}/2\sigma^2)$ and $\kappa_2(\mathbf{x}, \mathbf{y}) = \mathbf{x}^T\mathbf{y}/\sigma^2$, where $\sigma^2 > 0$. Using the rule 2. and rule 4. above, we can construct a *Gaussian kernel* in the following way

$$\kappa(\mathbf{x},\mathbf{y}) = \frac{\exp(\mathbf{x}^T\mathbf{y}/\sigma^2)}{\exp(\mathbf{x}^T\mathbf{x}/2\sigma^2)\exp(\mathbf{y}^T\mathbf{y}/2\sigma^2)} = \exp(-\|\mathbf{x}-\mathbf{y}\|^2/2\sigma^2). \tag{3.35}$$

Decomposition of the Gaussian kernel on $\mathbb{R} \times \mathbb{R}$ can be written as

$$\begin{aligned} \kappa(x,y) &= \exp(-(x-y)^2/2\sigma^2) \\ &= \exp(-x^2/2\sigma^2)\exp(-y^2/2\sigma^2)\sum_{i=0}^{\infty}\frac{(xy/\sigma^2)^i}{i!}, \end{aligned} \tag{3.36}$$

where $\gamma_i = (1/\sigma^2)^i/i!$ and $\varphi_i(x) = x^i \exp(-x^2/2\sigma^2)$.

3.3 MERCER KERNELS

A set of functions $\{\psi_i\}_i$ in a Hilbert space $\mathcal{H}$ is called *orthonormal basis*, if

1. $\langle\psi_i,\psi_j\rangle_{\mathcal{H}} = 0$, when $i \neq j$ (orthogonality),
2. $\langle\psi_i,\psi_i\rangle_{\mathcal{H}} = 1$, for all i,
3. $f = \sum_{i=1}^{\infty}\langle f,\psi_i\rangle_{\mathcal{H}}\psi_i$, for all $f \in \mathcal{H}$,

Properties 1. and 2. define *orthonormal set* in space $\mathcal{H}$. It can be shown that $\{\sqrt{\gamma_i}\varphi_i\}_{i=1}^{\infty}$, where $\psi_i = \sqrt{\gamma_i}\varphi_i$, is orthonormal basis of $\mathcal{H}$ in (3.25) with respect to the inner product $\langle\cdot,\cdot\rangle_{\mathcal{H}}$ in (3.27). Properties 1. and 2. are valid, since $\langle\sqrt{\gamma_i}\varphi_i,\sqrt{\gamma_j}\varphi_j\rangle_{\mathcal{H}} = 1$, iff $i = j$. Property 3. follows, since for all $f \in \mathcal{H}$, $f = \sum_{i=1}^{\infty} c_i\sqrt{\gamma_i}\varphi_i$ and $c_i = \langle f,\sqrt{\gamma_i}\varphi_i\rangle_{\mathcal{H}}$.

If a (symmetric) kernel $\kappa : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ is continuous with property

$$\int_{\mathcal{X}}\int_{\mathcal{X}} \kappa^2(\mathbf{x},\mathbf{y})d\mathbf{x}d\mathbf{y} < \infty. \tag{3.37}$$

then it is called a *Mercer kernel*. Let us define a linear operator $L_\kappa : L^2(\mathcal{X}) \to \mathcal{C}(\mathcal{X})$ as a mapping

$$(L_\kappa f)(\mathbf{x}) = \int_{\mathcal{X}} \kappa(\mathbf{x},\mathbf{y})f(\mathbf{y})d\mathbf{y}. \tag{3.38}$$

The continuity of the values of the operator L_κ follows from the continuity of the kernel κ [26], p. 33. A scalar γ is an eigenvalue and function φ is the corresponding eigenfunction of the operator L_κ, if

$$(L_\kappa \varphi)(\mathbf{x}) = \int_{\mathcal{X}} \kappa(\mathbf{x}, \mathbf{y})\varphi(\mathbf{y})d\mathbf{x} = \gamma\varphi(\mathbf{x}). \tag{3.39}$$

It is known from the Spectral theorem ([26], p. 27.) and Mercer theorem ([26], p. 34.), that the Mercer kernel can be decomposed to uniformly and absolutely convergent series (3.22), where $\gamma_1 \geq \gamma_2 \geq \ldots \geq 0$ are the eigenvalues and $\{\varphi_i\}_{i=1}^{\infty} \subset L^2(\mathcal{X})$ is a set of orthonormal (in space $L^2(\mathcal{X})$) eigenfunctions of the operator L_κ. Using (3.37) it follows that

$$\int_{\mathcal{X}} \int_{\mathcal{X}} \kappa^2(\mathbf{x}, \mathbf{y})d\mathbf{x}d\mathbf{y} = \sum_{i=1}^{\infty} \gamma_i^2 < \infty. \tag{3.40}$$

The eigenfunctions corresponding to Mercer kernel are continuous, because the kernel κ in (3.38) is continuous. The RKHS $\mathcal{H}$ of (3.25) is now defined by the set $\{\sqrt{\gamma_i}\varphi_i\}_{i=1}^{\infty}$ consisting of continuous functions. The functions in $\mathcal{H}$ are linear combinations of continuous eigenfunctions and therefore continuous by the uniform convergence of the series in (3.25).

In the following it is assumed that $\gamma_i > 0$ for $i = 1, 2, \ldots$. If some of the eigenvalues are zero, the RKHS is a linear subspace of $L^2(\mathcal{X})$ spanned by the eigenvectors corresponding to non-zero eigenvalues [26].

For Mercer kernels, the RKHS $\mathcal{H}$ in (3.25) is a subspace of $L^2(\mathcal{X})$, and by the uniqueness of the Fourier coefficients $\langle f, \varphi_i \rangle$, we can identify $d_i = \langle f, \varphi_i \rangle = c_i\sqrt{\gamma_i}$. The RKHS in (3.25) can now be characterized as follows:

$$\mathcal{H} = \left\{ f \in L^2(\mathcal{X}) \,|\, f(\mathbf{x}) = \sum_{i=1}^{\infty} d_i\varphi_i(\mathbf{x}),\ \sum_{i=1}^{\infty} d_i^2/\gamma_i < \infty \right\}. \tag{3.41}$$

This shows that the squared coefficients $d_i^2 \to 0$ faster than $\gamma_i \to 0$.

3.3.1 Function space identified by a Mercer kernel

Let κ be a Mercer kernel on $\mathcal{X} \times \mathcal{X}$. Let us now consider a set H_0, where the elements are defined as a linear combination of Mercer kernels

$$H_0 = \left\{ f \mid f(\mathbf{x}) = \sum_{i=1}^{l} a_i \kappa(\mathbf{x}_i, \mathbf{x}) : \mathbf{x}_i \in \mathcal{X}, i = 1, ..., l \right\} \tag{3.42}$$

Inner product between functions $f = \sum_{i=1}^{l} a_i \kappa(\mathbf{x}_i, \cdot) \in H_0$ and $g = \sum_{j=1}^{k} b_i \kappa(\mathbf{z}_j, \cdot) \in H_0$ is defined as

$$\langle f, g \rangle_H = \sum_{i=1}^{l} \sum_{j=1}^{k} a_i b_j \kappa(\mathbf{x}_i, \mathbf{z}_j). \tag{3.43}$$

This is inner product since it is real valued, symmetric, bilinear and the property $\langle f, f \rangle_H \geq 0$ follows from the positive definiteness of the kernel. Furthermore, assuming that $\langle f, f \rangle_H = 0$, then

$$|f(\mathbf{x})|^2 = |\langle f, \kappa(\mathbf{x}, \cdot) \rangle_H|^2 \leq \langle f, f \rangle_H \langle \kappa(\mathbf{x}, \cdot), \kappa(\mathbf{x}, \cdot) \rangle_H = 0 \tag{3.44}$$

for all $\mathbf{x} \in \mathcal{X}$ using Cauchy-Schwartz inequality. Clearly for any $f \in H_0$, we have the reproducing property

$$\langle f(\cdot), \kappa(\mathbf{x}, \cdot) \rangle_H = f(\mathbf{x}). \tag{3.45}$$

Norm induced by this inner product is written as $\|f\|_H = \sqrt{\langle f, f \rangle_H}$.

By completing H_0 with respect to limits of converging sequences, we obtain a unique Hilbert space H with the inner product $\langle \cdot, \cdot \rangle_H$, such that: (1) For all $\mathbf{z} \in \mathcal{X}$, $\kappa(\mathbf{x}, \mathbf{z})$ belongs to H as a function of $\mathbf{x}$, (2) For all $\mathbf{x} \in \mathcal{X}$ and $f \in H$ we have the reproducing properties with respect to the inner product $\langle \cdot, \cdot \rangle_H$ and the kernel κ, (3) The set $\{\kappa(\mathbf{x}, \cdot) \mid \mathbf{x} \in \mathcal{X}\}$ is dense in H, (4) The space H consists of continuous functions. Detailed proof of 1.-4. can be found from [26], pp. 35–36.

From the discussion above we see that the elements in H and $\mathcal{H}$ can be defined using two different representations

$$f = \sum_j a_j \kappa(\cdot, \mathbf{x}_j) = \sum_i c_i \sqrt{\gamma_i} \varphi_i, \tag{3.46}$$

if we identify $c_i = \sqrt{\gamma_i} \sum_j a_j \varphi_i(\mathbf{x}_j)$. Following result can be proved

Theorem 1. *([26], p. 37.) The Hilbert spaces $\mathcal{H}$ and H are identical.*

3.3.2 Feature map associated to the kernel

This section was started with a discussion of a "feature map" Φ and a "feature space" $\mathcal{F}$, which would allow a modeling approach for different machine learning tasks. The kernel based machine learning literature usually associates two feature maps to the kernel as follows [151]

- The kernel decomposition (3.22) induces a *feature space* of square summable series l^2 and the corresponding *feature map*

  $$\Phi : \mathcal{X} \to l^2 : \mathbf{x} \mapsto \Phi(\mathbf{x}) = (\sqrt{\gamma_i}\varphi_i(\mathbf{x}))_{i=1}^{\infty} \tag{3.47}$$

  The property $\Phi(\mathbf{x}) \in l^2$ is valid, because the sum $\sum_i \gamma_i \varphi_i^2(\mathbf{x})$ converges to $\kappa(\mathbf{x},\mathbf{x})$ by (3.22). Now the representation for the function can be written in a *primal form*

  $$f = \mathbf{c}^T\Phi(\cdot). \tag{3.48}$$

- A *reproducing kernel feature map* is defined as

  $$\Phi_\kappa : \mathcal{X} \to \mathbb{R}^{\mathcal{X}} : \mathbf{x} \mapsto \kappa(\mathbf{x},\cdot), \tag{3.49}$$

  where $\mathbb{R}^{\mathcal{X}} := \{f : \mathcal{X} \to \mathbb{R}\}$. Now the representation for the function can be written in a *dual form*

  $$f = \mathbf{a}^T\mathbf{k}, \tag{3.50}$$

  $\mathbf{a} = (a_1, a_2, \ldots)^T$ and $\mathbf{k} = (\Phi_\kappa(\mathbf{x}_1), \Phi_\kappa(\mathbf{x}_2), ...)^T$ by (3.46).

By using the above feature maps, we can use complex non-linear algorithm in input space via simple linear representation (3.48) or (3.50) in feature spaces. The feature space associated to kernels is a Hilbert space, which allows the formulation of geometric relationships between the elements. In the RKHS framework the feature space corresponding to map Φ can be infinite dimensional, so that we cannot implement the primal form representation (3.48) explicitly in practical computations, e.g. in regression and classification. For example, this is the case for the Gaussian kernel with decomposition (3.36). The information from a feature map Φ can be efficiently exploited in the learning algorithms by using the corresponding kernel $\kappa(.,.)$ in closed form, which corresponds to an inner product in the feature space via the decomposition (3.22)

$$\kappa(\mathbf{x},\mathbf{y}) = \Phi(\mathbf{x})^T\Phi(\mathbf{y}) = \langle\Phi(\mathbf{x}),\Phi(\mathbf{y})\rangle_{l^2}. \tag{3.51}$$

The property (3.51) is called as a "kernel trick" in the machine learning context and in practical calculations leads to the dual form function representations (3.50) using finite kernel spans. In fact, every algorithm which can be transformed to consider only inner products between data points, can be also "kernelized" using the dual form representation via kernel [151]. The "kernelization" in this context means that the inner product in the original data domain $\mathcal{X}$ is replaced with the kernel evaluation (3.51) between the data points.

In many cases, only the kernel is defined explicitly without consideration of the feature map Φ. Furthermore, the feature map associated to the kernel is not unique. For example, the decomposition given in (3.36) doesn't correspond to the eigendecomposition of the Gaussian kernel. However, for a given domain $\mathcal{X}$, the feature maps from $\mathcal{X}$ into Hilbert spaces are essentially equivalent. There is *isometric isomorphism* between the subspaces of different feature spaces, i.e. a linear bijective mapping between spaces exists which preserves inner products [107], p. 8.

In the following, we introduce kernels which are used in this dissertation. For a detailed discussion about kernels and their construction, see [106], [107], [127]– [129], [133], [139], [143], [151], [163], [169].

3.3.3 Gaussian kernel

The Gaussian kernel was defined as

$$\kappa(\mathbf{x},\mathbf{z}) = \exp\left(-\frac{\|\mathbf{x}-\mathbf{z}\|_2^2}{2\varsigma^2}\right), \tag{3.52}$$

where $\mathbf{x}, \mathbf{z} \in \mathcal{X}$ and $\varsigma \neq 0$ is the length-scale parameter. It can be shown that the Gaussian kernel is strictly positive definite in $\mathcal{X}$ for all $\varsigma^2 > 0$ [26], [106]. This kernel is *isotropic*, i.e.

$$\kappa(\mathbf{x}, \mathbf{z}) = g(\|\mathbf{x} - \mathbf{z}\|). \tag{3.53}$$

From isotropy it follows that the function is *translation invariant* (stationary), i.e. $\kappa(\mathbf{x} + \mathbf{s}, \mathbf{z} + \mathbf{s}) = \kappa(\mathbf{x}, \mathbf{z}) \quad \text{for all} \quad \mathbf{s} \in \mathcal{X}$ and *unitary invariant*, i.e. $\kappa(U\mathbf{x}, U\mathbf{z}) = \kappa(\mathbf{x}, \mathbf{z}) \quad \text{for all} \quad U^T = U^{-1}$. Due to unitary invariance, representations of data in different orthonormal bases lead to same kernel values. Isotropy is a natural property in contexts, in which the origin and the coordinate directions are arbitrary. A possible feature map for the Gaussian kernel in scalar case was presented in (3.36).

A Gaussian kernel can more generally be defined as ([1], [127])

$$\kappa(\mathbf{x}, \mathbf{z}) = \exp\left(-\frac{\|\mathbf{x} - \mathbf{z}\|_A^2}{2\varsigma^2}\right), \tag{3.54}$$

where $\|\mathbf{x} - \mathbf{z}\|_A^2 = (\mathbf{x} - \mathbf{z})^T A(\mathbf{x} - \mathbf{z})$ and A is some symmetric, positive semi-definite matrix. If $A \neq I$, the kernel is called *anisotropic*.

3.3.4 Polynomial kernel

A polynomial kernel of degree d is defined as

$$\kappa(\mathbf{x}, \mathbf{z}) = (a\mathbf{x}^T\mathbf{z} + b)^d, \tag{3.55}$$

where $\mathbf{x}, \mathbf{z} \in \mathcal{X} \subset \mathbb{R}^k$, $a > 0$, $b \geq 0$ and $d = 0, 1, 2, \ldots$ are free parameters. In the following, we use notation $\mathbf{x}^\alpha = x_1^{\alpha_1} \cdots x_k^{\alpha_k}$, $\alpha! = \alpha_1! \cdots \alpha_k!$ and $|\alpha| = \alpha_1 + \ldots + \alpha_k$ for $\alpha = (\alpha_1, \ldots, \alpha_k)$. For $a = 1$, $b = 1$ an inhomogenous polynomial feature map of total degree d consists of N monomial terms $\{\mathbf{x}^\alpha \| \alpha| \leq d\}$ and is written as ([26])

$$\Phi : \mathcal{X} \to \mathbb{R}^N : \mathbf{x} \to \left(1, (\mathbf{x}^{\alpha_i}(C_{\alpha_i}^d)^{1/2})_{i=1}^{N-1}\right), \tag{3.56}$$

where $\alpha_i \in \mathbb{R}^k$ and $1 \leq |\alpha_i| \leq d$, for $i = 1, \ldots, N-1$ and

$$N = \begin{pmatrix} k+d \\ k \end{pmatrix} = \frac{(k+d)!}{k!d!}, \quad C_{\alpha_i}^d = \frac{d!}{\alpha_i!}. \tag{3.57}$$

The monomials $\mathbf{x}^{\alpha}$, $\alpha \in \mathbb{N}_0^d$, are linearly independent [169], p. 20.

For example, if we choose $a = 1$, $b = 1$ and $d = 2$ for $\mathcal{X} \subset \mathbb{R}^3$, we have ten dimensional feature space and the components of the feature map are represented as

$$(1, (\mathbf{x}^{\alpha_i})_{i=1}^9) = (1, x_1, x_2, x_3, x_1^2, x_2^2, x_3^2, x_1x_2, x_1x_3, x_2x_3), \tag{3.58}$$

with weights

$$(1, ((C_{\alpha_i}^d)^{1/2})_{i=1}^9) = (1, \sqrt{2}, \sqrt{2}, \sqrt{2}, 1, 1, 1, \sqrt{2}, \sqrt{2}, \sqrt{2}). \tag{3.59}$$

The feature map above, however, does not necessarily correspond to the eigendecomposition of integral operator. The properties of the feature map depend on the domain $\mathcal{X}$. In the unit sphere $\mathcal{X} = S^{k-1}$ in $\mathbb{R}^k$ and $a = 1$, $b = 1$ and $d = 2$, we have eigenfunctions [107]

$$\{\varphi_i(\mathbf{x})\}_{i=1}^5 = \left(\frac{1}{\sqrt{2\pi}}, \frac{x_1}{\sqrt{\pi}}, \frac{x_2}{\sqrt{\pi}}, \frac{2x_1x_2}{\sqrt{\pi}}, \frac{x_1^2 - x_2^2}{\sqrt{\pi}}\right), \tag{3.60}$$

eigenvalues

$$\{\gamma_i\}_{i=1}^3 = \left(3\pi, 2\pi, \frac{\pi}{2}\right) \tag{3.61}$$

and the feature map

$$\Phi(\mathbf{x}) = \left(\sqrt{\frac{3}{2}}, \sqrt{2}x_1, \sqrt{2}x_2, \sqrt{2}x_1x_2, \frac{x_1^2 - x_2^2}{\sqrt{2}}\right). \tag{3.62}$$

The polynomial feature map is widely used in regression due to its practicality. The usage of polynomial kernel can be motivated by its efficiency to compute the combinatorially large number of monomial features. Some possible "drawbacks" for polynomial kernels are listed as: (1) the RKHS induced by polynomial kernel of finite degree d is finite dimensional and not dense in $\mathcal{C}([0,1]^k)$, so it is not possible to approximate continuous functions arbitrarily well [26], [129]; (2) although convenient in practice, increase in the degree of polynomial kernel is defined in discrete steps and cannot be controlled continuously; (3) the kernel is not translation invariant; (4) interpolation of data becomes difficult, because the high degree polynomials leads to oscillation in the regions between the interpolation points [102], [125]. This well known phenomenon is called as *polynomial wiggle* [102], p. 278.

The smoothing properties of RKHS norms $\|f\|^2_{\mathcal{H}}$ induced by the Gaussian kernel and polynomial kernels have been discussed in [107] and [151], pp. 99–101. In this context, the spectral decomposition (Mercer decomposition) of a kernel has a significant role. Especially, under certain conditions, explicit decompositions (3.22) with $L^2(\mathcal{X})$-orthonormal eigenfunctions $\{\varphi_i\}_i$ and eigenvalues $\{\gamma_i\}_i$ have been presented for these kernels [107]. For both kernels, eigenvalue spectra γ_i decreases and the eigenfunctions φ_i have decreasing smoothness properties. The Gaussian kernel has more sharper smoothing properties, with exponentially decaying eigenvalues [107].

3.4 SPLINE KERNELS

In the following polynomial feature map of total degree d is denoted as

$$\Psi_d(\mathbf{x}) = (\psi_1(\mathbf{x}), \ldots, \psi_{N_d^k}(\mathbf{x})), \tag{3.63}$$

where ψ_i are monomial terms of total degree less or equal than d and $N_d^k = (k+d)!/k!d!$ is the binomial coefficient. Kernel κ is defined as *d-conditionally (strictly) positive definite* (d-cpd kernel) in $\mathcal{X}$ iff it is symmetric and if for all finite collections $\{\mathbf{x}_i\}_{i=1}^l \subseteq \mathcal{X}$, real numbers $\{b_i\}_{i=1}^l$ and for Ψ_{d-1}, we have $\sum_{i,j}^l b_i b_j \kappa(\mathbf{x}_i, \mathbf{x}_j) \geq 0$, when

$$\sum_{i=1}^{l} b_i \psi_j(\mathbf{x}_i) = 0 \tag{3.64}$$

for $j = 1, \ldots, N_{d-1}^k$ and $\mathbf{x}_1, \ldots, \mathbf{x}_l \in \mathcal{X}$. According to the constraint (3.64), the coefficients $\mathbf{b} = (b_1, \ldots, b_l)$ annihilate all the polynomials of total degree less than d, i.e.

$$\sum_{i=1}^{l} b_i \mathbf{c}^T \Psi_{d-1}(\mathbf{x}_i) = 0. \tag{3.65}$$

Using matrix notation, the d-cpd property (3.64) of a kernel can be written as

$$\mathbf{b}^T K\mathbf{b} \geq 0, \quad \text{when} \quad \Psi_{d-1}(X)^T\mathbf{b} = \mathbf{0}, \tag{3.66}$$

where $K_{ij} = \kappa(\mathbf{x}_i, \mathbf{x}_j)$ and the matrix $\Psi_{d-1}(X)$ consists of all the polynomial terms up to total degree $d-1$

$$\Psi_{d-1}(X) = [\Psi_{d-1}(\mathbf{x}_1), \ldots, \Psi_{d-1}(\mathbf{x}_l)]^T \in \mathbb{R}^{l\times N_{d-1}^k}. \tag{3.67}$$

It is assumed that the points $\{\mathbf{x}_i\}_{i=1}^l$ are distinct and define matrix $\Psi_{d-1}(X)$ with full column rank. In this case a least squares regression with the points $\{\mathbf{x}_i\}_{i=1}^l$ and a polynomial of total degree of $d-1$ is unique.

3.4.1 Natural cubic spline

Assume that $a < x_1 < \cdots < x_l < b$, for $[a, b] \subset \mathbb{R}$ and $f(x_i) = g(x_i)$ for $i = 1, \ldots, l$. A *cubic spline* f is defined as a twice continuously differentiable function, that interpolates the data $\{x_i, g(x_i)\}_{i=1}^l$, so that $f(x_i) = g(x_i)$ for $i = 1, \ldots, l$. The cubic spline is a third-order polynomial in each interval $[x_i, x_{i+1}]$, $i = 1, \ldots, l-1$ and is called a *natural cubic spline*, if it is a linear polynomial on outer intervals $[a, x_1]$ and $[x_l, b]$ [169], pp. 8–11. In this case, its second and third derivatives vanish at x_1 and x_l, i.e. $f''(x_1) = f'''(x_1) = 0$ and $f''(x_l) = f'''(x_l) = 0$. A natural cubic spline minimizes the squared L^2 norm of the second derivative

$$\|f\|_{\mathcal{H}^2}^2 = \int_a^b |f''(x)|^2 dx. \tag{3.68}$$

It is known that every natural cubic spline can be alternatively written as a combination of parametric and non-parametric part

$$f(x) = c_1 + c_2 x + \frac{1}{12}\sum_{i=1}^{l} b_i|x - x_i|^3, \text{ s.t. } \sum_{i=1}^{l} b_i = \sum_{i=1}^{l} b_i x_i = 0. \tag{3.69}$$

The parametric part of the natural cubic spline is a first order polynomial and the non-parametric part is defined by the kernel span, where the kernel of the spline is defined as

$$\kappa(x, z) = \frac{1}{12}|x - z|^3. \tag{3.70}$$

3.4.2 Thin plate splines

In space $\mathbb{R}^k$, the interpolating function is defined as function which passes through the given data points $\{\mathbf{x}_i, g(\mathbf{x}_i)\}_{i=1}^l$. It is assumed that $\{\mathbf{x}_i\}_{i=1}^l$ is such that least squares regression on monomials $\psi_1, \ldots, \psi_{N_d^k}$ of total degree less than d is unique (see (3.63)). In the following $\mathcal{H}^d$ is a space of functions whose partial derivatives of total order d are in $L^2(\mathbb{R}^k)$ [167]. A *natural thin plate spline of degree d* (or *Duchon's spline* or *Meinguet's surface spline*) [31], [45], [104], [167], leads to an unique interpolating function in $\mathcal{H}^d$ which minimizes the following functional

$$\|f\|^2_{\mathcal{H}^d} = \sum_{\alpha_1+\ldots+\alpha_k=d} \frac{d!}{\alpha_1!\ldots\alpha_k!} \int_{\mathbb{R}^k} \left(\frac{\partial^d f}{\partial x_1^{\alpha_1} \ldots \partial x_k^{\alpha_k}} \right)^2 d\mathbf{x} \tag{3.71}$$

This functional defines a *semi-norm* in the space $\mathcal{H}^d$, i.e. $\|f\|_{\mathcal{H}^d} \geq 0$, $\|af\|_{\mathcal{H}^d} = |a|\|f\|_{\mathcal{H}^d}$ and $\|f+g\|_{\mathcal{H}^d} \leq \|f\|_{\mathcal{H}^d} + \|g\|_{\mathcal{H}^d}$, for all scalars a and $f, g \in \mathcal{H}^d$. The space $\mathcal{H}^d$ endowed with a semi-norm $\|f\|_{\mathcal{H}^d}$ is a RKHS [167]. A more detailed discussion on space $\mathcal{H}^d$ is given in [104].

The functional (3.71) has a null space of multivariate polynomial functions of total degree less than d that evaluate to zero, i.e. $\|\mathbf{c}^T \Psi_j(\mathbf{x})\|^2_{\mathcal{H}^d} = 0$, for $j = 1, \ldots, d-1$. For $\mathbb{R}^2$ and $d = 2$ the functional is

$$\|f\|^2_{\mathcal{H}^2} = \int_{\mathbb{R}} \int_{\mathbb{R}} \left(\frac{\partial^2 f}{\partial x_1^2} \right)^2 + \left(\frac{\partial^2 f}{\partial x_2^2} \right)^2 + 2 \left(\frac{\partial^2 f}{\partial x_1 \partial x_2} \right)^2 dx_1 dx_2 \tag{3.72}$$

and the null-space consist of $\mathcal{C}(\mathbb{R}^2)$ functions

$$f(\mathbf{x}) = c_0 + c_1 x_1 + c_2 x_2 = \mathbf{c}^T \Psi_1(\mathbf{x}). \tag{3.73}$$

Writing f_x for the partial derivative of f with respect to x, we can write the functional for $\mathbb{R}^3$ and $d = 2$ as

$$\|f\|^2_{\mathcal{H}^2} = \int_{\mathbb{R}^3} (f^2_{x^1x^1} + f^2_{x^2x^2} + f^2_{x^3x^3} + 2[f^2_{x^1x^2} + f^2_{x^1x^3} + f^2_{x^2x^3}]) d\mathbf{x}. \tag{3.74}$$

The null space consists of $\mathcal{C}(\mathbb{R}^3)$ functions of the form

$$f(\mathbf{x}) = c_0 + c_1 x_1 + c_2 x_2 + c_3 x_3 = \mathbf{c}^T \Psi_1(\mathbf{x}). \tag{3.75}$$

A *natural thin plate spline of degree d* is written as

$$\begin{aligned} &f(\mathbf{x}) = \mathbf{c}^T\Psi_{d-1}(\mathbf{x}) + \textstyle\sum_{i=1}^{l} b_j\kappa_i(\mathbf{x},\mathbf{x}_i) \\ &\text{s.t. } \Psi_{d-1}(X)^T\mathbf{b} = \mathbf{0}, \end{aligned} \tag{3.76}$$

where κ is a d-conditionally strictly positive definite function and the matrix of $\Psi_{d-1}(X)$ polynomial terms of total degree $d-1$ has full column rank [31], [104], [167]. For $d = 2$ the spline in $\mathbb{R}^k$ is the analog of natural cubic spline in one dimension and is called as *natural thin plate spline*. The function κ is isotropic on $\mathbb{R}^k$ and is defined for odd $2d-k > 0$:

$$\kappa(\mathbf{x},\mathbf{z}) = c_{d,k}\|\mathbf{x}-\mathbf{z}\|^{2d-k},\ c_{d,k} = \frac{\Gamma(k/2-d)}{2^{2d}\pi^{k/2}(d-1)!} \tag{3.77}$$

and for even $2d-k > 0$:

$$\kappa(\mathbf{x},\mathbf{z}) = c_{d,k}\|\mathbf{x}-\mathbf{z}\|^{2d-k}\log(\|\mathbf{x}-\mathbf{z}\|), \tag{3.78}$$

where

$$c_{d,k} = \frac{(-1)^{k/2+1+d}}{2^{2d-1}\pi^{k/2}(d-1)!(d-k/2)!}, \tag{3.79}$$

and Γ denotes Gamma-function [45], [167], not to be confused with Γ in (2.17). In this dissertation, κ is called *thin plate spline kernel of degree d*. The requirement $2d-k > 0$ means that the form of regularization functional that can be allowed depends on the dimension of the space $\mathbb{R}^k$ [31], [104]. For example the regularization functional based on second derivatives can be only used, when $k \leq 3$.

As an example, for $\mathbb{R}^2$ and $d = 2$, the kernel is written as

$$\kappa(\mathbf{x},\mathbf{z}) = \frac{1}{16\pi}\|\mathbf{x}-\mathbf{z}\|^2\log(\|\mathbf{x}-\mathbf{z}\|) \tag{3.80}$$

and for $\mathbb{R}^3$ and $d = 2$

$$\kappa(\mathbf{x},\mathbf{z}) = -\frac{\sqrt{\pi}}{4\pi^{3/2}}\|\mathbf{x}-\mathbf{z}\|. \tag{3.81}$$

3.4.3 Splines with infinite number of knots

Vapnik has introduced splines with infinite number of knots [163]. A spline of order d with infinite number of knots in $(0, a)$, $0 < a < \infty$, can be defined as

$$f(x) = \sum_{i=0}^{d} a_i x^i + \int_0^a a(t)(x-t)_+^d dt, \tag{3.82}$$

where $a_0, \dots, a_d, a(t) \in \mathbb{R}$, $(x - t_k)_+^d = (x - t_k)^d$, if $x > t_k$ and $(x - t_k)_+^d = 0$ otherwise. This is a generalization of a piecewise polynomial function with finite number of knots

$$f(x) = \sum_{i=0}^{d} a_i^* x^i + \sum_{i=1}^{m} a_i (x - t_i)_+^d \tag{3.83}$$

where $t = (t_1, \dots, t_m)$, $t_i = ia/m$ and $i = 1, \dots, m$.

A one dimensional spline kernel corresponding to (3.82) is defined as ([163], p. 465.)

$$\begin{aligned} \kappa(x, z) &= \int_0^a (x-t)_+^d (z-t)_+^d dt + \sum_{i=0}^{d} x^i z^i \\ &= \sum_{i=0}^{d} \frac{\binom{d}{i}}{2d - i + 1} \min(x, z)^{2d-i+1} |x - z|^i + \sum_{i=0}^{d} x^i z^i. \end{aligned} \tag{3.84}$$

A multidimensional spline kernel for $\mathbf{x} = (x_1, \dots, x_k)^T$ and $\mathbf{z} = (z_1, \dots, z_k)^T$ in $\mathcal{X} \subseteq \mathbb{R}^k$ can be defined as a product

$$\kappa(\mathbf{x}, \mathbf{z}) = \prod_{i=1}^{k} \kappa_i(x_i, z_i), \tag{3.85}$$

where κ_i's are one-dimensional kernels [163], p. 463.

4 Kernel based reflectance estimation and color calibration

We will estimate the reflectance information pixelwise, using a simulated or real multispectral digital camera with response functions $\{s_i\}_{i=1}^{k}$, where $k < 10$. We mainly concentrate on three-channel multispectral responses, which are obtained using one or two illumination sources. The reflection model is assumed to be (2.11) and imaging is based on sensor model (2.17), where $\Lambda = [\lambda_1, \lambda_2]$ denotes the wavelength interval. In the following $\langle \cdot, \cdot \rangle$ denotes the inner-product of $L^2(\Lambda)$.

The goal of estimating reflectance properties from multispectral measurements is to estimate the vectorial representation for measured reflectance factors that are obtained from a hyperspectral device with narrow-band characteristics. It is assumed, that the multi- and hyperspectral devices cover the same wavelength interval Λ and the same spatial area on the object surface.

Let $\mathbf{x} = (x_1, \ldots, x_k)^T$ denote the vector of measured multispectral responses according to model (2.17), i.e

$$x_i = \Gamma_i(t_g t_e \langle r, l s_i \rangle), \quad i = 1, \ldots, k, \tag{4.1}$$

where $s_i \in L^2(\Lambda)$ is the i^{th} spectral response function, Γ_i is a typically non-linear function, t_g and t_e are fixed scalars corresponding to illumination and view geometry and exposure time according to (2.12) and (2.17), $r \in L^2(\Lambda)$ is the reflectance spectrum, and $l \in L^2(\Lambda)$ is the spectral power distribution of the light source.

By ignoring the non-linear functions of the channels, the corresponding hyperspectral reflectance factors $\mathbf{q} = (q_1, \ldots, q_n)^T$ via system $\{\tilde{s}_j\}_{j=1}^{n}$ according to model (2.17) and (2.19) are written as

$$q_j = a_g \langle r, l \tilde{s}_j \rangle / \langle l, \tilde{s}_j \rangle, \quad j = 1, \ldots, n, \tag{4.2}$$

where a_g is (possibly) a term corresponding to illumination and viewing geometry according to (2.14).

From one point of view, we are looking for a mapping $\mathbf{x} \to \mathbf{q}$ between two spectral representations, where the hyperspectral representation is normalized with the measurement from diffuse standard. Standard CIEXYZ color calibration of multispectral responses is closely related to the reflectance estimation. In CIEXYZ calibration, the response functions $\tilde{s}_j$ of hyperspectral device are replaced in (4.2) with the three CIEXYZ color matching functions $\{\tilde{s}_j\}_{j=1}^{3} = \{\overline{x}, \overline{y}, \overline{z}\}$ and the normalization term is defined by the color matching function $\overline{y}$ according to equations in (2.20).

In the following we focus mainly on the case of reflectance estimation (specifically spectral color calibration), when $n >> k$, and it is assumed that

$$q_j \approx r(\lambda_j), \quad j = 1, \ldots, n, \tag{4.3}$$

where $\{\lambda_j\}_{j=1}^{n}$ corresponds to a partition of a interval Λ into n subintervals of equal length. We make the assumption (4.3) in order to calculate CIEXYZ color values (2.20) from the hyperspectral data. Especially, we assume that functions r, l, s_i and $\tilde{s}_i$ are defined pointwise and continuous. In this case, a more natural spectral space would be a RKHS space rather than $L^2(\Lambda)$ [122]. However, in this dissertation we don't investigate this further. A functional analytic framework for spectral color spaces has been discussed in [96].

We consider the case of empirical reflectance estimation, where the system $\{s_i\}_{i=1}^{k}$, non-linearities $\{\Gamma_i\}_{i=1}^{k}$ and light source l are unknown, but fixed. Empirical modeling will be performed using training data consisting of m samples. First, the training samples are measured with the hyper- and multispectral devices keeping the measurement geometry and the illuminations fixed. Then, a regression model is constructed to empirically describe the relationship between the measurements $\{\mathbf{x}_i\}_{i=1}^{m}$ and $\{\mathbf{q}_i\}_{i=1}^{m}$. The calculated model coefficients are then used to infer the reflectance data $\mathbf{q}$ in future data, from multispectral responses $\mathbf{x}$ that have been measured under the same lighting and measurement geometry.

As an example, in the painting imaging, the reflectance estimation has been performed by training a regression model by using multispectral and hyperspectral measurements of samples from a color palette (e.g a palette used by an artist) or samples from some color calibration chart [72], [70]. These training measurements can be also obtained from the painting itself, so that m amount of pointwise (pixelwise) measurements from appropriately chosen locations are performed. After the measurements are performed and the regression model is formulated, the obtained model parameters can then be used to estimate the spectral reflectance image of the whole painting, by using the multispectral images. By spectral reflectance image we mean an image, where each pixel is represented by a spectral reflectance vector $\mathbf{q}$. As a final application, the estimated reflectance image can be used in the digital archiving for museums or in the artwork reproduction and restoration.

The multispectral devices usually have large amount of pixels in the sensor chip and therefore they can be used to estimate reflectance images (with high spatial resolution) of objects faster that would be possible via hyperspectral devices (see Table 2.1 in chapter 2). For example, it has been suggested that a system, which combines an RGB color camera (Sinarback 54H) and a pair of absorption filters, is a practical choice for imaging cultural heritage [14]. The reflectance estimation results for GretagMacbeth ColorChecker target by using this 6-band system were found to provide equal accuracy with a hyperspectral 31-band LCTF system [14].

In the following we introduce the kernel based regression models for scalar valued case. The reflectance estimation and color calibration are presented as a vector valued regression in 4.1.6.

4.1 KERNEL BASED REGRESSION

In this section we consider a regression model, which relies on the feature space map Φ induced by a kernel as defined in (3.47). Assume that we have a training set of scalar outputs y and vector inputs $\mathbf{x}$

$$S = \{(\mathbf{x}_1, y_1), ..., (\mathbf{x}_m, y_m)\} \subset \mathbb{R}^k \times \mathbb{R}, \tag{4.4}$$

obtained from a random sampling of some density. The purpose of the regression is to estimate the conditional expectation of y given $\mathbf{x}$ using the data in S, or $y = f(\mathbf{x}) + \varepsilon$, where the expectation of the residual is zero; formally $E[y|\mathbf{x}] = f(\mathbf{x})$. A natural assumption is that the function of interest has some smoothness properties, so that similar inputs map to similar outputs. Intuitively, smoothness refers to the lack of wiggliness of the function [41], [50], [127], [128].

Recall that $\mathcal{H}$ denotes a RKHS space induced by a kernel κ as defined in chapter 3. We assume that f is an element of $\mathcal{H}$. In other words, the regression function is assumed to be of the form

$$f(\mathbf{x}) = \mathbf{c}^T \Phi(\mathbf{x}), \tag{4.5}$$

where $\mathbf{c} = (c_1, c_2, \ldots)$ and $\Phi(\mathbf{x}) = (\sqrt{\gamma_1}\varphi_1(\mathbf{x}), \sqrt{\gamma_2}\varphi_2(\mathbf{x}), \ldots)^T$, see (3.47). It is assumed that the feature map Φ is induced by a kernel function κ, so that the coefficients γ and functions φ correspond to the decomposition (3.22) with linearly independent set $\{\varphi_i\}_{i=1}^{\infty}$. For Mercer kernels this decomposition can correspond to the eigenvalues and eigenvectors of operator associated to the kernel κ according to (3.38).

The coefficients $\{c_i\}_{i=1}^{\infty}$ can be found by using the available training set S and minimizing the empirical error via some loss function $\mathcal{L}$

$$\arg\min_{f \in \mathcal{H}} \sum_{j=1}^{m} \mathcal{L}(y_j, f(\mathbf{x}_j)). \tag{4.6}$$

Clearly, if the number of coefficients c_i with $c_i \neq 0$, is infinite or exceeds m, it is possible that there is no unique solution to (4.6). Even if the number of non-zero coefficients is less than or equal to m, but large, the minimizing function function may be close to an interpolator of the training data, i.e. $f(\mathbf{x}_j) = y_j$, for $j = 1, \ldots, m$ and it may *overfit*. The estimated curve can be quite wiggly and therefore generalize poorly in some regions. For such reasons additional restrictions need to be put on the minimizing function. This is called *regularization* and is done by constraining the smoothness properties of the minimizing function. RKHSs allow us to do this in a powerful way by using the norm of the $\mathcal{H}$ as the measure for function smoothness.

We formulate an empirical minimization problem with regularization functional

$$\arg\min_{f\in\mathcal{H}} V[f] = \sum_{j=1}^{m} \mathcal{L}(y_j, f(\mathbf{x}_j)) + \sigma^2 \|f\|^2_{\mathcal{H}}. \tag{4.7}$$

The term in the right hand side measures the fit of the function to the training data under the loss function $\mathcal{L}$. The regularization functional $\|f\|^2_{\mathcal{H}}$ in this case is the norm or semi-norm of the function in the RKHS space $\mathcal{H}$.

In applications, the kernel κ is typically chosen in such a way that for smooth functions this regularization functional gives small values. The explicit form of the regularization functional was shown for the thin plate spline kernel in chapter 3. It was shown that this functional includes calculation of differentials, which measure the smoothness properties. The parameter σ^2 controls the trade-off between data fit and smoothness. A large value of σ^2 imposes a large penalty for function with a large norm.

4.1.1 Form of the regression estimator

For an arbitrary nonnegative and differentiable loss function $\mathcal{L}$, the solution of (4.7) can be written using the dual form $f = \sum_{j=1}^{m} a_j \kappa(\mathbf{x}_j, .)$ [151], pp.90-91. To show this, first note that for all $f \in \mathcal{H}$,

$$f(\mathbf{x}) = \sum_{j=1}^{m} a_j \kappa(\mathbf{x}, \mathbf{x}_j) + f_\perp, \tag{4.8}$$

where $f_\perp \in \mathcal{H}$ and $\langle f_\perp, \kappa(\mathbf{x}_j, \cdot)\rangle_{\mathcal{H}} = 0$ for all $j = 1, \ldots, m$. By using the reproducing property (3.18), $f(\mathbf{x}_i) = \langle f, \kappa(\mathbf{x}_i, \cdot)\rangle_{\mathcal{H}}$. Substituting (4.8) into this and using (3.19), we get that $f(\mathbf{x}_i) = \sum_{j=1}^{m} a_j \kappa(\mathbf{x}_i, \mathbf{x}_j)$, for $i = 1, \ldots, m$ and therefore the loss function term in (4.7) is independent of $f_\perp$. Second, $\|f\|^2_{\mathcal{H}} = \|\sum_{j=1}^{m} a_j \kappa(\mathbf{x}_j, .)\|^2_{\mathcal{H}} + \|f_\perp\|^2_{\mathcal{H}}$ due to the orthogonality of $\sum_{j=1}^{m} a_j \kappa(\mathbf{x}_j, .)$ and $f_\perp$. Therefore the minimum of (4.7) can be obtained by choosing $f_\perp = 0$.

Consider the special case of quadratic loss, i.e $\mathcal{L}(y_j, f(\mathbf{x}_j)) = (y_j - f(\mathbf{x}_j))^2$. Using the dual form we have

$$f = \sum_{j=1}^{m} a_j \kappa(\cdot, \mathbf{x}_j), \quad \|f\|_{\mathcal{H}}^2 = \sum_{i=1}^{m} \sum_{j=1}^{m} a_j a_i \kappa(\mathbf{x}_i, \mathbf{x}_j) \tag{4.9}$$

and the minimization problem (4.7) is written as

$$\arg\min_{\mathbf{a}} \|\mathbf{y} - K\mathbf{a}\|_2^2 + \sigma^2 \mathbf{a}^T K \mathbf{a}, \tag{4.10}$$

where matrix $K \in \mathbb{R}^{m \times m}$ is the kernel matrix of training data, with $K_{ij} = \kappa(\mathbf{x}_i, \mathbf{x}_j)$. To find the solution of (4.10), take the partial derivatives with respect to a_j for $j = 1, \ldots, m$ and set to zero

$$\frac{\partial}{\partial \mathbf{a}} (\mathbf{y}^T \mathbf{y} - 2\mathbf{a}^T K \mathbf{y} + \mathbf{a}^T K^2 \mathbf{a} + \sigma^2 \mathbf{a}^T K \mathbf{a}) = \mathbf{0}. \tag{4.11}$$

For the solution, we have

$$K(K\mathbf{a} + \sigma^2 \mathbf{a} - \mathbf{y}) = \mathbf{0}, \tag{4.12}$$

or equivalently

$$(K + \sigma^2 I_m)\mathbf{a} = \mathbf{y}, \tag{4.13}$$

where I_m is $m \times m$ identity matrix. Matrix $K + \sigma^2 I_m$ is positive definite when $\sigma^2 > 0$ and the unique dual form solution is

$$\mathbf{a} = (K + \sigma^2 I_m)^{-1} \mathbf{y}. \tag{4.14}$$

Function evaluation for a new test point $\mathbf{x}$ is formulated as

$$\hat{y} = \mathbf{k}_{\mathbf{x}}^T \mathbf{a} = \mathbf{k}_{\mathbf{x}}^T (K + \sigma^2 I_m)^{-1} \mathbf{y}, \tag{4.15}$$

where $\mathbf{k}_{\mathbf{x}}^T$ is a row vector containing the kernel evaluations between the training set and test point, with element i, $(\mathbf{k}_{\mathbf{x}}^T)_i = \kappa(\mathbf{x}_i, \mathbf{x})$.

The properties of the calculated regression function depend on the chosen kernel. For polynomial kernel (3.55) with $d < m$, the matrix K doesn't have full rank and the model approaches least squares polynomial fit, when $\sigma^2 \to 0$. It has been proven that the matrix K is invertible for distinct $\{\mathbf{x}\}_{i=1}^m$, when the elements have representation $K_{ij} = \kappa(\|\mathbf{x}_i - \mathbf{x}_j\|)$, and function κ is chosen to be a radial basis function with some properties [106]. This result is known as a Micchelli's Theorem, and it applies to many radial basis functions such as Gaussian kernel (3.52), Hardy's Multiquadrics and Inverse Multiquadrics [53]. For example, for the Gaussian kernel with non-zero scale parameter, the solution approaches the interpolator of the data, when $\sigma^2 \to 0$. The regularization parameter $\sigma^2 > 0$ guarantees, that the system (4.14) is numerically well-conditioned for every kernel. Regardless of chosen kernel, the solution approaches zero, when $\sigma^2 \to \infty$.

Previously in (4.5) we formulated the regression function in primal form by using the coefficients c_i, $i = 1, 2, \ldots$ and the feature map Φ in l^2 space. By using the decomposition (3.22) of the kernel, the regression function can be formulated in this infinite dimensional form

$$f(\mathbf{x}) = \sum_{j=1}^{m} a_j \sum_{i=1}^{\infty} \gamma_i \varphi_i(\mathbf{x}_j) \varphi_i(\mathbf{x}) = \sum_{i=1}^{\infty} c_i \sqrt{\gamma_i} \varphi_i(\mathbf{x}), \tag{4.16}$$

where

$$c_i = \sum_{j=1}^{m} a_j \sqrt{\gamma_i} \varphi_i(\mathbf{x}_j). \tag{4.17}$$

Consider the matrix $\Phi(X) = [\Phi(\mathbf{x}_1) \ldots \Phi(\mathbf{x}_m)]^T \in \mathbb{R}^{m \times \infty}$ and with $\mathbf{a} = (a_1, \ldots, a_m)^T$ and $\mathbf{c} = (c_1, c_1, \ldots)^T \in l^2$. Then, (4.17) can be written as

$$\mathbf{c} = \Phi(X)^T \mathbf{a}. \tag{4.18}$$

This shows that the practical solution in this infinite dimensional space can be given in terms of a_j and the coefficients c_i are only used implicitly.

For the quadratic loss in (4.7), the coefficients c_i can also be viewed as a solution to

$$\arg\min_{\mathbf{c}} \|\mathbf{y} - \Phi(X)\mathbf{c}\|_2^2 + \sigma^2 \mathbf{c}^T \mathbf{c}, \tag{4.19}$$

The usual least squares solution for $\sigma^2 > 0$ is $\hat{y} = \Phi(\mathbf{x})^T \mathbf{c}$, where

$$\mathbf{c} = (\Phi(X)^T \Phi(X) + \sigma^2 I_\infty)^{-1} \Phi(X)^T \mathbf{y}. \tag{4.20}$$

One can show that primal form solution (4.20) is equivalent to (4.18), where **a** is as in (4.14) ([P3], appendix).

4.1.2 Semi-parametric form

The model for the regression function can be generalized to semi-parametric form as follows [151], [167].

Let $\Psi(\mathbf{x}) = (\psi_1(\mathbf{x}), \ldots, \psi_N(\mathbf{x}))^T$ denote some feature map in $\mathcal{X} \subset \mathbb{R}^k$, with linearly independent elements. Let us formulate a space $\mathcal{H}$ with elements

$$f = g_1 + g_2,\ g_1 \in \mathcal{H}_1,\ g_2 \in \mathcal{H}_2, \tag{4.21}$$

where $\mathcal{H}_1 = \text{span}\{\psi_1, \ldots, \psi_N\}$ is Hilbert space and $\mathcal{H}_2$ denotes a RKHS induced by positive definite kernel κ. As an example, the most widely used semi-parametric model is formulated by using only constant feature, i.e. $N = 1$ and $\Psi(\mathbf{x}) = 1$. For a detailed discussion of this case, see [129]. In the case of $\Psi(\mathbf{x}) = 1$ and quadratic loss, the minimizer of (4.7) has been called a *Least Squares Support Vector Machine* [155].

By replacing $\|f\|^2_{\mathcal{H}}$ with $\|g_2\|^2_{\mathcal{H}_2}$ in (4.7), the form of the minimizer can be found as follows. For distinct $\{\mathbf{x}_j\}_{j=1}^m$, the functions in $\mathcal{H}$ can be written as

$$f = \sum_{j=1}^{N} c_j \psi_j + \sum_{j=1}^{m} b_j \kappa(\cdot, \mathbf{x}_j) + f_2, \tag{4.22}$$

where $f_2 \in \mathcal{H}_2$, $g_2 := \sum_{j=1}^m b_j \kappa(\cdot, \mathbf{x}_j) + f_2$ and $\langle f_2, \kappa(\cdot, \mathbf{x}_j)\rangle_{\mathcal{H}_2} = 0$, for $j = 1, \ldots, m$. By using the reproducing property (3.18) in $\mathcal{H}_2$, $f(\mathbf{x}_i) = \sum_{j=1}^N c_j \psi_j(\mathbf{x}_i) + \langle g_2, \kappa(\mathbf{x}_i, \cdot)\rangle_{\mathcal{H}}$. Substituting g_2 from (4.22) into this and using (3.19), we have $f(\mathbf{x}_i) = \sum_{j=1}^N c_j \psi_j(\mathbf{x}_i) + \sum_{j=1}^m b_j \kappa(\mathbf{x}_i, \mathbf{x}_j)$, $i = 1, \ldots, m$. This shows that the loss function term in (4.7) is independent of f_2. Second, $\|g_2\|^2_{\mathcal{H}_2} = \| \sum_{j=1}^m b_j \kappa(\mathbf{x}_j, .)\|^2_{\mathcal{H}_2} + \|f_2\|^2_{\mathcal{H}_2}$ due to the orthogonality of $\sum_{j=1}^m b_j \kappa(\mathbf{x}_j, .)$ and f_2 in $\mathcal{H}_2$. Therefore, by choosing $f_2 = 0$, the minimizer of (4.7) can be written as

$$f(\mathbf{x}) = \sum_{j=1}^{m} b_j \kappa(\mathbf{x}, \mathbf{x}_j) + \sum_{j=1}^{N} c_j \psi_j(\mathbf{x}). \tag{4.23}$$

In general, the regularization functional in (4.7) can be modified to take into account also the parametric part, e.g. by using a sum of two regularization functionals $\sigma_1^2 \|g_1\|_{\mathcal{H}_1}^2 + \sigma_2^2 \|g_2\|_{\mathcal{H}_2}^2$. In the following we concentrate on the regularization of g_2 only.

By using (4.23), the minimization problem (4.7) with quadratic loss and regularization functional $\|g_2\|_{\mathcal{H}_2}^2$, can be written as

$$\arg\min_{\{\mathbf{b},\mathbf{c}\}} V[\mathbf{b},\mathbf{c}] = \|\mathbf{y} - K\mathbf{b} - \Psi(X)\mathbf{c}\|_2^2 + \sigma^2 \mathbf{b}^T K \mathbf{b}, \tag{4.24}$$

where the matrix $\Psi(X) \in \mathbb{R}^{m \times N}$ with full column rank has elements $[\Psi(X)]_{ji} = \psi_i(\mathbf{x}_j)$, $m > N$ and K is kernel matrix with (i,j) elements $\kappa(\mathbf{x}_i, \mathbf{x}_j)$. To find the solution, take the partial derivatives of (4.24) with respect to b_j, $j = 1, \ldots, m$ and c_i, $i = 1, \ldots, N$ and set to zero to get

$$\begin{aligned} &(K + \sigma^2 I)\mathbf{b} + \Psi(X)\mathbf{c} - \mathbf{y} = \mathbf{0}, \\ &\Psi^T(X)\Psi(X)\mathbf{c} + \Psi^T(X)K\mathbf{b} - \Psi^T(X)\mathbf{y} = \mathbf{0} \end{aligned} \tag{4.25}$$

By multiplying the upper equation in (4.25) with $\Psi^T(X)$ from the left, it can be written $\Psi^T(X)(K + \sigma^2 I)\mathbf{b} + \Psi^T(X)\Psi(X)\mathbf{c} = \Psi^T(X)\mathbf{y}$. Substituting the $\Psi^T(X)\mathbf{y}$ to the lower equation in (4.25), it simplifies to $\sigma^2 \Psi^T(X)\mathbf{b} = \mathbf{0}$. The solution can now be calculated from the equations

$$\begin{aligned} &(K + \sigma^2 I)\mathbf{b} + \Psi(X)\mathbf{c} = \mathbf{y}, \\ &\Psi(X)^T \mathbf{b} = \mathbf{0} \end{aligned} \tag{4.26}$$

4.1.3 Semi-parametric form via d-cpd kernels

The kernel in semi-parametric form (4.24) can be modified to be d-conditionally (strictly) positive definite (d-cpd) according to conditions (3.64). Specifically, we concentrate on the d-cpd thin-plate spline kernel (3.77) (or (3.78)) in $\mathbb{R}^k$. The parametric part $\mathcal{H}_1$ above is now chosen to be spanned by the polynomials of maximal degree $d-1$, so that the $\mathcal{H} := \mathcal{H}^d$ is a RKHS of functions whose partial derivatives of total order d are in the $L^2(\mathbb{R}^k)$ [31], [167].

In this case, the space $\mathcal{H}$ is a direct sum of N^k_{d-1}-dimensional space $\mathcal{H}_1$ and a RKHS $\mathcal{H}_2$ [104], [167], p. 34. In the following $N := N^k_{d-1}$ and $\Psi(X) := \Psi_{d-1}(X)$. The regularization functional in (4.7) is (3.71), which is a norm in $\mathcal{H}_2$ and semi-norm in $\mathcal{H}$. The polynomials in $\mathcal{H}_1$ now belong to the null space of the regularization functional, i.e. $\|g_1 + g_2\|^2_{\mathcal{H}} = \|g_2\|^2_{\mathcal{H}}$. The minimization (4.7) with quadratic loss is written as

$$\begin{aligned} &\arg\min_{\{\mathbf{b},\mathbf{c}\}} \left(\|\mathbf{y} - K\mathbf{b} - \Psi(X)\mathbf{c}\|_2^2 + \sigma^2 \mathbf{b}^T K \mathbf{b}\right), \\ &\text{s.t.} \quad \Psi(X)^T \mathbf{b} = \mathbf{0} \end{aligned} \tag{4.27}$$

where $\Psi(X) = [\Psi(\mathbf{x}_1) \ldots \Psi(\mathbf{x}_m)]^T \in \mathbb{R}^{m \times N}$, $m > N$ is the matrix (3.67) of polynomial terms corresponding to the $\mathcal{H}_1$ and K is the kernel matrix with (i, j) elements $\kappa(\mathbf{x}_i, \mathbf{x}_j)$. The solution is unique and can be calculated by using the equations (4.26) [31], [167].

The spline model above can be generalized to *partial spline models*, the same way as the generalization to semi-parametric model was done above. In this case two parametric spaces are used in addition to kernel induced space: one correspond to some well-defined functions and the other is the polynomial part. These parametric spaces can be constrained to be in the null space of the kernel induced regularization functional as above [167], p. 75.

4.1.4 Calculation of semi-parametric solution

The coefficients in (4.27) can be calculated from a block matrix equation with full rank ([44], p. 148)

$$\begin{pmatrix} K + \sigma^2 I & \Psi(X) \\ \Psi(X)^T & 0 \end{pmatrix} \begin{pmatrix} \mathbf{b} \\ \mathbf{c} \end{pmatrix} = \begin{pmatrix} \mathbf{y} \\ \mathbf{0} \end{pmatrix}. \tag{4.28}$$

The coefficients $\mathbf{b}$ and $\mathbf{c}$ in (4.27) can be expressed by separating the residual to orthogonal directions (For original calculation via QR-decomposition, see [167], p. 33.). Let $\Psi(X) \in \mathbb{R}^{m \times N}$ be a matrix of rank N. From singular value decomposition

$$\Psi(X) = USV^T = [U_1 U_2] \begin{pmatrix} S_1 & 0 \\ 0 & S_2 \end{pmatrix} \begin{pmatrix} V_1 \\ V_2 \end{pmatrix}^T, \tag{4.29}$$

where $U \in \mathbf{R}^{m\times m}$, $V \in \mathbf{R}^{N\times N}$, $S \in \mathbf{R}^{m\times N}$, $U_1 \in R^{m\times N}$, $U_2 \in \mathbb{R}^{m\times(m-N)}$, S_2 is zero matrix and $S_1 \in \mathbb{R}^{N\times N}$ is diagonal with non-negative entries σ_i, where $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_N > 0$ [17]. The matrices V and U are orthonormal and therefore $V^T V = I$ and $U^T U = I$.

By (4.27) $\Psi(X)^T\mathbf{b} = VSU^T\mathbf{b} = \mathbf{0}$, and therefore $\mathbf{b} = U_2\mathbf{d}$, for some $\mathbf{d} \in \mathbb{R}^{m-N}$. Due to orthogonality, for any $\mathbf{z} \in \mathbb{R}^m$, we can separate the norm, i.e. $\|\mathbf{z}\|^2 = \|U_1U_1^T\mathbf{z} + U_2U_2^T\mathbf{z}\|^2 = \|U_1^T\mathbf{z}\|^2 + \|U_2^T\mathbf{z}\|^2$. By using this separation for (4.27), it simplifies

$$\begin{aligned}\|U_2^T\mathbf{y} - U_2^TKU_2\mathbf{d}))\|_2^2 \;&+\; \|U_1^T\mathbf{y} - U_1^TKU_2\mathbf{d} - S_1V_1^T\mathbf{c})\|_2^2 \\ &+\; \sigma^2\mathbf{d}^TU_2^TKU_2\mathbf{d}. \end{aligned} \tag{4.30}$$

For the minimizers $\hat{\mathbf{d}}, \hat{\mathbf{c}}$ we can write

$$\begin{cases} U_2^T\mathbf{y} = (U_2^TKU_2 + \sigma^2 I)\hat{\mathbf{d}} \Leftrightarrow \hat{\mathbf{d}} = (U_2^TKU_2 + \sigma^2 I)^{-1}U_2^T\mathbf{y}, \\ S_1V_1^T\hat{\mathbf{c}} = U_1^T\mathbf{y} - U_1^TKU_2\hat{\mathbf{d}} \Leftrightarrow \hat{\mathbf{c}} = V_1S_1^{-1}(U_1^T\mathbf{y} - U_1^TKU_2\hat{\mathbf{d}}) \end{cases} \tag{4.31}$$

By using $\hat{\mathbf{b}} = U_2\hat{\mathbf{d}}$, the estimate for $\mathbf{x}$ is obtained as

$$\hat{y} = \hat{\mathbf{b}}^T\mathbf{k}_{\mathbf{x}} + \hat{\mathbf{c}}^T\Psi(\mathbf{x}). \tag{4.32}$$

For $\sigma^2 \to \infty$, we have that $\hat{\mathbf{b}} \to 0$ and $\hat{\mathbf{c}} \to V_1S_1^{-1}U_1^T\mathbf{y}$ and the solution approaches to the polynomial least squares fit. For $\sigma^2 \to 0$ and full rank K, the solution approaches to interpolating model.

4.1.5 Connections to Gaussian processes and other models

We have presented models for reflectance estimation using functional analytic RKHS framework. These methods are well-known in the context of *approximation and interpolation with irregularly spaced data* [3], [8], [31], [104], [106], [140], [169]. However, the models are known in several fields of statistics and machine learning. Especially, the dual form regression (4.15) has been referred e.g. as *Regularization Network* [16], [126]– [128], *Radial Basis Function network* [53], *Gaussian process regression* [133] , *ridge regression in dual variables* [139] and *Least Squares Support Vector Machine* (with the additional constant term) [155]. Discussion about these connections has been presented e.g. in [33], [128]. Gaussian process framework for semi-parametric models is discussed in [118], [133].

In particular, Gaussian processes are related to RKHS models since (strictly) positive definite kernels correspond to the covariance of the Gaussian process [133]. In the following we provide alternative derivation of the result (4.15), which is classical in the field of Gaussian processes.

Theorem 2. *([132], p. 522) Let* $\mathbf{u}_1$ *and* $\mathbf{u}_2$ *be two normal random vectors. Then the probability distribution of* $\mathbf{u} = [\mathbf{u}_1^T \ \mathbf{u}_2^T]^T$ *is defined as*

$$\mathbf{u} = \begin{bmatrix} \mathbf{u}_1 \\ \mathbf{u}_2 \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} \right), \tag{4.33}$$

and the conditional distribution of $\mathbf{u}_2$ *given* $\mathbf{u}_1$*, is*

$$\mathbf{u}_2 | \mathbf{u}_1 \sim \mathcal{N}(\mu_2 + \Sigma_{21}\Sigma_{11}^{-1}(\mathbf{u}_1 - \mu_1), \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}), \tag{4.34}$$

provided that the inverses exist.

Let us consider scalar valued regression

$$y = f(\mathbf{x}) + \eta, \quad f(\mathbf{x}) = \Phi(\mathbf{x})^T \mathbf{c} \tag{4.35}$$

where

$$\mathbf{c} \sim \mathcal{N}(\mathbf{0}, \Sigma_c), \quad \eta \sim \mathcal{N}(0, \sigma_\eta^2). \tag{4.36}$$

A *Gaussian process* $f(\mathbf{x}) \sim \mathcal{GP}(E[f(\mathbf{x})], E[(f(\mathbf{x}) - E[f(\mathbf{x})])(f(\mathbf{x}) - E[f(\mathbf{x})])])$ is defined as a collection of random variables with a joint Gaussian distribution. Consider a training set $\{(\mathbf{x}_i, y_i)\}_{i=1}^m$ with function values $\mathbf{y} = (y_1, \ldots, y_m)^T$, where $y_i = f(\mathbf{x}_i) + \eta_i$, and the η_i's are independently and identically distributed (i.i.d.). Suppose such m values are jointly Gaussian. Then, we have the mean and covariance

$$E[f(\mathbf{x})] = \Phi(\mathbf{x})^T E[\mathbf{c}] = 0, \tag{4.37}$$

$$E[y_i y_t] = \Phi(\mathbf{x}_i)^T \Sigma_c \Phi(\mathbf{x}_t) + \sigma_\eta^2 \delta_{it} = \kappa(\mathbf{x}_i, \mathbf{x}_t) + \sigma_\eta^2 \delta_{it}, \tag{4.38}$$

where δ_{it} denotes Kronecker delta and κ is the covariance function of the process. Covariance matrix of vector $\mathbf{y}$ is

$$\mathrm{Cov}(\mathbf{y}) = \Phi(X)\Sigma_c\Phi(X)^T + \sigma_\eta^2 I = K(X, X) + \sigma_\eta^2 I, \tag{4.39}$$

where $K_{ij} = \kappa(\mathbf{x}_i, \mathbf{x}_j)$.

We can define the joint distribution of training points and test point y as

$$\begin{pmatrix} \mathbf{y} \\ y \end{pmatrix} \sim \mathcal{N}\left(\mathbf{0}, \begin{pmatrix} K(X,X) + \sigma_\eta^2 I & \mathbf{k}(\mathbf{x}) \\ \mathbf{k}(\mathbf{x}) & \kappa(\mathbf{x},\mathbf{x}) \end{pmatrix}\right), \tag{4.40}$$

where $(\mathbf{k}(\mathbf{x}))_i = \Phi(\mathbf{x})^T \Sigma_c \Phi(\mathbf{x}_i)$, $i = 1, \ldots, m$. Using Theorem 2 we have

$$y|\Phi(\mathbf{x}), \Phi(X), \mathbf{y} \sim \mathcal{N}(\overline{y}, \mathrm{Cov}(y)) \tag{4.41}$$

where the mean value correspond to (4.15)

$$\overline{y} = \mathbf{k}(\mathbf{x})^T (K + \sigma_\eta^2 I)^{-1} \mathbf{y} \tag{4.42}$$

and

$$\mathrm{Cov}(y) = \kappa(\mathbf{x},\mathbf{x}) - \mathbf{k}(\mathbf{x})^T (K + \sigma_\eta^2 I)^{-1} \mathbf{k}(\mathbf{x}). \tag{4.43}$$

This framework allows to consider positive semi-definite RKHS kernels as a covariance function $\kappa(\mathbf{x}_i, \mathbf{x}_t)$ of the process. Especially, kernel with $0 < \kappa(\mathbf{x}_i, \mathbf{x}_t) \leq 1$ can be considered as a correlation of function values $f(\mathbf{x}_i)$ and $f(\mathbf{x}_t)$. As an example, we can consider the Gaussian kernel (3.52), where the correlation approaches to 1 for $\varsigma \to \infty$ and vanishes for $\varsigma \to 0$.

4.1.6 Estimation as a vector valued kernel regression

Next we continue the discussion of reflectance estimation and (spectral) color calibration. Assume that $\mathbf{x} \in \mathbb{R}^k$ and $\mathbf{q} \in \mathbb{R}^n$ correspond to multi- and hyperspectral measurements (4.1) and (4.2), and that we have a training set of size m

$$\{(\mathbf{x}_1, \mathbf{q}_1), ..., (\mathbf{x}_m, \mathbf{q}_m)\} \subset \mathbb{R}^k \times \mathbb{R}^n. \tag{4.44}$$

In color calibration, $\{\mathbf{q}_i\}_{i=1}^m \subset \mathbb{R}^3$ correspond to some arbitrary color coordinates, such as CIE colors. The goal is to estimate a mapping $\mathbf{x} \to \mathbf{q}$ by using a regression model. Let us define a matrix $Q = [Q_1 \ldots Q_n] \in \mathbb{R}^{m \times n}$ with column vectors $Q_s \in \mathbb{R}^m$ as

$$Q = [\mathbf{q}_1 \ldots \mathbf{q}_m]^T = [q_{js}], \; j = 1, \ldots, m, s = 1, \ldots, n. \tag{4.45}$$

In the simplest case, a vector valued regression model for mapping $\mathbf{x} \to \mathbf{q}$ is formulated with independent scalar regressions (4.7) for components of vector $\mathbf{q} = (q_1, \ldots, q_n)^T$. We focus on this approach by using the squared error loss and same model for all the n components ([P1]–[P3]). In this case, there are n minimization problems (4.10) to find $\mathbf{a}_s \in \mathbb{R}^m$, $s = 1, \ldots, n$, for components of $\mathbf{q}$. These minimizations can be concatenated for $s = 1, \ldots, n$ and written as

$$\arg\min_{A \in \mathbb{R}^{m \times n}} V(A) = \sum_{s=1}^{n} (\|Q_s - K\mathbf{a}_s\|^2 + \sigma^2 \mathbf{a}_s^T K \mathbf{a}_s), \tag{4.46}$$

where $A = [\mathbf{a}_1 \ldots \mathbf{a}_n]$. The minimization above is equivalent to

$$\arg\min_{A \in \mathbb{R}^{m \times n}} V(A) = \|Q - KA\|_F^2 + \sigma^2 \mathrm{Tr}(A^T K A), \tag{4.47}$$

where $\mathrm{Tr}(\cdot)$ is the trace and $\|\cdot\|_F$ is the Frobenius norm, i.e. $\|Q\|_F^2 = \sum_{j=1}^{m} \sum_{s=1}^{n} |q_{js}|^2$. The solution for (4.47) is written as

$$\hat{A} = [\hat{\mathbf{a}}_1 \ldots \hat{\mathbf{a}}_n] = (K + \sigma^2 I_m)^{-1} Q \tag{4.48}$$

and estimate for $\mathbf{x}$ is evaluated as

$$\hat{\mathbf{q}} = \hat{A}^T \mathbf{k}_{\mathbf{x}} = Q^T (K + \sigma^2 I_m)^{-1} \mathbf{k}_{\mathbf{x}}. \tag{4.49}$$

The estimate (4.49) is a linear combination of training spectra, i.e. $\hat{\mathbf{q}} = \sum_{j=1}^{m} p_j(\mathbf{x}) \mathbf{q}_j$, where $\mathbf{p}(\mathbf{x}) = (K + \sigma^2 I_l)^{-1} \mathbf{k}_{\mathbf{x}} \in \mathbf{R}^m$ denotes coefficient vector corresponding to measurement $\mathbf{x}$.

For semi-parametric models, the minimization problems (4.27) concatenated for $s = 1, \ldots, n$ are written as

$$\begin{aligned} &\arg\min_{B \in \mathbb{R}^{m \times n}, C \in \mathbb{R}^{N \times n}} \left(\|Q - KB - \Psi(X)C\|_F^2 + \sigma^2 \mathrm{Tr}(B^T K B)\right) \\ &\text{s.t.} \quad \Psi(X)^T B = \mathbf{0}, \end{aligned} \tag{4.50}$$

with solutions

$$\begin{aligned} \hat{B} &= U_2[\hat{\mathbf{d}}_1 \ldots \hat{\mathbf{d}}_n] = U_2(U_2^T K U_2 + \sigma^2 I)^{-1} U_2^T Q, \\ \hat{C} &= [\hat{\mathbf{c}}_1 \ldots \hat{\mathbf{c}}_n] = V_1 S_1^{-1} (U_1^T Q - U_1^T K \hat{B}), \end{aligned} \tag{4.51}$$

where matrices $U_1, U_2, V1$ and S_1 correspond to (4.29). Estimate for $\mathbf{x}$ is

$$\hat{\mathbf{q}} \;=\; \hat{B}^T \mathbf{k}_{\mathbf{x}} + \hat{C}^T \Psi(\mathbf{x}). \tag{4.52}$$

4.1.7 Computational cost

In 4.1.1, the primal and dual formulations for the estimation model with positive definite kernels were presented. If the RKHS in (4.7) is finite dimensional (e.g. consists of polynomials) and if we can explicitly evaluate the finite dimensional feature map Φ, the primal form calculation can be used. Coefficients in primal form (4.20) and dual form (4.14) can be calculated using standard least squares techniques ([17]) and computationally, the most efficient form depends on the feature space, cost of the kernel evaluation and the number of training samples.

Assume $T(n)$ is the number of operations needed in computation. The computational cost is denoted as $T = O(t)$ by assuming that there is $C > 0$ and function t, such that $T(n) \leq Ct(n)$ for $n \to \infty$. Let us assume that $\mathbf{x} \in \mathbb{R}^k$, $\Phi(\mathbf{x}) \in \mathbb{R}^N$, the size of the training is m, the evaluation cost of Φ is $O(N)$, the evaluation cost of $\kappa(\cdot,\cdot)$ is $O(k)$ (in worst case $O(N)$) and the inversion cost of $m \times m$ matrix with a naive implementation is $O(m^3)$. In a general vector valued regression, there are totally n scalar regression estimations $f_s(\mathbf{x}) = \mathbf{c}_s^T\Phi(\mathbf{x})$, $s = 1, \ldots, n$, corresponding to dimension of vector output. In primal form, the computational cost for one scalar regression is

$$\text{solve } \mathbf{c}_s : O(N^3 + mN), \text{ evaluate } f_s(\mathbf{x}) : O(N) \tag{4.53}$$

In dual form $f_s(\mathbf{x}) = \sum_{j=1}^{m} a_{sj}\kappa(\mathbf{x}_j, \mathbf{x})$, $s = 1, \ldots, n$ with $a_{sj} \neq 0$, for $j = 1, ..., m$. The cost of one scalar regression in dual form is

$$\text{solve } \mathbf{a}_s : O(m^3 + m^2k), \text{ evaluate } f_s(\mathbf{x}) : O(mk) \tag{4.54}$$

Several approximation methods for the solution (4.14) have been presented (e.g. [46], [53], [127], [133], [155]). As an example, the system in (4.14) is positive definite and iterative *conjugate gradient* method can be used to approximate $\mathbf{a}_s$ with computational cost $O(vm^2)$, where v is the number of iteration steps of the algorithm [46], p. 78.

For conditionally positive kernels, the solutions f_s, $s = 1, \ldots, n$, have representation $f_s(\mathbf{x}) = \mathbf{c}_s^T \Psi(\mathbf{x}) + \sum_{j=1}^{m} b_{sj} \kappa(\mathbf{x}_j, \mathbf{x})$, and the computational cost is

$$\text{solve}\, \mathbf{c}_s, \mathbf{b}_s : O(N_p^3 + mN_p + m^3 + m^2 k),\ \text{evaluate}\, f_s(\mathbf{x}) : O(mk + N_p), \tag{4.55}$$

where N_p correspond to dimension of the polynomial feature $\Psi(\mathbf{x})$. In this form, the computational cost of solving the coefficients is usually dominated by the inversion of kernel matrix. It has been shown for the thin plate splines that the solution can be approximated using $O(m \log m)$ operations [8].

As an practical example, the color calibration of RGB device consists of two phases. First, the measured data are used as a training data to formulate a model. The size m of the training set is often $9 \times 9 \times 9$ samples and consists of measured color targets with known color characteristics [88], p.395. The formulated estimation model is then used to populate a 3-D look-up-table e.g with $33 \times 33 \times 33$ samples [66], [88]. For final image processing, the look-up-table is used to interpolate color values for image. The population of look-up-table can be done off-line with the presented methods, but the calibration of all pixels in an image has to be fast and is performed with some interpolation method.

Similarly, reflectance estimation model in artwork imaging might have several minutes or even hours for image processing. However, there are also applications where high computational efficiency is needed. One example is the spectral video imaging for real-time transmission [171].

4.1.8 Measurement noise in camera responses

In practice, the measurements $\mathbf{x}$ and $\mathbf{q}$ from multispectral and hyperspectral devices contain also noise, which has effect on the estimation. Lack of control over the noise is one shortcoming of the empirical models (4.49) and (4.52), since no explicit assumptions about noise are included in these models.

One standard assumption for the noise in multispectral values is that of additive normally distributed noise with independently distributed over different channels [43], [113], [148]. In this case the noise can be filtered with two simple ways: 1) spatial averaging over pixels in the image, when the measured sample contains spatially homogeneous material or 2) in the measurement phase, by using temporal averages of multiple measurements of the same spatial position, when the measurement object is not moving. These are the approaches which were used for the data in this dissertation.

It is also possible to extend the spatial averaging in case 1) by using filtering techniques [115], [162]. For empirical reflectance estimation models, spatial noise removal filtering can be considered as a separate preprocessing step. It has been suggested, that preprocessing via a spatial filter is efficient for images where the SNR of the image channels is at moderate level ($>$ 35 dB) [115].

4.1.9 Model training for regression estimation

In the models (4.15) and (4.32), the kernel parameters and the regularization parameter need to be chosen carefully to achieve good accuracy and generalization properties. In the limiting case, we can choose the parameters to construct interpolators, which may be numerically ill-conditioned or overfit. Similarly we don't want to choose parameters that underfit the data.

For some applications, it is useful if the parameters can be chosen by using automatic techniques. For example, in the case of empirical reflectance estimation, we might be tempted to decompose the estimation to separate scalar regressions for every output dimension, so that the complete model is too complex to fine tune with subjective choices. In this dissertation, we promoted the simplicity of the models and focused on the use of (4.49) and (4.52), which lead to optimization of a single pair of kernel and regularization parameters. Similarly, automatic techniques are needed if there are several data sets, possibly measured with different devices, in which the estimations are applied.

Structural risk minimization (SRM) is a principle for model selection using finite data sets, which has been presented in the framework of statistical learning theory by V. Vapnik [163]. An overview of the statistical learning theory and SRM are presented in [33]. The idea of SRM is that the best model needs to be constrained to some appropriate hypothesis space (RKHSs) via regularization functionals, so that the minimizer of (4.7) is close to the minimizer of expected risk. In the SRM, the regularization parameter σ^2 is optimized as a function of finite training data set. If the hypothesis space corresponding to parameter σ^2 is too "large" with respect to properties of training data, the SRM theory indicates that model has poor generalization properties.

According to [33], there are practical difficulties of implementing the SRM framework. In practice, the search of the model parameters is done using approximative methods, such as *cross-validation* [44], [50], [167]. In N-fold cross validation, the training set is first separated into N equal-sized parts. The model is then trained using $N-1$ parts and tested with remaining portion. This procedure is repeated for all the N parts separately, by changing the training and test data. By using the search grids for model parameters, the average values of the chosen error value is calculated from all N test set separations. The parameters corresponding to the minimum average error are chosen to be the optimal values.

4.2 TRANSFORMATIONS OF REFLECTANCE SPECTRA

It is useful to include a priori knowledge to the estimation by using known physical properties of the reflectance spaces.

4.2.1 Logit function

In many cases it is reasonable to expect that the values of the reflectance spectrum $\mathbf{q}$ are constrained to region $[0,1]$. Especially, for the Lambertian surfaces and measurements with viewing solid angle of 2π, all the measurements are in this range due to physical reasons. However, also in many other cases for practical measurements and surfaces, the data seem to have the values in region $[0,1]$.

It is possible that for the presented reflectance estimation models (4.49) and (4.52), the $\hat{\mathbf{q}}$ may contain values, which are outside the region [0,1]. In order to address this issue, we introduce a transformation for the output training data $\{\mathbf{q}_i\}_{i=1}^m$ using the *logit function*

$$\tilde{\mathbf{q}} = \log\left(\frac{\mathbf{q}}{1-\mathbf{q}}\right), \tag{4.56}$$

where $\log :]0, +\infty[\to \mathbb{R}$ is the natural logarithm and can be evaluated elementwise for $\mathbf{q} \in]0, 1[^n$. The inverse transformation, i.e. the restoration from non-linear features is performed after the evaluation of the regression models and is defined using componentwise *logistic function*

$$\mathbf{q} = \frac{\exp(\tilde{\mathbf{q}})}{1+\exp(\tilde{\mathbf{q}})}. \tag{4.57}$$

In [P3], the experimental results were calculated using the transformation $\tilde{\mathbf{q}} = \operatorname{arctanh}(2\mathbf{q}-1)$, where $\operatorname{arctanh} : [-1, 1] \to \mathbb{R}$ is calculated elementwise. The inverse transformation is written as $\mathbf{q} = (1+\tanh(\hat{\mathbf{q}}))/2$. However, since we can write $\operatorname{arctanh}(2\mathbf{q} - 1) = 1/2\log(\mathbf{q}/(1-\mathbf{q}))$ this is equal to (4.56) up to a scalar.

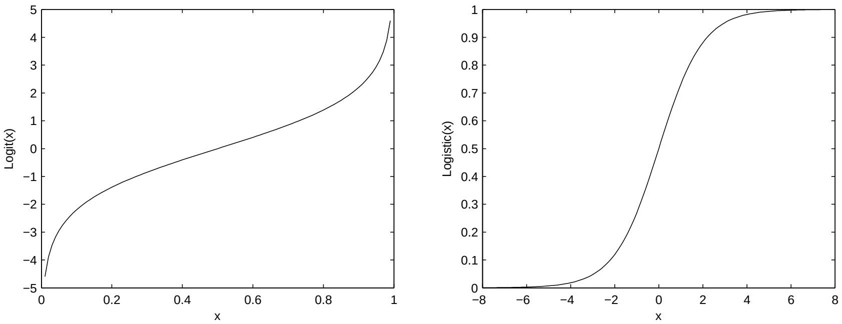

Figure 4.1: Left: Logit function (4.56), Right: Logistic function (4.57).

4.2.2 Principal component analysis

Principal component analysis (PCA) is a classical, statistically motivated method for data transformation and analysis [74]. It is widely used method also in the analysis and approximation of spectral data (e.g. [21], [57], [79], [94], [95], [100], [120], [123], [137]). PCA can be considered as a unsupervised method, in a sense that the data are projected to directions, which are not based on some fixed basis, but on the available data set itself. – From the point of view of matrix theory, the Singular Value Decomposition leads to the same representation of data as the PCA.

In PCA, we search for the direction $\mathbf{u}$ (unit vector), which maximizes the variance of the data $\{\mathbf{q}_i\}_{i=1}^m$

$$\arg\max_{\mathbf{u}} \mathrm{Var}(\mathbf{u}^T\mathbf{q}) \approx \arg\max_{\mathbf{u}} \sum_{i=1}^{m} (\mathbf{u}^T(\mathbf{q}_i - \mathbf{q}'))^2 = \arg\max_{\mathbf{u}} \mathbf{u}^T\hat{\Sigma}\mathbf{u}, \tag{4.58}$$

where $\mathbf{q}'$ is the estimated mean vector and $\hat{\Sigma}$ is the m-sample covariance matrix of the available data. It can be shown that the direction that maximizes the variance is given by the eigenvector corresponding to the largest eigenvalue of the covariance matrix [74]. Similarly, the second largest variance direction is defined by the eigenvector corresponding to second largest eigenvalue, and so on. Assuming that the rank of $\hat{\Sigma}$ is n, the coordinates of vector $\mathbf{q} \in \mathbb{R}^n$ in orthonormal PCA eigenvector basis $\{\mathbf{u}_i\}_{i=1}^n$ are defined as

$$\mathbf{q}^T\mathbf{u}_1, \ldots, \mathbf{q}^T\mathbf{u}_n, \tag{4.59}$$

and are called as *principal components* (PCs). In data approximation, only the $k << n$ largest eigenvalues and corresponding eigenvectors are chosen. Assuming zero mean data set, the approximation for element $\mathbf{q}$ is formulated as an orthogonal projection to the k-dimensional subspace

$$\hat{\mathbf{q}} = \sum_{i=1}^{k} \mathbf{u}_i \mathbf{q}^T \mathbf{u}_i. \tag{4.60}$$

The effective PCA dimension of hyperspectral reflectance data is usually significantly lower than the dimension of measurements. This has been empirically validated for several reflectance ensembles and hyperspectral images (e.g. [21], [77], [79], [92], [123], [158]).

Thererefore, PCA approximation is widely used in spectral data compression and noise removal applications, but also in the reflectance estimation to reduce large number of hyperspectral dimensions in vector valued regression. In the latter case, the reflectance measurements in $\mathbb{R}^n$ are replaced with corresponding truncated vectors of PCs. The representation in terms of PCs improves the practicality and computational efficiency, especially when vector valued regression is calculated via separate scalar regressions. In this case, the dual forms (4.15) and (4.32) lead to construction and storage of kernel matrices $K_i \in \mathbb{R}^{m \times m}$, $i = 1, \dots, n$ for every reflectance dimension, which may be impractical for large n.

Furthermore, the logit-logistic transformations (4.56)-(4.57) can be combined with PCA approximation to constrain the approximated reflectance values to the region [0,1]. In this case the data $\{\mathbf{q}_i\}_{i=1}^m$ is logit transformed via (4.56) before the calculation of covariance matrix. Inverse transformation (4.57) is calculated for the orthogonally projected subspace elements (4.60) in $\mathbb{R}^n$.

4.3 EXPERIMENTAL RESULTS AND DISCUSSION

First we give results for PCA approximation of Munsell reflectance ensemble (1269 samples in 400–700 nm, 5 nm sampling [161]) [116]. The approximation is calculated for logit transformed data as discussed above. The approximations were evaluated, by using the ΔE color errors (2.25) under D65 illumination and spectralwise root mean square error (RMSE). The RMSE between hyperspectral reflectance measurement $\mathbf{q} \in \mathbb{R}^n$ and estimate $\hat{\mathbf{q}} \in \mathbb{R}^n$ is defined as

$$\mathrm{RMSE}(\mathbf{q}, \hat{\mathbf{q}}) = \sqrt{\|\mathbf{q} - \hat{\mathbf{q}}\|^2 / n}. \tag{4.61}$$

These color and spectral errors are presented in Figures 4.2 and 4.3 as average and maximal values for the Munsell set. The results indicate fast improvement in accuracy, when the subspace dimension increases. The results for logit transformed data show that the average and maximal RMSE and ΔE color errors are lower when compared to standard PCA approximation. The RMSE errors are approximately 5-10% lower, when compared to standard PCA approximation for fixed dimensionality. The ΔE errors indicate even larger reduction from the errors for standard PCA approximations. The ΔE results for 3- and 4-dimensional subspaces show significant improvement, when compared to the standard PCA.

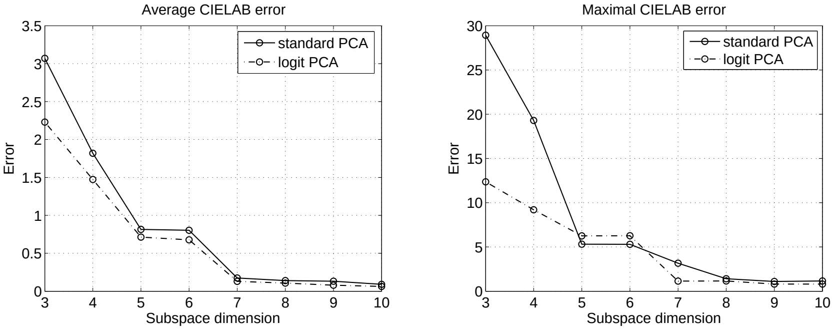

Figure 4.2: Average and maximum of CIELAB error as a function of subspace dimension. Data: Munsell Matte collection.

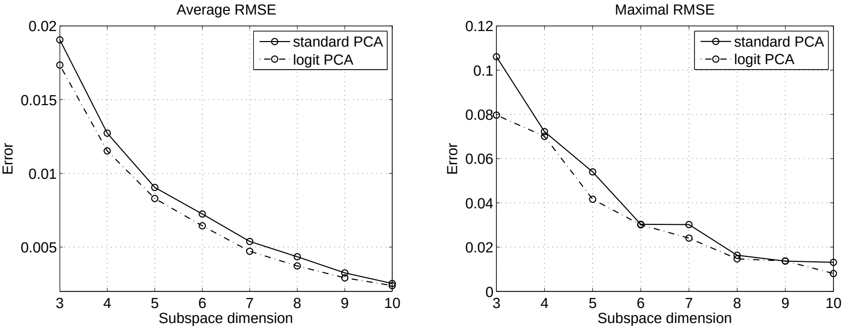

Figure 4.3: Average and maximum of RMS error as a function of subspace dimension. Data: Munsell Matte collection.

As a second experiment, we calculated reflectance estimation results by using model (4.49) with the Gaussian kernel and model (4.52) with the thin plate spline (TPS). The estimations were calculated by using real RGB data $\{\mathbf{x}_i\}_{i=1}^{1164} \subset \mathbb{R}^3$ of the Munsell Matte collection and corresponding reflectance data $\{\mathbf{q}_i\}_{i=1}^{1164} \subset \mathbb{R}^{61}$. The same data was used in [P1]–[P3]. The sets were separated randomly to training set (669 samples) and test set (495 samples) ten times.

Table 4.1 presents the average results with different pre-processing methods for reflectance outputs (logit, PCA and PCA+logit). In all cases, the RGB data was scaled to $[0, 1]^3$ and the pre-processed reflectance data was centered to have zero mean in training set. The parameters of the models were optimized using 10-fold cross validation (10-fold CV) and minimization of average RMSE. In the case of Gaussian kernel (3.52), the kernel parameter ς and regularization parameter σ^2 were both found by using the 10-fold CV. The degree (d in (3.77)) of the TPS kernel was fixed to 2 and only the regularization parameter σ^2 was optimized by 10-fold CV.

The results in Table 4.1 show, that the logit transformation and truncation to PCs leads to similar performance when compared to the use of measurements $\mathbf{q}$ in regression. Therefore, via these two transformations, the estimations are in $[0, 1]^n$ and can be computed independently of the dimension n of (4.2). The average RMSE accuracies for these kernels are close to the accuracy of four dimensional PCA approximation of the complete Munsell set (Fig. 4.3).

It can be seen that the TPS leads to improved performance over the Gaussian kernel, although the degree of TPS kernel was not optimized. This property of the TPS model can be found to be appealing from the practical point of view. As a comparison, Kang and Anderson have reported that CIEXYZ color calibration model based on a neural network (Cascade Correlation Network) is difficult to train and may lead to overfitting [78]. Those evaluations were done for a flatbed scanner by using a Kodak Q60-C target (236 color samples were used) and several training- and test set separations. They also reported that a least squares polynomial fit was able to provide good generalization properties for the same color set [78].

As an example of a computation time (by using a simple MATLAB implementation), an evaluation for 1 × 5000 pixel line of six-channel data was calculated in 0.39 seconds. In this calculation we used the dual form (4.49) with the Gaussian kernel, pre-calculated model coefficients, training set of 600 samples and a post processing with PCA projection (10 components) and logistic transformation. The spectral dimension (n) of the estimated output was 61 samples.

Table 4.1: RMSE errors for the measured Munsell matte multispectral data from Fujifilm S1 RGB-camera under fluorescent light source. Kernel and regularization parameters were chosen using 10-fold CV and intervals $\varsigma^2 \in [2^{-8}, 2^3]$ *and* $\sigma^2 \in [2^{-10}, 2^3]$ *for the Gaussian kernel. For the TPS kernel* $d = 2$ *and* $\sigma^2 \in [2^{-10}, 2^3]$. *In table, (logit): logit transformation for regression targets, (PCA): Regression targets are vectors of 10 most significant principal components calculated using the training set.*

Train: Munsell I, Test: Munsell II.			
Model	Avg.	Std.	Max.
Gaussian	0.0143	0.0103	0.0821
Gaussian (logit)	0.0141	0.0104	0.0846
Gaussian (PCA)	0.0144	0.0103	0.0818
Gaussian (logit+PCA)	0.0142	0.0104	0.0843
TPS	0.0137	0.0107	0.0757
TPS (logit)	0.0134	0.0106	0.0764
TPS (PCA)	0.0139	0.0106	0.0755
TPS (logit+PCA)	0.0135	0.0105	0.0763

In this case, a six-channel image with 5000 × 5000 pixels could be evaluated (linewise) in 32 minutes. With a training set of 100 samples, the evaluation time reduces to 6 minutes.

In addition to the above results, we summarize the reflectance estimation and color calibration results, which are presented in [P1]–[P3]. All the experimental results for reflectance estimation from multispectral data have been calculated using the standard least squares polynomial regression (primal form), dual form (4.49) for the Gaussian, polynomial and spline (3.85) kernel and model (4.52) for the TPS kernel. In [P1]–[P3], we evaluated the models by using four data ensembles: Munsell Book of Color - Matte Finish Collection (1269 color samples), Natural Color System (NCS) atlas (1750 color samples), Pantone (922 color samples), Macbeth chart (24 color samples). These are widely used ensembles in color science and their properties have been discussed in several works (see [79], [80], [91], [92] for references). We used four sensor systems (Fujifilm Finepix S1, Canon A20 Powershot, simulated Canon

EOS 10D and simulated Gaussian six-band system) and two light sources (Fluorescent Philips DeLuxe 90 TLD 18W/965 6300K source and Tungsten 100W source (see Fig. 2.4)). The error evaluations were done by using the RMSE (4.61), CIE ΔE^* (2.25) and CIE ΔE^*_{00} ([88], pp. 113–115.) by using CIE D65 illuminant.

4.3.1 Estimation with polynomials (P1)

In [P1], the estimation of reflectance spectra and CIE L*a*b* values are performed using regularized polynomial least squares fitting and real multispectral data from two RGB cameras. The estimation models were evaluated by using model parameters which minimized the average CIE ΔE^* error in a subset of Munsell set. The final evaluation of models was done for a separate set, which was not used in the parameter optimization.

The results suggest that for color engineering purposes, the reflectance estimation may provide good alternative over the estimation CIE L*a*b* coordinates. Although the average CIE color accuracy in reflectance estimation for Fujifilm Finepix S1 camera is 31.3% lower than in direct CIE L*a*b* calibration, the results suggest that the degradation might be tolerable when the light source independence of the data is taken into account. Interestingly, it was also found that the average color error for reflectance estimation using Canon A20 camera were 19.2% lower than the average color error in direct CIE L*a*b* estimation.

4.3.2 Estimation with kernel methods (P2)

In [P2], the reflectance estimation procedure is represented by using the kernel based framework, where regularized regression models were proposed, consisting of Thin plate spline (TPS), tensor splines, polynomials and Gaussian kernel regression. The estimation was performed by using the Munsell Matte collection, RGB data under D65 illumination and training sets with 100 and 600 samples. The training, parameter optimization and the final evaluation of models were done by using separate sets.

The differences between regularized polynomial model and polynomial and tensor spline (3.85) kernel models were found to be small for both training set sizes. These results suggest that the model choice in these cases should be based on computational efficiency. The Gaussian kernel (avg. RMSE of 0.0165, max. RMSE of 0.0688) and TPS kernel with degree 3 (avg. RMSE of 0.0154, max. RMSE of 0.0699) improved the results from other models for training set with 600 samples. When the TPS results for test set were compared to the best performing polynomial model (avg. RMSE of 0.0172, max. RMSE of 0.0855), the relative improvement in average RMSE accuracy was 10.5% and the relative improvement in maximal RMSE accuracy was 18.5%. In the case of training set with 100 samples, it was found that the regularized TPS (avg. RMSE of 0.0207, max. RMSE of 0.0883) gave similar performance in test set with the positive definite kernels in terms of average RMSE. However, the maximal RMSE of test set was smallest for the TPS kernel. As an example, in this case the average RMSE of 0.0214 and the maximal RMSE of 0.0981 were obtained for the Gaussian kernel.

It was found that the regularization of the solution is highly beneficial in order improve the generalization properties of the models. In the case of TPS, an interpolator (avg. RMSE of 0.0184, max. RMSE of 0.1074) and a model with regularization term (avg. RMSE of 0.0154, max. RMSE of 0.0699) were compared, in favor of the latter. The regularization also improves the numerical stability of primal and dual formulations, when the matrices in estimation models would otherwise be effectively ill-conditioned.

In a related work, Sharma and Shaw constructed a TPS model for color calibration of printer [144]. They suggested that a TPS model can slightly improve accuracy of a local regression model. Zhang, et.al. [172] generalized (4.7) with a different loss function. They presented reflectance estimation results by using a Support Vector Machine regression (SVR), which is based on (4.7) with epsilon insensitive loss function, i.e. $\mathcal{L}(y, f(\mathbf{x})) = |y - f(\mathbf{x})|_\epsilon$, where $|y - f(\mathbf{x})|_\epsilon = \max\{0, |y - f(\mathbf{x})| - \epsilon\}$. However, only small differences were found between the model (4.49) and a SVR model by using a training set with 600 Munsell samples and several test sets [172]. For a training set with 100 samples, it was found that a SVR model can improve the performance of (4.49) for some sets.

4.3.3 Evaluation and unification of methods (P3)

In publication [P3], we evaluated the reflectance estimation for real and simulated RGB data under multiple light sources by using RKHS models and the transformation in (4.56), (4.57). The simulations were based on numerical approximation of (4.1). An effort was made to emphasize the connections between different estimation models. Properties of the used data sets and some previously proposed estimation methods, such as locally defined regression, were also discussed and evaluated in this publication.

We presented estimation results by using kernel model (4.49) with the Gaussian kernel, different training data and different response systems. In the evaluation of training sets and response systems, we used the same kernel- and regularization parameters in the training of the model (4.49). The parameters were fixed in order to show that there exists parameters that can be used for different data sets and different multispectral sensors systems to obtain accurate, albeit suboptimal, reflectance estimation performance. The goodness of the estimation was evaluated relative to the performance of standard linear methods and PCA approximation. The logit transformation (4.56) was introduced and shown to lead to physically realistic estimations with good accuracy.

In the simulated case, the kernel matrix K in (4.48) was constructed by using the response functions of multispectral camera and spectral power distribution of light source. In this case, the estimates (4.48) with a linear kernel are identical to estimates from widely used linear physical model and the regularization parameter corresponds to the variance of normal i.i.d. noise in the responses. In this context, significant improvements can be obtained by replacing the linear kernel with a non-linear one. In terms of the relative increase in mean accuracy, the Gaussian kernel improved the RMSE results for linear kernel by 38.3%, when the training and test data were from the Munsell set.

The usage of multiple light sources improves the estimation accuracy significantly in simulated case. In a noise-free case, the average RMSE decreased 61.2% relative to the single light source, in a case where the training and test set were from the Munsell ensemble. In the simulated noisy case with 43 dB SNR gaussian noise, the improvement reduced to 23.0%. As a comparison, it was shown that the robustness to noise increased by using a another simulated six-band system instead of the RGB-two-light source system. The real RGB-two-light source system gave improvement of 14.0% relative to results with single tungsten source. The results indicate that in practice, a careful choice of light sources is important. The differences between the real and simulated data are partly explained by the use of non-optimal jpeg format for real measurements.

In Table 4.1, we reported estimation results for the real Munsell Matte RGB data by using the Gaussian kernel, logit transformation and 10-fold cross-validation routine in parameter optimization. The results in Table 4.1 suggest that the average RMSE accuracy improves 18% with fluorescent source (relative to the comparable results with fixed model parameters in [P3], Table 8.).

The model choice, optimization and performance are dependent on properties of training set. In [P3], we heuristically analyzed differences between training and test reflectance data by using *principal angles* between PCA subspaces [164], [42], pp. 603–604. It was found that the estimation results were significantly more accurate for sets with similar reflectance subspaces. In order to optimize the accuracy, the training targets may need to be tailored via *a priori* hyperspectral imaging for object or scene of interest. This can be seen as a *spectral colorization* (cf. [9], [114], [131]). The term *colorization* usually refers to estimation of complete RGB color image from a monochromatic image with partial color areas.

5 *Kernel based classification of spectral data*

In this chapter we focus on the classification of multispectral measurements from pine, spruce and birch tree species by using the kernel based feature spaces, presented in chapter 4. The idea of classification is to use a decision function to label a given feature vector to some predefined class. The construction of the decision function is done in a supervised way, i.e. by using a training set with known feature-label pairs. In the following, the multispectral measurements are simulated from hyperspectral data to correspond to measurements from photogrammetric airborne camera tailored for remote sensing applications.

5.1 REMOTELY SENSED TREE DATA

Remote sensing is an information acquisition technique where satellites or airborne vehicles are used to derive information about objects in the ground, properties of the earth surface itself or the atmosphere. Measurement of surface reflected electromagnetic spectrum in some wavelength region is one possible approach for the information extraction for objects in the ground. This is usually performed using recording device attached to the vehicle, such as digital multispectral or hyperspectral camera [19], [141]. If we compare the remote sensing to spectral measurements in laboratory conditions, it is evident that remote sensing data is affected by several parameters that are difficult to control, such as the atmospheric state and solar irradiance incident on the ground.

Tree species classification is an evident bottleneck in current remote sensing of forests, in spite of the ample research carried out into the use of airborne laser scanning and the digital aerial multispectral cameras [83]– [86]. It has been realized that in acquired images there exist severe radiometric normalization problems because trees are geometrically complex surface elements with varying reflectance behaviour (e.g. [27]). The structural effects of reflectance are due to branches, shoots and needles (or leaves), which result in illumination- and view geometry dependent surface Bidirectional Spectral Reflectance Distribution Function (BSRDF) [97]. The spectral mixing from some spatial neighborhood and background are also causing significant amount of normalization problems. The mutual shading and height variation of trees prevent from observing all trees in direct light and most of the smaller trees remain undetected. The constantly changing sunlit-shadow variation due to self- and neighbour shading causes the radiometric characterization of trees to be a difficult problem [83]– [86].

Recently, it has been suggested that multispectral sensors tailored originally for photogrammetric applications could be used also in quantitative remote sensing [39], [62]. Multispectral sensors can provide additional benefits over hyperspectral sensors owing to their high spatial resolution. In this context it is assumed that it is possible to construct an all-purpose sensor with small number of spectral response functions, which would allow adequate information extraction from the ground objects, both radiometrically and geometrically. The ADS40 sensor systems by the Leica Geosystems are constructed to follow this philosophy [11], [12], [39]. The pushbroom architecture of ADS40 reduces the variations in viewing direction and therefore should provide a good foundation for imaging of ground objects with viewing geometry dependent BSRDF. The ADS40 provides absolutely calibrated measurements in two fixed viewing directions with four spectral bands per direction (Fig. 5.1). The absolute calibration can be utilized in estimation of band-averaged reflectance values and BSRDF normalization [10], [12].

In this dissertation, the objects of interest are single trees in the ground. In Finland, there are three tree species, that are commercially significant: Scots pine (Pinus sylvestris L.), Norway spruce (Picea abies (L.) H. Karst.) and birch (Betula pubescens Ehrh., Be-

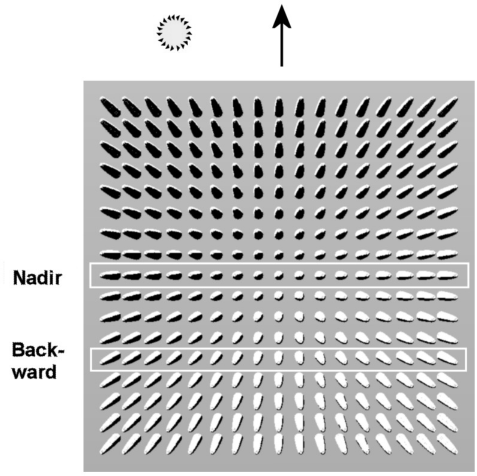

Figure 5.1: Illustration of the view-illumination geometry of trees in the focal plane of a nadir-looking aerial camera. The flying direction is upwards and the direction of the Sun is 26° to the left of the upward direction. The white rectangles depict the nadir- (0° along-track) and backward-viewing (16° along-track) CCDs in the ADS40 line sensor. (Figure: Ilkka Korpela).

tula pendula Roth). It has been suggested that the species classification accuracy should reach 90–95% to be adequate for stand management and local wood procurement planning [84]. Photogrammetric cameras are not customized for forestry applications but for surveying and mapping purposes. There is still a substantial lack of basic research into the spectral characteristics distinguishing given forest objects, and such information would be valuable for specifying optimal sensors designed for forest use [83]– [86].

The classification of objects in images is based on their geometrical or spectral features. At the single-tree level, important structures contribute to image textures only at high spatial resolutions, and due to low flight altitude, such images are often too expensive to acquire over large areas [83], [84]. As an example, the Ground Sampling Distance (GSD) of ADS40 system is 40 cm from the altitude of 4 km and suggests that this sensor is suitable for single tree imaging in a cost efficient way. In the following, we will focus entirely on features derived from pointwise multispectral measurements (pixel-based classification). These measurements are simulated from hyperspectral data to correspond to a ADS40 system.

5.2 CLASSIFICATION USING SEPARATING HYPERPLANES

Let us consider a two-class classification problem with training data

$$S = \{(\mathbf{x}_1, y_1), ..., (\mathbf{x}_m, y_m)\} \subset \mathcal{X} \times \{-1, 1\}. \tag{5.1}$$

We concentrate on the *hyperplane classifier* and assume that the representation of the *decision function* $f : \mathcal{X} \to \{-1, 1\}$ is

$$f(\mathbf{x}) = \text{sign}(\mathbf{c}^T \Phi(\mathbf{x}) + b), \tag{5.2}$$

where $\mathbf{c} = (c_1, c_2, \ldots)$ and $\Phi(\mathbf{x}) = (\sqrt{\gamma_1}\varphi_1(\mathbf{x}), \sqrt{\gamma_2}\varphi_2(\mathbf{x}), \ldots)^T$ correspond to RKHS feature map (3.47).

An affine hyperplane $\mathcal{P}$ which is associated to the decision function has the representation

$$\mathcal{P} = \{\Phi(\mathbf{x}) | \mathbf{c}^T \Phi(\mathbf{x}) + b = 0, \mathbf{x} \in \mathcal{X}\}. \tag{5.3}$$

In the following we formulate the Support Vector Machine (SVM) algorithm for finding a separating hyperplane $\mathcal{P}$ which maximizes the margin between two classes [163].

5.2.1 Hard margin Support Vector Machine

If the data can be separated by using a hyperplane in the feature space, it can be written that $f(\mathbf{x}_i) \geq +1$, when $y_i = +1$ and $f(\mathbf{x}_i) \leq -1$, when $y_i = -1$, for all $i = 1, \ldots, m$. In other words, we have a decision function which correctly classifies all the training points. The hyperplane $\mathcal{P}$ is called as a *canonical hyperplane* with respect to $\{\Phi(\mathbf{x}_i)\}_{i=1}^m$, when the points closest to the hyperplane have $|\mathbf{c}^T \Phi(\mathbf{x}_i) + b| = 1$, [151], p. 190. The unit normal vector of the surface is $\mathbf{c}^T / \|\mathbf{c}\|_2$ and has orthogonality property $\mathbf{c}^T / \|\mathbf{c}\|_2 (\Phi(\mathbf{x}) - \Phi(\mathbf{z})) = 0$ for the all the points $\Phi(\mathbf{x}), \Phi(\mathbf{z}) \in \mathcal{P}$. Furthermore $\mathbf{c}^T \Phi(\mathbf{x}) = -b$ for all $\Phi(\mathbf{x}) \in \mathcal{P}$. Now for a point $\Phi(\mathbf{v})$

not in $\mathcal{P}$, we can define distance to the point $\Phi(\mathbf{x})$ in $\mathcal{P}$ as the difference of the projections of these two points to the unit normal of the hyperplane

$$\left| \frac{\mathbf{c}^T}{\|\mathbf{c}\|_2}(\Phi(\mathbf{v}) - \Phi(\mathbf{x})) \right| = \frac{1}{\|\mathbf{c}\|_2} |\mathbf{c}^T \Phi(\mathbf{v}) + b| = \frac{1}{\|\mathbf{c}\|_2}, \tag{5.4}$$

where the last equality is obtained when the point $\Phi(\mathbf{v})$ from the class is closest to the hyperplane. The distance above is equivalent to the perpendicular distance of the point $\Phi(\mathbf{v})$ from the hyperplane $\mathcal{P}$. Therefore we have that the *margin* (i.e. smallest distance) between the two classes is $2/\|\mathbf{c}\|_2$.

If the data can be separated by a hyperplane, it can be written that

$$y_i(\mathbf{c}^T \Phi(\mathbf{x}_i) + b) \geq 1, \quad i = 1, \ldots, m. \tag{5.5}$$

The idea of SVM is to construct a hyperplane $\mathcal{P}$, so that Eq. (5.5) is satisfied and the margin between the two classes is maximized. Equivalently, by using (5.4), it is possible to minimize $\|\mathbf{c}\|^2/2$, which is the squared RKHS norm (3.27) of the decision function. The minimization problem is written as

$$\begin{aligned} &\min\nolimits_{\mathbf{c},b} \tfrac{1}{2}\mathbf{c}^T\mathbf{c}, \\ &\text{s.t.} \quad y_i(\mathbf{c}^T \Phi(\mathbf{x}_i) + b) \geq 1 \quad i = 1, \ldots, m, \end{aligned} \tag{5.6}$$

5.2.2 Soft margin Support Vector Machine

A more widely used variant of SVM (by Cortes and Vapnik [25]) is based on the assumptions that the two classes cannot be separated by a hyperplane, but they have some overlap in the feature space l^2. In this case the SVM model (*C-SVM*, *Soft margin SVM*) is derived as the solution to [33]

$$\min_{\mathbf{c},b} \frac{1}{2}\mathbf{c}^T\mathbf{c} + \frac{C}{m}\sum_{i=1}^{m} |1 - y_i(\mathbf{c}^T \Phi(\mathbf{x}_i) + b)|_+, \tag{5.7}$$

where the loss function is

$$|x|_+ = \begin{cases} x, & x > 0; \\ 0, & \text{otherwise.} \end{cases} \tag{5.8}$$

This minimization problem corresponds to (3.4), with $\sigma^2 = m/2C$. The (5.7) is equivalent to a quadratic optimization problem

$$\begin{aligned} &\min_{\mathbf{c},b,\xi} \tfrac{1}{2}\mathbf{c}^T\mathbf{c} + \tfrac{C}{m}\textstyle\sum_{i=1}^m \xi_i, \\ &\text{s.t.} \quad y_i(\mathbf{c}^T\Phi(\mathbf{x}_i) + b) \geq 1 - \xi_i \quad i = 1, \ldots, m, \\ &\text{and} \quad \xi_i \geq 0 \quad i = 1, \ldots, m, \end{aligned} \tag{5.9}$$

where the parameter C controls the penalization of the samples located on the incorrect side of the decision boundary, and $\{\xi_i\}_{i=1}^m$ are *slack variables* which indicate misclassification of sample $\mathbf{x}_i$ when $\xi_i > 1$ [33]. Since $\mathbf{c}^T\Phi(\mathbf{x}_i) + b$ is proportional to the distance of the point $\Phi(\mathbf{x}_i)$ from the hyperplane, it can be seen that the points far away from the decision boundary have small effect for the shape of the boundary, if they are on the correct side of the hyperplane (the distance to hyperplane is larger or equal than the margin and $\xi_i = 0$). The points which are on the wrong side of the margin have more effect to the optimization functional and to the final form of the decision boundary. For $C \to \infty$, this model approaches to the separable case (5.6), where misclassifications are not tolerated.

By following [151], the *Lagrangian* for this problem is written as

$$\begin{aligned} \mathcal{L}(\mathbf{c}, b, \xi; \alpha, \nu) &= \frac{1}{2}\mathbf{c}^T\mathbf{c} + \frac{C}{m}\sum_{i=1}^m \xi_i - \sum_{i=1}^m \alpha_i(y_i(\mathbf{c}^T\Phi(\mathbf{x}_i) + b) - 1 + \xi_i) \\ &\quad - \sum_{i=1}^m \nu_i\xi_i, \end{aligned} \tag{5.10}$$

with *Lagrange multipliers* $\alpha_i \geq 0$ and $\nu_i \geq 0$. For this Lagrangian we have saddle point

$$\max_{\alpha,\nu} \min_{\mathbf{c},b,\xi} \mathcal{L}(\mathbf{c}, b, \xi; \alpha, \nu), \tag{5.11}$$

which gives the solution for the problem. By taking the partial derivatives with respect to components of $\mathbf{c}$, b and ξ_i, $i = 1, \ldots, m$ and setting them to zero, we have

$$\begin{aligned}
&\mathbf{c} = \textstyle\sum_{i=1}^{m} \alpha_i y_i \Phi(\mathbf{x}_i), \\
&\textstyle\sum_{i=1}^{m} \alpha_i y_i = 0, \qquad (5.12) \\
&0 \leq \alpha_i \leq C/m \quad i = 1, \ldots, m.
\end{aligned}$$

By identifying $\mathbf{c}^T\mathbf{c} = \sum_{i,j=1}^{m} \alpha_i \alpha_j y_i y_j \Phi(\mathbf{x}_i)^T \Phi(\mathbf{x}_j)$, the solution is obtained from the convex quadratic optimization problem using the dual space of Lagrange multipliers

$$\begin{aligned}
&\min_\alpha \tfrac{1}{2}\alpha^T Q \alpha - \mathbf{1}^T \alpha, \\
&\text{s.t.} \quad \mathbf{y}^T \alpha = 0 \qquad (5.13) \\
&\text{and} \quad 0 \leq \alpha_i \leq C/m \quad i = 1, \ldots, m,
\end{aligned}$$

where $Q_{ij} = y_i y_j K_{ij}$ and for the i, j entries of kernel matrix K we have $\kappa(\mathbf{x}_i, \mathbf{x}_j) = \Phi(\mathbf{x}_i)^T \Phi(\mathbf{x}_j)$ by kernel property (3.51). In this way decision boundaries in the input space are defined without having to make explicit use of the possibly infinite dimensional features in space l^2. The slack variables vanish from the dual problem and only the constant C constrains the Lagrange multipliers. The term b can be derived from Karush-Kuhn-Tucker conditions [151], p. 206.

The decision function for the SVM can be written as a sparse dual representation of training points

$$f(\mathbf{x}) = \operatorname{sign}\Big(\sum_{i=1}^{m_s} \alpha_i y_i \kappa(\mathbf{x}, \mathbf{x}_i) + b\Big), \qquad (5.14)$$

where $\{\alpha_i\}_{i=1}^{m_s}$ are the calculated Lagrange multipliers and κ is the selected positive definite kernel function [151]. In the summation $m_s < m$, because the coefficients α_i are non-zero only for some training points. The training points $i = 1, \ldots, m_s$ with the above property are called as *support vectors*. The sparse representation can be a significant property in some applications, where fast evaluation is needed. Also the memory requirements for large training sets can be alleviated by this property.

In our case, the motivation to use the C-SVM was that there is large amount of empirical evidence that they produce excellent classification results for remotely sensed data. For example in case of hyperspectral AVIRIS data, it has been suggested that an SVM provides higher accuracy when compared to Regularized Radial Basis Function Networks (Including optimization of centers, variances and weights for units in hidden layer), kernel Fisher Discriminant and Regularized Ada-boost classifiers [18]. Also other studies in the remote sensing field suggest that SVMs compete well with the best available classifiers [18], [38], [65], [103], [105]. It has been suggested that SVM classifiers are suitable for remote sensing applications, since they are robust to noise and high-dimensional data [18]. Recent review arcticle shows that SVM models are currently one of the most actively studied methods in the field of remote sensing with over one hundred published works (April, 2010) [111].

5.3 SIMULATED MULTISPECTRAL RESPONSES

In order to study the classification of single trees, we simulated ADS40 responses by using the sensor model (2.17), available hyperspectral data of tree species and a daylight spectral power distribution [P4]. By considering only the nadir scan line (Fig. 5.1), the ADS40 is a four band system with wavelength supports in 428–492, 533–587, 608–662 and 833–887 nm [11], [90]. The corresponding spectral response functions (SRFs) were approximated by using rectangular functions.

The hyperspectral data were ground level reflectance factor measurements collected from Finland and Sweden. These measurements were carried out by Jääskelainen et. al, by using a PR 713/702 AM spectroradiometer in the range of 390–1070 nm with 4 nm sampling [76]. They report, that each radiance measurement represented an average spectrum of thousands of leaves (or needles) on a growing tree at a distance of 50 m. The measured crowns were thick in order to minimize effects of background illumination. In the measurement setting the sun was always behind the measurement direction, with a clear path towards to object. The solar vector had almost constant zenith angle, but the azimuth angle with respect to measurement direction had variation (see Fig. 2.1).

5.4 SVM CLASSIFICATION RESULTS FOR SIMULATED TREE DATA (P4)

In [P4], we performed the classification of spectral measurements from the canopy of pine, spruce and birch trees using the RKHS feature space mappings and the C-SVM. The C-SVM is a binary classifier and the ″one against one″ classification was used for this multiclass classification case ([P4], [103]). The McNemar's test was used to test statistical significances between classification results [30]. The data were simulated to corresponded to a: (1) 4-band ADS40 sensor, (2) 5-band sensor and (3) 93-band hyperspectral sensor. We compared the predictive power of the non-linear Gaussian- (3.52) and Mahalanobis kernels to the linear first degree polynomial kernel (3.55). The Mahalanobis kernel is an anisotropic kernel (3.54), where $A = \Sigma^{-1}$ and Σ is a covariance matrix. This kernel can be identified as the Gaussian kernel (3.52), with data *whitened* in input domain. In whitening, the data are first represented in orthogonal directions defined by the eigenvectors of Σ (training set) and scaled to have equal variance in this representation.

The results for the simulated 4-band ADS40 sensor suggested that it is highly beneficial to use a non-linear kernel. The use of a non-linear kernel compensated for the poor spectral accuracy of the system, and led to an improvement of 5–23 percentage points in the misclassification performance when it was compared to the first degree polynomial kernel. However, the high information content of 93-band data allowed accurate classification also with the use of linear kernel. In this case it may be advisable to avoid the use of non-linear kernels, in order to avoid over-fitting effects.

We studied how an additional SRF would improve the classification performance of the ADS40 system. In the first case, an additional SRF located in 705–755 nm was chosen to correspond to a existing 5-band system [11]. In the second case, the band selection was considered by using maximal average distance between probability density functions of simulated responses. The band selection method was based on the maximization of average *Jeffries-Matusita distance* between the three tree classes [160]. In this method, it is assumed that the class conditional probability density functions are normal. During the calculations, two fixed bandwidths, 15 nm and 30 nm were used and the position of new SRFs were constrained to be in the 662–833 nm region, which is not supported by the 4-band ADS40. The band selection led to two positions: (1) 695–725 (30 nm bandwidth) nm and (2) 710–725 nm (15 nm bandwidth). The calculations led to the same positions for the two bands, if the constraint for wavelength region was removed, but it was assumed instead, that there was no overlap with the original bands. The locations of all the SRFs were in the ″red edge″, which refers to the 680–730 nm region of change in reflectance of chlorophyll [134].

The results suggested that significant improvements for classification performance can be obtained by using an additional fifth SRF. In our case, the band in 710–725 nm region gave the best results, with an average misclassification ratio of 15%.

It was assumed that the classification performance in simulated ADS40 response space decreased due to outliers in the class samples. It can be assumed that some outliers in data were due to the varying measurement conditions (e.g. variable sensor-object-sun geometry) or measurement errors, for instance. However, the set of outliers include also samples due to natural BSRDF variation within- and between species. In practice it may be difficult to remove disrupting data variation automatically in an way that no essential information is removed from training sets. In our case, the results suggested that the Mahalanobis kernel with $\varsigma = 1$ processed the outliers automatically and increased the classification accuracy.

6 Discussion and conclusions

In this dissertation we have studied (1) the color calibration of a camera, (2) the estimation of reflectance spectra, and (3) the supervised classification of spectral information, based on reproducing kernel Hilbert space methods (RKHS). We considered reflectance estimation as a generalized color calibration problem and mainly focused on an empirical regression approach that assumes relatively large ensembles of training data. Although the physical characteristics of the camera and the light source are not explicitly used in model construction, the empirical approach can be motivated for many applications by its simplicity, cost-efficiency and practicality.

The performance of the estimation and classification models depends on the number of spectral response functions, spectral properties of the objects, properties of the available training data, and on measurement conditions. A unifying characteristic of our data was that there was only a small number of individual broad-band spectral response functions available. It was shown that non-linear feature maps via RKHS kernels improved the accuracy of estimation and classification markedly as compared to standard linear models. The latter usually work well for measurement devices with a high number of spectral bands with narrow-band characteristics.

In [P1] we started with the evaluation of polynomial models for reflectance estimation and color calibration. Using machine learning terminology, these correspond to primal form kernel models. These were the starting point for the dual form kernel models in publications [P2] and [P3]. The regularization of polynomial models was found to be important, when the training sets were small and the degree of polynomials increased. In [P2] we introduced and evaluated the performance of several kernel functions in estimation and interpolation. We found that the spectral and CIE color accuracies of regularized kernel models with the Gaussian and thin plate splines (TPS) were higher than the accuracies of widely used polynomial models. In [P3] we discussed the connections between several empirical and physical reflectance estimation models. It was shown that the kernel based models can effectively utilize the spectral characteristics of measurement devices and information from measurements under multiple light sources. The evaluations in [P3] suggested that the accuracy of kernel methods were higher than the accuracies of classical, physical and empirical linear estimation models. Here, the evaluations of models were performed by using real or simulated data (based on simple BSRDF model) of widely used Munsell Matte, NCS, Pantone and Macbeth color ensembles. The real RGB data of Macbeth and Munsell data in [P1]–[P3] were utilized by using the JPEG format due to the limitations of utilized RGB devices. It is expected that the estimation accuracy improves further if RGB devices capable to produce raw sensor measurements are utilized.

The RKHS based regression and interpolation models include subclasses that have previously been used for reflectance estimation and color calibration, such as linear and polynomial least squares fits and radial basis function networks. Despite this, their connection to the more general RKHS model has not been discussed adequately. A central aim of the core text of the dissertation has been to fill this gap.

Similarly, regularization is not currently used in most reflectance estimation and color calibration applications. Based on our calculations, it can be concluded that regularized learning has a high potential in these areas to produce accurate and numerically stable estimates. As an example, the regularized thin plate splines (TPS) were shown ([P2]) to lead to better accuracy in reflectance estimation than the interpolating TPS. To the best of our knowledge this is also the first study, that investigates the use of TPS for reflectance estimation.

Physical a priori knowledge of reflectance data was also introduced in order to improve the efficiency and accuracy of reflectance estimation. We presented details of how combination of logit transformation and Principal Component Analysis (PCA) can be used as preprocessing tool in reflectance estimation.

The kernel models have primal and dual formulations, that can both in principle be used in practical computation. In most cases the dual form is simpler to solve. However, if the feature map is finite dimensional, the primal form can also be effectively solved using least squares. We have primarily intended the reflectance estimation methods for off-line applications, but they may also be used to form Look-Up-Tables for fast on-line computing.

The pixel-based classification of tree species (Birch, Pine and Spruce) was carried out using simulated data based on multispectral airborne camera in [P4]. A significant amount of redundancy exists in spectral radiance from natural objects and intelligent signal measurement is appropriate. This was achieved using 4-band and 5-band systems based on an ADS40 sensor. The classifier was a Support Vector Machine (SVM) combined with RKHS feature space mappings. It was shown that large differences in performance exist between linear, Gaussian and Mahalanobis kernels. Our results suggest that a higher number of bands, or new positioning and modification of the bands are needed to improve the accuracy of classification.

Our findings suggest several topics for future work. (1) In reflectance estimation, we concentrated on the use of squared error loss function. Overall, it can be concluded that the kernel methods with squared error loss are relatively simple to implement in practice. In this case only the solution of linear equations are needed. However, it is known that the squared error loss may have problems with robustness. (2) The Gaussian process formulation can also be utilized to derive confidence intervals for the estimated spectra and classification probabilities. (3) It seems possible to include spatial and textural information to estimation and classification in order to reduce noise and improve accuracy (cf. [114], [115] and [162]). (4) In simulated tree species classification and sensor optimization, future work is needed to improve the quality of reflectance data sets and to model the atmosphere more accurately. The simulations were based on ground-level measurements, and a natural extension is to repeat the simulations with real airborne hyperspectral data. (5) It is also important to investigate differences between the performance of simulated and real multispectral measurements in reflectance estimation and classification. We are currently comparing the tree species classification results obtained for simulated data to results for real ADS40-SH52 multispectral data [55], [86].

Bibliography

[1] Abe, S., "Training of Support Vector Machines with Mahalanobis kernel," in *ICANN 2005, Lecture Notes on Computer Science* **3697**, Springer-Verlag, 571–576, 2005.

[2] Alsam, A., Lenz, R. "Calibrating color cameras using metameric blacks" *J. Opt. Soc. Am. A* **24**(1), 11–17 (2007).

[3] Amidror, I., "Scattered data interpolation methods for electronic imaging systems: a survey," *J. Electron. Imaging* **11**(2), 157–176 (2002).

[4] Arai, Y., Nakauchi, S. and Usui, S. "Color Correction Method Based on the Spectral Reflectance Estimation using a Neural Network" In *Proceedings of Fourth Color Imaging Conference Color Science, Systems, and Applications*, 5–9, Scottsdale, Arizona, November, 1996.

[5] Aronszajn, N., "Theory of reproducing kernels," *Transactions of the American Mathematical Society* **68**(3), 337–404 (1950).

[6] Bala, R. *Digital Color Imaging Handbook, Chapter 5: Device Characterization*, Gaurav Sharma Ed. (CRC Press, 2003).

[7] Barnard, K., Funt, B. "Camera Characterization for Color Research", *Color Res. Appl.* **27**(3), 153–164 (2002).

[8] Beatson R.K, Cherrie J.B. and Mouat C.T. "Fast fitting of radial basis functions: Methods based on preconditioned GMRES iteration" *Adv. Comput. Math.* **11**(2–3), 253–270 (1999).

[9] Bochko, V., Parkkinen, J. "A spectral color analysis and colorization technique" *IEEE Computer Graphics and Applications* **26**(5), 74–82, 2006.

[10] Beisl, U., Woodhouse, N. “Correction of atmospheric and bidirectional effects in multispectral ADS40 images for mapping purposes,” in *Proc. XXth Congress of the ISPRS*, Istanbul, Turkey, 12–23, July, 2005.

[11] Beisl, U. “Absolute spectroradiometric calibration of the ADS40 sensor,” *Congrs ISPRS Commission Technique I. Symposium*, Marne-la-Valle, France, 14–18., 2006.

[12] Beisl, U., Telaar, M., Schönemark, V. “Atmospheric correction, reflectance calibration and BRDF correction for ADS40 image data,” *International Archives of Photogrammetry and Remote Sensing*, Vol. 27, Part B1, 7–12, 2008.

[13] Berns, R.S. *Billmeyer and Saltzman's Principles of Color Technology, 3rd ed.* (Wiley & Sons. Inc., New York, USA, 2000).

[14] Berns, R.S., Taplin, L.A., Nezamabadi, M., Mohammadi, M., Zhao, Y., “Spectral imaging using a commercial color-filter array digital camera,” *ICOM, 14th Triennal ICOM-CC meeting*, Hague, Netherlands, 743–750, September, 2005.

[15] Berns, R.S., “Color-accurate image archives using spectral imaging”, in *Scientific Examination of Art: Modern Techniques in Conservation and Analysis*, 105–119 (National Academy Press, Washington, USA, 2005).

[16] Bertero, M., Poggio, T. and Torre, V. “Ill-posed problems in early vision,” in *Proceedings of the IEEE* **76**, 869–889 (1988).

[17] Björck, Å *Numerical Methods for Least Squares Problems* (SIAM, Philadelphia, 1996).

[18] Camps-Valls, G., Bruzzone, L. “Kernel-Based Methods for Hyperspectral Image Classification”, *IEEE Trans. Geosci. Remote Sens.* **43**(6), 1351–1362 (2005).

[19] Chang, C-I. *Hyperspectral Imaging: Techniques for Spectral Detection and Classification.* (Kluwer academic/Plenum Publishers, New York, 2003).

[20] Cheung, V., Westland, S., Connah, R. And Ripamonti, C. "A Comparative Study of The Characterization of Colour Cameras by Means of Neural Network and Polynomial Transforms," *Journal of Coloration Technology* **120**(1), 19–25 (2004).

[21] Cohen, J. "Dependency of the spectral reflectance curves of the Munsell color chips", *Psychon. Sci.* **1**, 369–370 (1964).

[22] Cohen, J.B and Kappauf, W.E "Metameric color stimuli, fundamental metamers and Wyszecki's metameric blacks", *American Journal of Psychology* **95**, 537–564 (1982).

[23] Cohen, J.B and Kappauf, W.E "Color mixture and fundamental metamers: Theory, algebra, geometry, application" *American Journal of Psychology* **98**, 171–259 (1985).

[24] Connah, D. R., Hardeberg, J. Y. Spectral recovery using polynomial models, *Proceedings of the SPIE Vol. 5667*, 65–75, 2005.

[25] Cortes, C. and Vapnik, V. "Support vector networks," *Machine Learning* **20**, 273–297 (1995).

[26] Cucker, F. and Smale, S. "On the mathematical foundations of learning," *Bulletin of the American Mathematical Society* **39**, 1–49 (2002).

[27] Deering, D.W., Eck, T.F, Banerjee, B. "Characterization of the reflectance anisotropy of three boreal forest canopies in spring-summer," *Remote Sens. Environ.* **67**(2), 205–229, 1999.

[28] Dicarlo, J. M., Xiao, F. and Wandell, B. A. "Illuminating illumination," *The 9th Color Imaging Conference: Color Science and Engineering: Systems, Technologies, Applications*, 27–34 (Scottsdale, Arizona, USA, 2001).

[29] Dicarlo, J. M. and Wandell, B. A. "Spectral estimation theory: beyond linear but before Bayesian," *J. Opt. Soc. Am. A* **20**, 1261–1270 (2003).

[30] Dietterich, T.G. "Approximate Statistical Tests for Comparing Supervised Classification Learning Algorithms," *Neural Computation* **10**(7), 1895–1923 (1998).

[31] Duchon, J. "Spline Minimizing Rotation-Invariant Semi-Norms in Sobolev Space," *Constructive Theory of Functions of Several Variables*, Springer-Verlag, Berlin (1977).

[32] Ebner, M. *Color Constancy* (Wiley, 2007).

[33] Evgeniou, T., Pontil, M. and Poggio, T. "Regularization networks and support vector machines," *Adv. Comput. Math.* **13**, 1–50 (2000).

[34] Fairchild, M.D. *Color Appearance Models* (Addison-Wesley, Reading, MA, 1998).

[35] Farrell, J. E., Cupitt, J., Saunders, D. and Wandell, B.A. "Estimating spectral reflectances of digital artwork," in *Proc. of the International Symposium on Multispectral Imaging and Color Reproduction for Digital Archives*, Chiba University, Chiba, Japan, 1999, 42–49.

[36] Fellman, A., Heikkinen, V. Salminen, A., Hiltunen, J., Martinkauppi, B, Hauta-Kasari, M. "The effect of parameter change in CO2 laser welding on spectra," *25th International Congress on Applications of Lasers and Electro Optics*, Scottsdale, Arizona, USA, Oct. 30-Nov. 2, 2006.

[37] Finlayson, G.D., Hordley, S.D. and Hubel, P.M. "Color by correlation: A simple unifying framework for color" *IEEE Trans. Pattern Anal. Mach. Intell.* **23**(11), 1209–1221 (2001).

[38] Foody, G. M., Mathur, A. "A relative evaluation of multiclass image classification by support vector machines," *IEEE Trans. Geosci. Remote Sens.* **42**(6), 1335–1343 (2004).

[39] Fricker, P. "Raising the bar for multi-band, high-resolution airborne imagery", In *Photogrammetric Week 07*; Fritsch, D. (ed.). Wichmann Verlag. Heidelberg, Germany, 71–79, 2007.

[40] Fält, P., Hiltunen, J., Hauta-Kasari, M., Sorri, I., Kalesnykiene, V., and Uusitalo, H. "Extending Diabetic Retinopathy Imaging from Color to Spectra" In *SCIA 2009, Lecture Notes in Computer Science* **5575**, Springer-Verlag, 149–158, 2009.

[41] Girosi, F., Jones, M. and Poggio, T. "Regularization theory and neural network architectures," *Neural Computation* **7**, 219–269 (1995).

[42] Golub, G.H. and Van Loan, C.F. *Matrix Computations* (Johns Hopkins University Press, Baltimore, 1996).

[43] Gonzales, R. C. and Woods R. E. *Digital Image Processing*, 2^{nd} *edition* (Prentice Hall, New Jersey, 2002).

[44] Green, P. and MacDonald, L. *Colour Engineering, Achieving Device Independent Colour* (John Wiley & and Sons, West Sussex, 2002).

[45] Green, P.J. and Silverman, B.W. *Nonparametric Regression and Generalized Linear Models. A Roughness Penalty Approach* (Chapman & Hall, London, 1994).

[46] Hamers, B. *Kernel Models for Large Scale Applications,* (PhD Thesis, Katholieke Universiteit Leuven, 2004).

[47] Haneishi, H., Hasegawa, T., Hosoi, A., Yokoyama, Y. Tsumura, N. and Miyake, Y. "System design for accurately estimating the reflectance spectra of art paintings," *Appl. Opt.* **39**, 6621–6632 (2000).

[48] Hardeberg, J.Y. *Acquisition and Reproduction of Color Images - Colorimetric and Multispectral Approaches* (Universal Publishers/Dissertation.com, Parkland, USA, 2001).

[49] Hardeberg, J.Y. "Filter selection for multispectral color image acquisition", *J. Imaging Sci. Technol.* **48**(2), 105–110 (2004).

[50] Hastie, T., Tibshirani, R. and Friedman, J. *The Elements of Statistical Learning: Data mining, Inference and Prediction* (Springer-Verlag, New York, 2001).

[51] Hauta-Kasari, M. *Computational Techniques For Spectral Image Analysis* (PhD Thesis, Lappeenranta University of Technology, 1999).

[52] Hauta-Kasari, M., Miyazawa, T., Toyooka S., Parkkinen, J. "Spectral vision system for measuring color images" *J. Opt. Soc. Am. A* **16**(10), 2353–2362 (1999).

[53] Haykin, S. *Neural Networks: a comprehensive foundation, 2nd ed.* (Prentice-Hall, New Jersey, 1999).

[54] Heikkinen, V., Jetsu, T., Parkkinen, J. and Jääskeläinen, T. "Estimation of Reflectance Spectra Using Multiple Illuminations," *CGIV 2008/MCS'08, 4th European Conference on Colour in Graphics, Imaging, and Vision*, Terrassa-Barcelona, Spain, June 9–13, 272–276, 2008.

[55] Heikkinen, V., Korpela, I., Honkavaara, E., Tokola, T., Parkkinen, J., "An SVM classification of tree species radiometric signatures based on the Leica ADS40 sensor," *IEEE Trans. Geosci. Remote Sens., in press*, 2011.

[56] Hernández-Andrés, J., Romero, J., García-Beltrán, A. and Nieves, J. L. "Testing Linear Models on Spectral Daylight Measurements," *Appl. Opt.* **37**, 971–977 (1998).

[57] Hernández-Andrés, J., Romero, J., Nieves, J. L., and Lee, R. L. "Color and spectral analysis of daylight in southern Europe," *J. Opt. Soc. Am. A* **18**(6), 1325–1335 (2001).

[58] Hernández-Andrés, J., Nieves, J. L., Valero, E. M., and Romero, J. "Spectral-daylight recovery by use of only a few sensors," *J. Opt. Soc. Am. A* **21**(1), 13–23 (2004).

[59] Herzog, P.G., Knipp, D., Stiebig, H. and König, F. "Colorimetric characterization of novel multiple-channel sensors for imaging and metrology", *J. Electron. Imaging* **8**, 342–353, 1999.

[60] Hyvärinen, T., Herrala, E., Jussila, J. "High Speed Hyperspectral Chemical Imaging," *1st European Conference on Process Analytics and Control Technology*, EuroPACT2008, Frankfurt, Germany, April, pp. 22–25, 2008.

[61] Hiltunen, J. *Accurate Color Measurement* (PhD Thesis, University of Joensuu, 2002).

[62] Honkavaara, E. 2008. *Calibrating digital photogrammetric airborne imaging systems using a test field* (PhD thesis, Helsinki University of Technology, 2008).

[63] Hong, G., Luo, M.R. and Rhodes, P.A. "A study of digital camera colorimetric characterization based on polynomial modeling," *Color Res. Appl.* **26**, 76–84 (2001).

[64] Horn, B.K.P and Sjoberg, R.W "Calculating the reflectance map," *Appl. Opt.* **18**(11), pp. 1770–1779 (1979).

[65] Huang, C., Davis, L. S. and Townshend, J. R. G. "An assessment of support vector machines for land cover classification," *Int. J. Remote Sens.* **23**(4), 725–749 (2002).

[66] Hung, P.C. "Colorimetric calibration in electronic imaging devices using a look-up table model and interpolations," *J. Electron. Imaging* **2**(1), 53–61 (1993).

[67] Hunt, R.W.G *Measuring Colour* (Fountain Press, Kingston-upon-Thames, England, 1998).

[68] Hurlbert, A. C. *The Computation of Color*, PhD Thesis (MIT, 1989).

[69] Imai, F. H. and Berns, R. S. "Spectral estimation using trichromatic digital cameras," in *Proceedings of the International Sympo-*

sium on Multispectral Imaging and Color Reproduction for Digital Archives, 42–49, (Chiba, Japan, 1999).

[70] Imai, F. H., Rosen, M. R. and Berns, R. S. "Comparison of spectrally narrow-band capture versus wide-band with a priori sample analysis for spectral reflectance estimation", in *Proceedings of Eighth Color Imaging Conference: Color Science and Engineering, Systems, Technologies and Applications*, IS&T, Springfield, 2000, 234–241.

[71] Imai, F. H., Berns, R.S and Tzeng, D. "A comparative analysis of spectral reflectance estimation in various spaces using a trichromatic camera system", *J. Imaging Sci. Technol.* **44**, 280–287 (2000).

[72] Imai, F.H. and Berns, R.S. "Spectral estimation of oil paints using multi-filter trichromatic imaging," in *Proceedings of the 9th Congress of the International Color Association, Rochester, USA,* Vol 4421, R.Y. Chung and A.B.J. Rodrigues, eds. (SPIE, 2001), 504–507, 2001.

[73] Jetsu, T. *Modeling color vision* (PhD Thesis, University of Eastern Finland, 2010).

[74] Jolliffe, I.T *Principal Component Analysis* (Springer, New York, 1986).

[75] Johnson, W.R., Wilson, D.W., Fink, W., Humayun, M., Bearman, G. "Snapshot hyperspectral imaging in ophthalmology", *J. Biomed. Opt.* **12**(1), 014036-1–014036-7 (2007).

[76] Jääskelainen, T., Silvennoinen, R., Hiltunen, J. and Parkkinen, J. "Classification of the reflectance spectra of pine spruce, and birch," *Appl. Opt.* **33**(2), 2356–2362 (1994).

[77] Kaarna, A. *Multispectral Image Compression Using The Wavelet Transform* (PhD Thesis, University of Lappeenranta, 2000).

[78] Kang, H.R. Anderson, P.G. ""Neural network applications to the color scanner and printer calibrations" *J. Electron. Imaging* **1**(2), 125–135 (1992).

[79] Kohonen, O., Parkkinen, J. and Jääskeläinen, T. "Databases for spectral color science," *Color Res. Appl.* **31**, 381–388 (2006).

[80] Kohonen, O. *Retrieval of Databased Spectral Images* (PhD Thesis, University of Joensuu, 2007).

[81] Kolmogorov, A.N. and Fomin, S.V. *Introductory real analysis* Translated and edited by A. Silverman (Dover, New York, 1975).

[82] Kolmogorov, A.N. and Fomin, S.V. *Elements of the Theory of Functions and Functional Analysis* (Dover, New York, 1999).

[83] Korpela, I. *Individual tree measurements by means of digital aerial photogrammetry.* (Silva Fennica Monographs 3, 2004).

[84] Korpela, I. and Tokola, T. "Potential of aerial image-based monoscopic and multiview single-tree forest inventory - a simulation approach," *Forest Science* **52**(2), 136–147 (2006).

[85] Korpela, I., Rohrbach, F. "Variation and anisotropy of reflectance of forest trees in radiometrically calibrated airborne line sensor images - implications to species classification," in *ISPRS TC VII Symp. – 100 Years ISPRS, IAPRS, Vol. XXXVIII, Part 7B*, W. Wagner, B. Székely Eds., Vienna, Austria, July 5–7, 2010, 342–347.

[86] Korpela, I., Heikkinen, V., Honkavaara, E., Rohrbach, F., Tokola, T. "Variation and directional anisotropy of reflectance at the crown scale - implications to species classification in digital aerial images," *Remote Sens. Environ., in press*, 2011.

[87] Laamanen, H. *Spectral color and spectral color image analysis,* (PhD Thesis, University of Joensuu, 2007).

[88] Lee, H.-C. *Introduction to Color Imaging Science* (Cambridge University Press, 2009).

[89] Lee, H.-C. "Modeling Light Reflection for Computer Color Vision," *IEEE Trans. Pattern Anal. Mach. Intell.* **12**(4), 402–409 (1990).

[90] Leica Geosystems, "2008 ADS80 datasheet," URL: http://www.leica-geosystems.com/.

[91] Lehtonen, J., Parkkinen, J. and T. Jääskeläinen, "Optimal sampling of color spectra," *J. Opt. Soc. Am. A* **23**, 2983–2988 (2006).

[92] Lehtonen, J. *Spectral sampling and spectral image compression,* PhD Thesis (University of Joensuu, 2009).

[93] Lenz, R. "Estimation of illumination characteristics" *IEEE Trans. Image Process.* **10**(7), 1031–1038 (2001).

[94] Lenz, R. "Two stage principal component analysis of color", *IEEE Trans. Image Process.* **11**(6), 630–635 (2002).

[95] Lenz, R. and Bui, T.H. "Statistical properties of color signal spaces", *J. Opt. Soc. Am. A* **22**(5), 820–827 (2005).

[96] Lenz., R. "Spectral color spaces: Their structure and transformations", (In book: *Advances in Imaging and Electron Physics* **138**, 1–67 (2005).

[97] Li, X., Strahler, A. H., "Geometric-optical bi-directional reflectance modeling of a coniferous forest canopy", *IEEE Trans. Geosci. Remote Sens.* **24**, 281–293 (1986).

[98] Luo, M. R., Cui, G. and Rigg, B. "The development of the CIE 2000 colour-difference formula," *Color Res. Appl.* **26**, 340–350 (2000).

[99] Maloney, L. T. and Wandell, B. A. "Color constancy: a method for recovering surface spectral reflectance," *J. Opt. Soc. Am. A* **3**, 29-33 (1986).

[100] Maloney, L.T. “Evaluation of linear models of surface spectral reflectance with small numbers of parameters,” *J. Opt. Soc. Am. A* **3**, 1673–1683 (1986).

[101] Mansouri, A., Marzani, F.S., Gouton, P. “Neural networks in two cascade algorithms for spectral reflectance reconstruction”, *IEEE International Conference on Image Processing (ICIP 2005)*, 718–721, 2005.

[102] Mathews, J.H. *Numerical Methods for Mathematics, Science and Engineering* (Prentice-Hall International, New Jersey, 1992).

[103] Melgani, F., Bruzzone, L. “Classification of hyperspectral remote sensing images with support vector machines,” *IEEE Trans. Geosci. Remote Sens.* **42**(8), 1778–1790 (2004).

[104] Meinguet, J. “Multivariate interpolation at arbitrary points made simple”, *Journal of Applied Mathematics and Physics (ZAMP)* **30**, 292–304 (1979).

[105] Mercier, G. and Lennon, M. “Support Vector Machines for Hyperspectral Image Classification with Spectral-based kernels,” *Geoscience and Remote Sensing Symposium*, IGARSS ’03. Proceedings (IEEE International), 2003.

[106] Micchelli, C.A. “Interpolation of scattered data: Distance matrices and conditionally positive definite functions”, *Constructive Approximation* **2**, 11–22 (1986).

[107] Minh, H. Q., Niyogi, P. and Yao, Y. “Mercer’s Theorem, Feature Maps, and Smoothing,” *Lecture Notes in Computer Science* **4005**, 154–168 (2006).

[108] Miyata, K., Laamanen, H., Jaaskelainen, T., Hauta-Kasari, M. and Parkkinen, J. “Application of Spectral Information to Investigate Historical Materials – Detection of Metameric Color Area in Icon Images” in *Lecture Notes in Computer Science, 2005* **3540**, 193–202, 2005.

[109] Miyata, K. “Development of practical investigation system for cultural properties based on a projector-camera system,” in *proceedings: Color Imaging XIV: Displaying, Processing, Hardcopy, and Applications* Eds. R. Eschbach, G. G. Marcu, S. Tominaga, Alessandro Rizzi, 2009.

[110] Miyazawa, K, Hauta-Kasari, M., and Toyooka, S., “Rewritable broad-band color filters for spectral image analysis”, *Optical Rev.* **8**(2), 112–119, 2001.

[111] Mountrakis, G., Im, J., Ogole, C. “Support vector machines in remote sensing: A review,” *ISPRS Journal of Photogrammetry and Remote Sensing, in press (online version available)*, 2010.

[112] Morovic, P. and Finlayson G.D. “Metamer-set-based approach to estimating surface reflectance from camera RGB” *J. Opt. Soc. Am. A* **23**(8), 1814–1822 (2006).

[113] Murakami, Y., Obi, T., Yamaguchi, M. and Ohyama, N. “Nonlinear estimation of spectral reflectance on Gaussian mixture distribution for color image reproduction”, *Appl. Opt.* **41**, 4840–4847 (2002).

[114] Murakami, Y., Letomi, K., Yamaguchi, M. and Ohyama, N. “Maximum a posteriori estimation of spectral reflectance from color image and multipoint spectral measurements”, *Appl. Opt.* **46**, 7068–7082 (2007).

[115] Murakami, Y., Fukura, K., Yamaguchi, M. and Ohyama, N. “Color reproduction from low-SNR multispectral images using spatio-spectral Wiener estimation”, *Optics Express* **16**(6), 4106–4120 (2008).

[116] Munsell Book of Color, Matte Finish Collection (Munsell Color, Baltimore, USA, 1976).

[117] Nicodemus, F.E, Richmond, J.C., Hsia, J.J., Ginsberg, I.W. and Limperis, T. “Geometrical considerations and nomencla-

ture for reflectance ," Monograph 160 (U.S Bureau Standards (U.S.), 1977).

[118] O'hagan, A., "Curve Fitting and Optimal Desing for Prediction," *Journal of the Royal Statistical Society B* **40**, 1–42 (1978).

[119] Ohta, N., Robertson, A.R. *Colorimetry: Fundamentals and Applications (The Wiley-IS & T Series in Imaging Science and Technology)* (John Wiley and Sons Ltd, 2005).

[120] Oja, E. *Subspace Methods of Pattern Recognition* (Research Studies Press LtD. Letchworth, England, 1983).

[121] Park, J-I., Lee, M-H., Grossberg, M. D. and Nayar, S. K. "Multispectral Imaging Using Multiplexed Illumination," in *ICCV 2007, Eleventh IEEE International Conference on Computer Vision*, 14–21, 2007.

[122] Parkkinen, J. *Subspace methods in two machine vision problems,* (PhD Thesis, University of Kuopio, Finland, 1989).

[123] Parkkinen, J., Hallikainen, J. and Jääskeläinen, T. "Characteristic spectra of Munsell colors," *J. Opt. Soc. Am. A* **6**, 318–322 (1989).

[124] Parthasarathy, K.R. and Schmidt, K. "Positive definite kernels, continuous tensor products, and central limit theorems of probability theory," *Lecture notes in Mathematics* **272**, Springer-Verlag, Berlin (1972).

[125] Phillips, G.M. and Taylor, P.J. *Theory and Applications of Numerical Analysis* (Academic Press, New York, 1973).

[126] Poggio, T., Torre, V., Koch, C., "Computational vision and regularization theory", *Nature* **317**, 314–319 (1985).

[127] Poggio, T. and Girosi, F. "Networks for approximation and learning," in *Proceedings of the IEEE* **78**, 1481–1497 (1990).

[128] Poggio, T. and Girosi, F. “Regularization algorithms for learning that are equivalent to multilayer networks,” *Science* **247**, 978–982 (1990).

[129] Poggio, T., Mukherjee, S., Rifkin, R., Rakhlin, A. and Verri, A. “B,” in *Uncertainty in Geometric Computations*, J. Winkler and M. Niranjan, eds., (Kluwer Academic Publishers, 2002), 131–141.

[130] Pratt, W.K. and Mancill, C.E. “Spectral estimation techniques for the spectral calibration of color image scanner”, *Appl. Opt.* **15**, 73-75 (1976).

[131] Quang, M.H., Kang, S.H., and Le, T.M. “Image and Video Colorization Using Vector-Valued Reproducing Kernel Hilbert Spaces” *J. Math. Imaging Vision* **37**(1), 49–65, 2010.

[132] Rao, C.R. *Linear Statistical inference and Its Applications, 2nd edition*. (John Wiley & Sons Ltd., 1973).

[133] Rasmussen, C.E. and Williams, C. *Gaussian Processes for Machine Learning*, (MIT Press, 2006).

[134] Rautiainen, M., Lang, M., Möttus, M., Kuusk, A., Nilson, T., Kuusk, J., Lukk, T. “Multi-angular reflectance properties of a hemiboreal forest: An analysis using CHRIS PROBA data,” *Remote Sens. Environ.* **112**(5), 2627–2642 (2008).

[135] Ribes, A., Schmitt, F., Brettel, H. “Reconstructing Spectral Reflectances of Oil Pigments with Neural Networks” in *Proc. 3rd Int. Conf. on Multispectral Color Science*, 9-12, Joensuu, Finland, June, 2001.

[136] Ribes, A., Schmitt, F. “A fully automatic method for the reconstruction of spectral reflectance curves by using mixture density networks”, *Pattern Recognit. Lett.* **24**(11), 1691–1701 (2003).

[137] Romero, J., García-Beltrán, A., Hernández-Andrés, J. “Linear bases for representation of natural and artificial illuminants,” *J. Opt. Soc. Am. A* **14**(5), 1007–1014 (1997).

[138] Saitoh, S. *Theory of Reproducing Kernels and Its Applications* (Longman scientific and technical, Harlow, UK, 1988).

[139] Saunders, C., Gammerman, A. and Vovk, V. "Ridge Regression Learning Algorithm in Dual Variables", In J. Shavlik editor, *Machine learning: Proceedings of the 15th International Conference (ICML '98)*. Morgan Kaufman, 1998, 515–521 (1998).

[140] Schaback, R. and Wendland, H. "Approximation by Positive Definite Kernels", in: Advanced Problems in Constructive Approximation (M.D. Buhmann and D.H. Mache, eds.), *International Series in Numerical Mathematics* **142**, 203–221 (2002).

[141] Schowengerdt, R. A. *Remote sensing: Models and methods for image processing* (3^{rd} Ed., Elsevier, 2007).

[142] Shafer, S.A. "Using color to separate reflection components," *Color Res. Appl.* **10**(4), 210–218 (1985).

[143] Shawe-Taylor, J. and Christianini, N. *Kernel Methods for Pattern Analysis* (Cambridge University Press, 2004).

[144] Sharma, G. and Shaw, M.Q. "Thin-plate splines for printer data interpolation," *14^{th} European Signal Processing Conference*, Florence, Italy, September, 4–8, 2006.

[145] Shen, H-L. and Xin, J. H. "Spectral characterization of a color scanner by adaptive estimation", *J. Opt. Soc. Am. A* **21**, 1125–1130 (2004).

[146] Shen, H-L. and Xin, J. H. "Spectral characterization of a color scanner based on optimized adaptive estimation", *J. Opt. Soc. Am. A* **23**, 1566–1569 (2006).

[147] Shimano, N. "Optimization of spectral sensitivities with Gaussian distribution functions for a color image aquisition device in the presence of noise," *Opt. Eng.* **45**(1), 013201, 2006.

[148] Shimano, N. “Recovery of Spectral Reflectances of Objects Being Imaged Without Prior Knowledge”, *IEEE Trans. Image Process.* **15**, 1848–1856 (2006).

[149] Shimano, N.,Terai K., and Hironaga, M. “Recovery of spectral reflectances of objects being imaged by multispectral cameras”, *J. Opt. Soc. Am. A* **24**, 3211–3219 (2007).

[150] Smith, V. C. & Pokorny, J., “Spectral sensitivity of the foveal cone photopigments between 400 and 500 nm”, *Vision Res.* **15**, 161–171 (1975).

[151] Schölkopf, B. and Smola, A.J. *Learning With Kernels* (MIT Press, Cambridge, 2002).

[152] Solli, M., Andersson, M., Lenz, R., Kruse, B. “Color Measurements with a Consumer Digital Camera Using Spectral Estimation Techniques”, *Lecture Notes on Computer Science* **3540**, 110–114, 2005.

[153] Stigell, P., Miyata, K. and Hauta-Kasari, M. “Wiener estimation method in estimation of spectral reflectance from rgb images,” *Pattern Recognit. Image Anal.* **15**, 327–329 (2005).

[154] Styles, I.B., Calcagni, A., Claridge, E., Orihuela-Espina, F., Gibson, J.M. “Quantative analysis of multi-spectral fundus images”, *Med. Image Anal.* **10**, 578–597 (2006).

[155] Suykens, J.A.K., Van Gestel, T., De Brabanter, J., De Moor, B. and Vandewalle J. *Least Squares Support Vector Machines* (World Scientific Publishing, 2002).

[156] Takeya, M., Tsumura, N., Haneishi, H. and Miyake, Y. “Estimation of Transmittance Spectra from Multiband Micrographs of Fungi and its Application to Segmentation of Conidia and Hyphae,” *Appl. Opt.* **38**(16), 3644–3650 (1999).

[157] Taplin, L.A. and Berns, R.S. “Practical spectral capture systems for museum imaging,” in *Proceedings of the 10th Congress*

of the International Color Association, Granada, Spain, J.L. Nieves and J. Hernandez-Andres, eds. (Comite Espanol del Color, Granada, Spain, 2005), 1287–1290.

[158] Thomson, M. and Westland, S. "Spectral color statistics of surfaces," in *Color Image Science: Exploiting Digital Media*, Edited by L. W MacDonald and M. R. Luo (John Wiley and Sons Ltd., 2002).

[159] Tominaga, S. "Color coordinate conversion via neural networks," in *Color Imaging. Vision and Technology*. Edited by L. W. MacDonald and M. R. Luo (John Wiley & Sons Ltd., 165-178, 1999).

[160] Tsai, F. and Philbot, W.D. "Derivative-Aided Hyperspectral Image Analysis System for Land-Cover Classification", *IEEE Trans. Geosci. Remote Sens.* **40**(2), 416–425 (2002).

[161] University of Eastern Finland Color Group, "Spectral Database", http://spectral.joensuu.fi/.

[162] Urban, P., Rosen, M. R. and Berns, R. S. "Spectral image reconstruction using an edge preserving spatio-spectral Wiener estimation", *J. Opt. Soc. Am. A* **26**(8), 1865–1875 (2009).

[163] Vapnik, V. *Statistical Learning Theory* (John Wiley and Sons, New York, 1998).

[164] Vora, P. L. and Trussell, H. J. "Measures of goodness of a set of color scanning filters", *J. Opt. Soc. Am. A* **10**(7), 1499–1508 (1993).

[165] Vrhel, M.J. and Trussel, H.J. "Color device calibration: A mathematical formulation," *IEEE Trans. Image Process.* **8**, 1796–1806 (1999).

[166] Vrhel, M.J. and Trussel, H.J. "Color scanner calibration via a neural network," in *Proceedings of the 1999 IEEE International Conference on Acoustics, Speech, and Signal Processing* **6**, pp. 3465–3468, 1999.

[167] Wahba, G. *Spline Models for Observational Data*, (SIAM CBMS-NSF Regional Conference series in Applied Mathematics, vol.59), (SIAM, Philadelphia, 1990).

[168] Wandell, B.A. *Foundations of Vision* (Sinauer Associates, Inc. Sunderland, MA, 1995).

[169] Wendland, H. *Scattered Data Approximation* (Cambridge Monographs on Applied and Computational Mathematics. Cambridge University Press, 2005).

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

[171] Yamaguchi, M., Haneishi, H., Ohyama, N. "Beyond Red-Green-Blue (RGB): Spectrum-based Color Imaging Technology," *J. Imaging Sci. Technol.* **52**(1), 010201-1–010201-25 (2008).

[172] Zhang, W-F. and Dai, D-Q. "Spectral reflectance estimation from camera responses by support vector regression and a composite model," *J. Opt. Soc. Am. A* **25**(9), 2286–2296 (2008).

Ville Heikkinen

Kernel methods for estimation and classification of data from spectral imaging

This study concentrated on color calibration of a camera, estimation of reflectance spectra, and supervised classification of spectral information, based on reproducing kernel Hilbert space (RKHS) methods. We mainly focused on an empirical regression approach that assumes relatively large ensembles of training data. Several RKHS models and transformations were introduced and evaluated in these tasks.

UNIVERSITY OF
EASTERN FINLAND

Publications of the University of Eastern Finland
Dissertations in Forestry and Natural Sciences

isbn 978-952-61-0423-2
issn 1798-5668